20 The Infant's Visual World

The Everyday Statistics for Visual Learning

Swapnaa Jayaraman and Linda B. Smith

The prowess of human vision is central to many domains of human intelligence (DiCarlo & Cox, 2007). We discriminate thousands of individual faces, recognize thousands of object categories, and excel under challenging visual conditions. We can become visual experts in recognizing birds, mathematical equations, art, and more. The developmental path to these achievements is protracted with mature levels of competence not fully reached until adolescence (Hadad, Maurer, & Lewis, 2011; Nishimura, Scherf, & Behrmann, 2009). The evidence also shows marked changes in infancy in the domains of face processing and object recognition. The evidence makes clear that visual experience itself is a significant driver of these changes (Maurer, Mondloch, & Lewis, 2007; Smith, 2009). A complete theory of learning in any domain requires an understanding of both the learning mechanisms and the experiences – the data – on which those mechanisms operate. Although much research is directed to determining the developing mechanisms, little has been directed to visual experience itself. Seminal research indicates that the changes in sensory-motor abilities in the first 2 years of postnatal life define a curriculum of changing visual tasks that spur developments in perception and cognition (Bertenthal, Campos, & Barrett, 1984; E. W. Bushnell & Boudreau, 1993; Iverson, 2010; Soska, Adolph, & Johnson, 2010). Accordingly, this chapter focuses on an emerging research that attempts to measure the natural statistics of infants' everyday visual environments and how they change with sensory-motor development.

20.1 Egocentric Vision

Considerable progress in understanding adult vision has been made by studying the visual properties of scenes represented by photographs of the natural (Geisler, 2008; Simoncelli, 2003) and artifact-filled (Brockmole & Henderson, 2006; Castelhano & Witherspoon, 2016; Im, Park, & Chong, 2015; Wolfe, Võ, Evans, & Greene, 2011) worlds. These studies analyze the statistical regularities in lower-level visual features such as spatial frequency, contrast, orientations, mid-level ensemble statistics, and higher-level contents such as object categories. The findings yield two general conclusions about mature visual prowess. First, the precision and sensitivity of adult visual processing from lower to higher levels in the visual system aligns closely to the regularities in these scenes (e.g., Geisler, 2008; Simoncelli, 2003). Second, adult visual attention, discrimination, and categorization exploit these predictive regularities to yield contextually nuanced visual intelligence that homes in on the relevant visual properties for specific tasks and context (Brockmole & Henderson, 2006; Wolfe et al., 2011). The photographs that are the bases for these analyses, however, are biased (Braddick & Atkinson, 2011; Smith, Yu, Yoshida, & Fausey, 2015) by a mature body and visual system that purposely holds the camera and selects the content and frames of the scene. What is needed to understand the visual statistics of everyday infant environments, given their bodies and sensory-motor abilities, are scenes captured from the developing infant's point of view (Franchak & Adolph, 2010; Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008; Yurovsky, Smith, & Yu, 2013).

Egocentric vision uses head cameras and head-mounted eye trackers to study vision from the perspective of freely moving individuals. Analyses of these first-person scenes show that the way people look at the world when they are moving and performing everyday tasks differs fundamentally from the way they take photographs and how they look at still images (Fathi, Ren, & Rehg, 2011; Foulsham, Walker, & Kingstone, 2011; Tatler, Hayhoe, Land, & Ballard, 2011). The rationale for an egocentric vision approach to studying infant visual environments is threefold (see Figure 20.1).

First, relevant visual information is not framed by a purposely held and still camera but is the image projected to the retina. This *proximal* image is determined by (1) the *intrinsic* properties of the to-be-perceived or *distal* object; and (2) *extrinsic* factors including the spatial relation of the object to the sensors, the nature of the illumination, and the journey taken by photons to reach the sensors. The most fundamental unsolved problem in all of human vision is how we perceive the intrinsic properties of the distal object as constant – a round cup is round – given the transformational effects of extrinsic factors on the retinal image. Developmental research on visual cognition has often skipped over this core problem to the higher-order tasks of discrimination and classification, and has used clean adult canonical images (an upright cup on a white background). But all vision begins with the image projected to the retina. Head-camera images capture a reasonable approximation (Yoshida & Smith, 2008) and force visual analyses to reckon with the extrinsic variability (Figure 20.1A). Second, an individual perceiver's view is highly selective. The bottle, the ball, and parts of the blocks and the car in Figure 20.1B are in the infant's view and are captured by the head camera. Many things in the room that are spatially near the infant – the dog, the train, her mother's face – are not in view unless the infant turns her head and looks. The perceiver's location, posture, and ability or motivation to change their posture systematically bias egocentric visual information. Third, if bodies, posture, and interests change with development, which they do, then the statistical regularities in

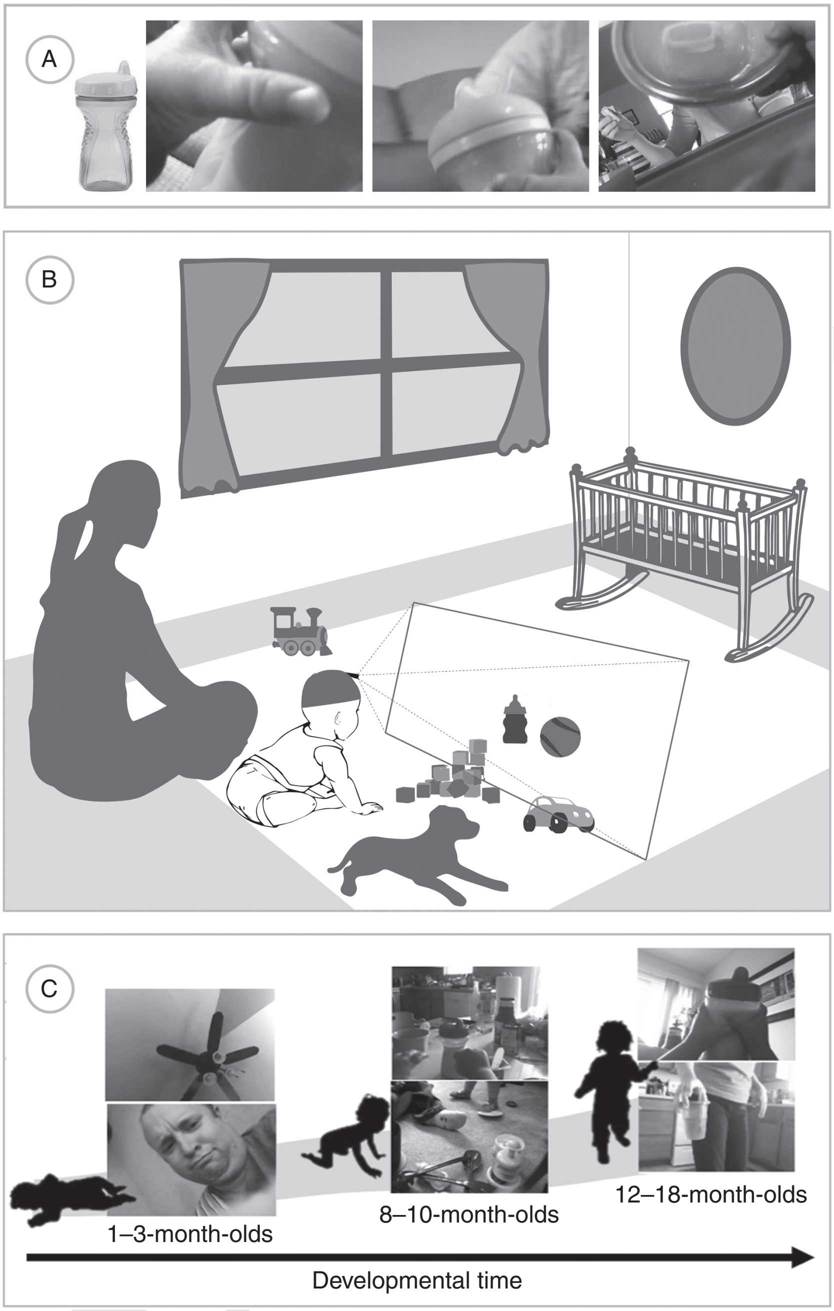

Figure 20.1. *An egocentric approach to the study of visual environments. (A) Canonical views of objects look different from egocentric views that more closely approximate the proximal images that fall on visual sensors. (B) Views of the environment are determined by the perceiver’s location, posture, and motorical abilities. (C) Perceivers’ views change systematically with development.*

the visual environment will also change. Each new sensory-motor achievement milestone – rolling over, reaching, crawling, walking, manipulating objects – opens and closes gates to specific visual experiences (Figure 20.1C). Newborns have limited vision and locomotion. Much of what they see depends on what caregivers put in front of and close to the infant's face, which is often their own face (Jayaraman, Fausey, & Smith, 2015, 2017; Sugden, Mohamed-Ali, & Moulson, 2014). An older crawling baby can see much further and can move to a distant object for a closer view. When moving, the crawler creates new patterns of dynamic visual information or optic flow (Gilmore, Baker, & Grobman, 2004; Higgins, Campos, & Kermoian, 1996) and sees only the floor. She actually has to stop crawling and sit up to see social partners (Franchak, Kretch, & Adolph, 2018). How do these egocentric experiences change and how do those changes inform visual development?

20.2 Biased Early Experiences

Human face perception is remarkable for its precision and for its relevance to species-important tasks (Haxby, Hoffman, & Gobbini, 2000). Although the exact mechanisms behind the development of face perception are still debated, it is understood that it develops over an extended period and is tuned by experience (Carey & Diamond, 1994; Gauthier & Tarr, 1997; Johnson & Morton, 1991; Kanwisher, 2000). The first 3 months of postnatal life appear especially important. At birth, infants show a bias towards high-contrast, low-spatial-frequency face-like patterns (Fantz, 1963; Johnson, Dziurawiec, Ellis, & Morton, 1991; Macchi, Turati, & Simion, 2004), which may be related to their nascent ocular structure that limits abilities to bring objects into focus (Dobson, Teller, & Belgum, 1978; Maurer & Lewis, 2001a, 2001b; Oruç & Barton, 2010). Despite these limitations, or perhaps partly because of them, infants during these first 3 months preferentially look at frontal views of faces close to their own and recognize and discriminate faces that are similar (in race, gender, age) to those of their caregivers (W. Bushnell, 2003; Pascalis et al., 2014; Scott, Pascalis, & Nelson, 2007; Scherf & Scott, 2012). Consistent with these early biases and learning, head-camera studies (Scherf & Scott, 2012; Sugden et al., 2014) of face experiences in everyday life indicate that infants primarily encounter faces that are frontal views, close, female, and of the same race as the infants themselves (Jayaraman et al., 2015; Sugden et al., 2014).

One series of studies is based on a large corpus of head-camera recordings of 51 infants (25 females) between the ages of 1 and 15 months (Fausey, Jayaraman, & Smith, 2016; Jayaraman et al., 2015, 2017; Jayaraman & Smith, 2018). Recordings of everyday scenes were collected as infants went about their daily activities with no experimenters present and in multiple locations (home, playground, store, church; see Jayaraman et al., 2015 for details of data collection). The average length of recordings from each infant was over 4 hours, resulting in a corpus of over 25 million frames.

Figure 20.2A shows the proportion of images in which a face was present as a function of the age of individual infants. For infants 3 months and younger, faces were present in over 15 minutes per recorded hour; for infants between 12 and 15 months, faces were present in only in about 6 minutes per recorded hour. The face views for all infants were mostly frontal but views of younger infants featured more up-close faces (Jayaraman et al., 2015; see Figure 20.2B) that were also temporally more enduring (Jayaraman & Smith, 2018; see Figure 20.2C). In brief, the data for learning about faces appears to have unique properties – frequency, proximity, and duration – for infants 3 months of age and younger.

Studies of infants whose everyday visual experiences were disrupted by institutionalization or early visual problems suggest that observed natural statistics of early face experiences may be crucial to mature face processing (Maurer, 2017; Maurer et al., 2007; Moulson, Westerlund, Fox, Zeanah, & Nelson, 2009). Institutionalized infants who lacked typical exposure to caretakers show measurable oddities when tested for their neural responses to face stimuli (Moulson et al., 2009). Individuals with congenital cataracts that were removed between 4 to 6 months of age also lack typical early visual experiences and show permanent deficits in configural face processing (Maurer, 2017; Maurer et al., 2007). Configural face processing – the sensitivity to second-order relations that adults can use effectively only with upright faces, and only when low spatial frequencies are present – is a late-developing property of the human visual system, one that only emerges in childhood and is not fully mature until adolescence (Scherf & Scott, 2012). Thus, these deficits in early visual experience are characterized as "sleeper effects," a late-emerging consequence of much earlier sensory deprivation (Maurer et al., 2007).

Infants' early and apparently crucial experiences of faces are substantial *but not massive*. By the end of the first 3 months, using the estimates of 15 minutes of face time per hour and 12 waking hours a day, an infant would have experienced 270 hours of predominantly close frontal views of faces (Jayaraman et al., 2015). The consequences of missing these 270 hours of experience appear to be permanent deficits not counteracted by a lifetime of seeing faces. Thus, the first 3 months of postnatal life may be characterized as a sensitive period of development in which specific experiences have an outsized effect on long-term outcome. While sensitive periods may reflect fundamental changes in neural plasticity (Oakes, 2017), developmentally changing visual environments may also play a role. Infants with cataracts removed at 4 months may not "catch up" in face processing not solely because of limited neural plasticity but because they do not encounter the same structured data set: dense close frontal views of faces enduring in time, experiences dependent on the infant's own sensory-motor, cognitive, and emotional abilities and the caregiver behaviors they elicit.

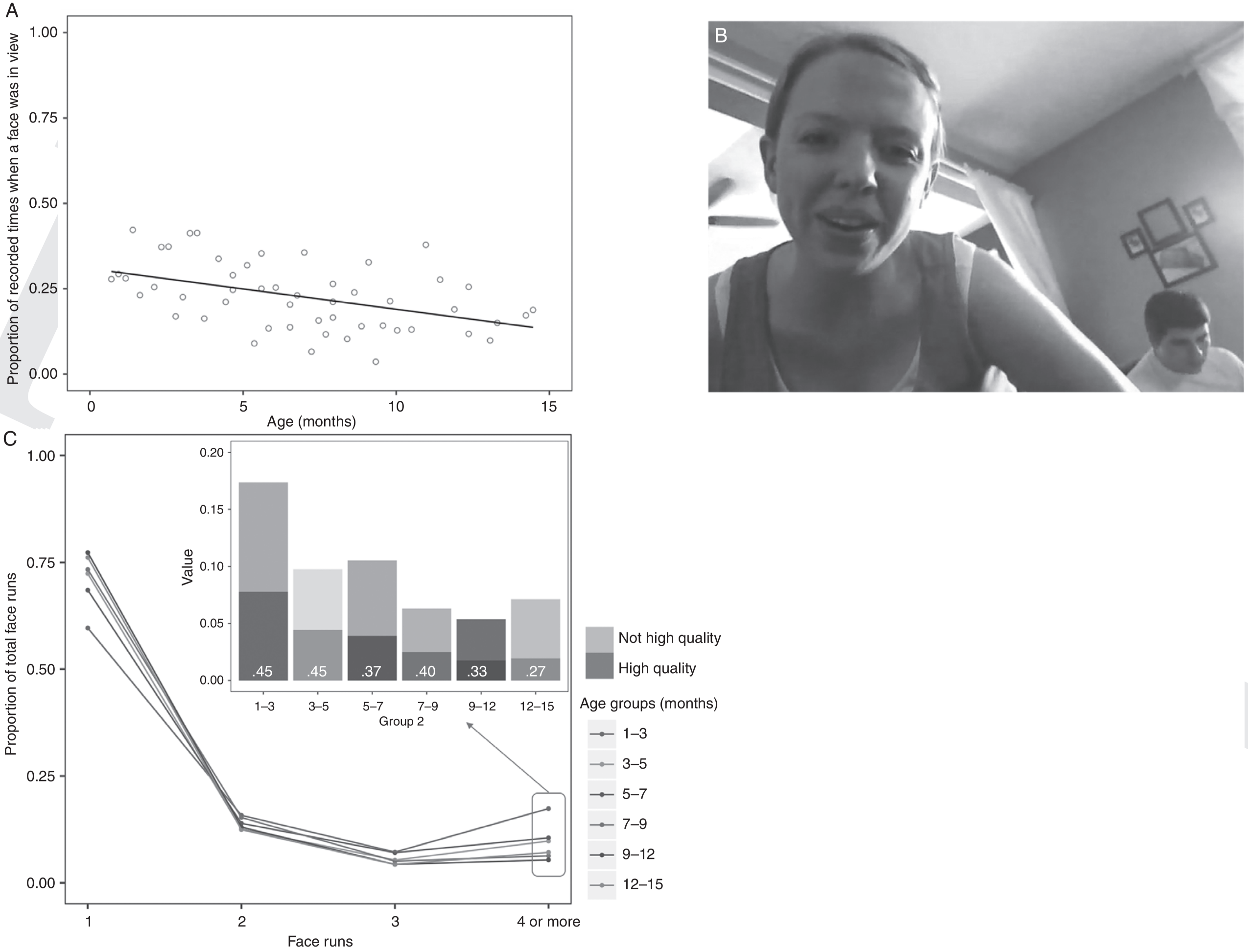

Figure 20.2. *Early face experiences. (A) Face experiences of younger infants are more frequent than those of older infants. (B) Younger infants predominantly experienced faces within 2 feet of the head camera. (C) Faces are temporally more enduring in the visual environments of younger infants.*

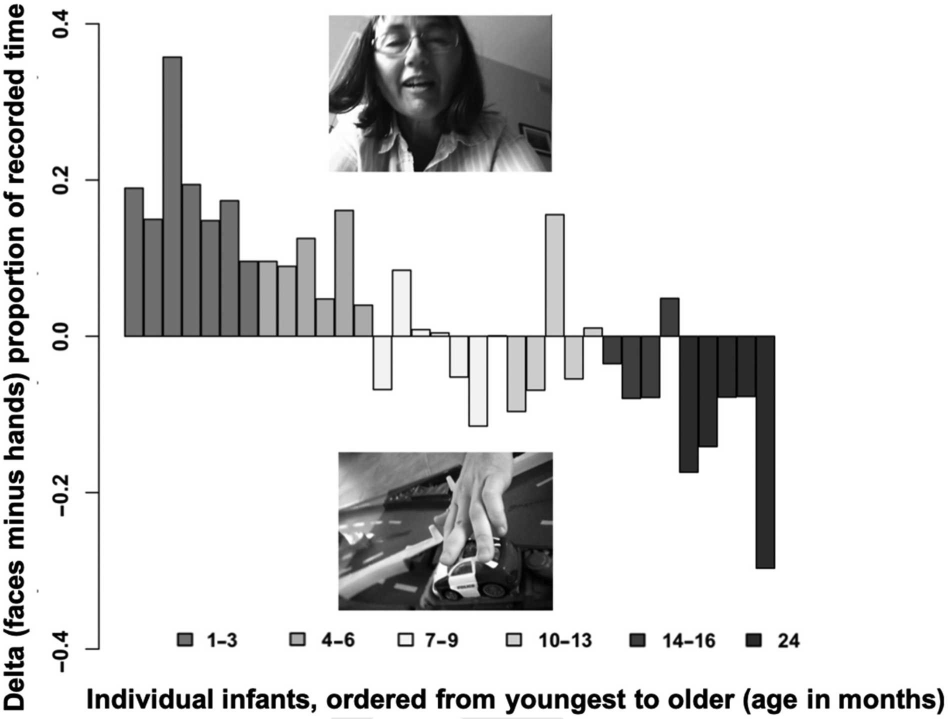

Figure 20.3. *Developmental transition from visual experiences dense in faces to those dense in hands.*

20.3 Developmentally Changing Content

The frequency of faces in infant egocentric images decreases with age, but the frequency of people does not. Other people's bodies are present in about 40% of head-camera images for all infants from birth to 24 months (Jayaraman et al., 2017). As shown in Figure 20.3, the developmental transition is from egocentric images with many faces to egocentric images with hands (Fausey et al., 2016). In 85% of all infant head cameras with hands in view, those hands are acting on objects – showing, giving, attaching, moving. The transition in scene content is thus from faces to hands acting on objects.

Before they are 6 months of age, infants show clear expectations about possible/typical versus impossible/atypical hand actions (Sommerville, Woodward, & Needham, 2005; Woodward, 1998). Older infants are highly sensitive to the causal and semantic structure of manual actions (Sommerville, Upshaw, & Loucks, 2012) and to how points, gestures, and manual actions guide visual attention to objects (e.g., Gogate, Bahrick, & Watson, 2000; Goldin-Meadow & Wagner, 2005; Yu & Smith, 2013). A growing body of findings indicates that infants' own hand actions play a key role in their understanding of other's actions (Brandone, 2015; Krogh-Jesperson & Woodward, 2018). This link is often viewed as interconnected neural representations for interpreting and performing actions and/or for building an understanding of intentions and goals (e.g., Falck-Ytter, Gredebäck, & von Hofsten, 2006; Flanagan & Johansson, 2003; Gerson, Meyer, Hunius, & Bekkering, 2017; Krogh-Jesperson & Woodward, 2018). But how could this work *visually*? From the egocentric view, images of one's own hand actions are very different from others (Bambach, Crandall, & Yu, 2015). One plausible hypothesis is that an infant's ability to perform a particular action invites others to jointly engage with objects (Striano & Reid, 2006) providing visual data from learning to match one's own hand actions to others. As infants' manual abilities become more sophisticated and their own hands appear more densely in view, so may the hands of others performing similar actions. This would provide direct visual evidence with regard to the common intrinsic properties of own versus other hand actions with the same purpose, and the visual basis for systematic expectations about others that are linked to one's own actions. The field's current understanding of these phenomena could benefit from the systematic study of visual experiences of hand actions "in the wild" of everyday life.

The developmental shift from a visual world dense in faces to one dense in hands seems highly relevant to social development. Very young infants follow another's gaze in highly restricted viewing contexts (e.g., Farroni, Johnson, Brockbank, & Simion, 2000; Farroni, Pividori, Simion, Massaccesi, & Johnson, 2004), but 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). The *spatial* complexity explodes as infants become more physically active and transition from interactions dominated by face-to-face play to interactions dominated by shared engagement with objects (see Striano & Reid, 2006). One study using simultaneous head-mounted eye trackers worn by toddlers and parents (Yu & Smith, 2013) found that 1-year-old infants coordinated their own gaze with that of the parent, not by following parent eye gaze, but by fixating on parent hand movements to objects (to which parent eye gaze was also dynamically coordinated), perhaps reflecting a shift from faces to hands in social interactions and social competency.

All of this is the tip of a very large and likely important factor in development. Because of the marked changes in infant sensory-motor abilities, cognitive abilities, and behavior, there are likely major changes from low-level to high-level properties of visual experiences. Moreover, the hierarchical structure of the neural visual pathways (Figure 20.4) suggest that it is highly unlikely that a domain-specific experience will have strictly domain-specific consequences (e.g., Hochstein & Ahissar, 2002; Yamins & DiCarlo, 2016). Instead, early experiences of all kinds will tune processes at early layers that underlie all visual judgments. In this way, learning about faces and about *nonface object categories* will both depend on the precision, tuning, and activation patterns

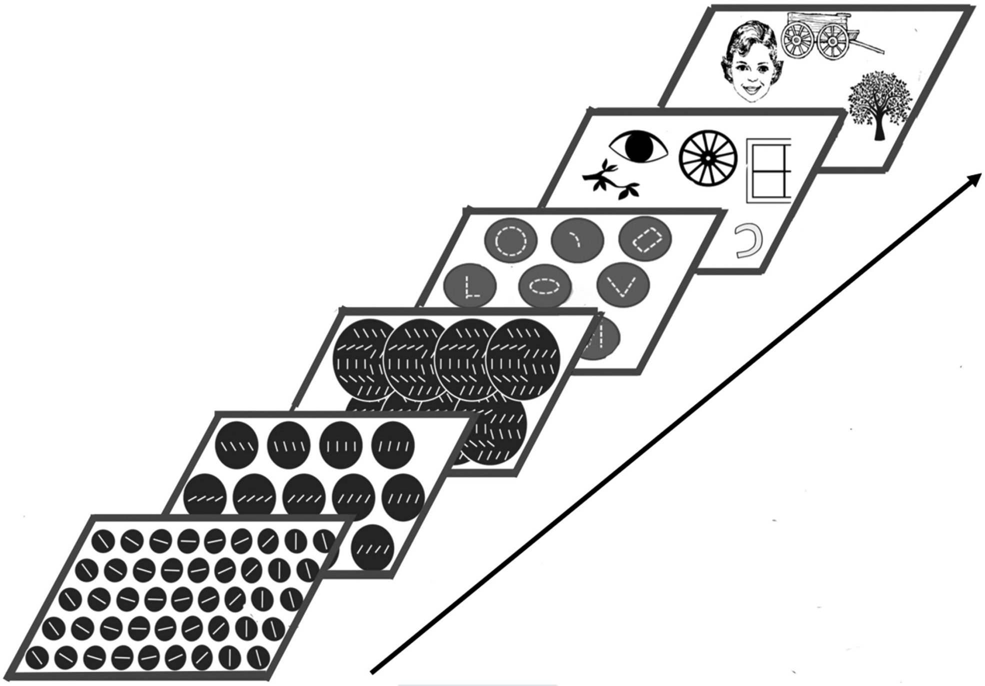

Figure 20.4. *A schematic of the cascade of feature abstraction in the human visual cortex, beginning with the spatially localized extraction of simple features in the lower layers and category-specific features and object categories in the upper layers.*

of the same lower layers. Simple visual discriminations at lower layers can have far-reaching generality across higher-level visual processes, and top-down connections between layers also drive lower changes that can affect processing across multiple domains (e.g., Ahissar & Hochstein, 1997; Cadieu et al., 2014).

Although Maurer et al. (2007) used the term "sleeper effects" to refer to *deficits* in experience, the role of early visual experience on later emerging achievements may be conceptualized both negatively and positively. Regularities in an individual's early experiences may set up potentially hidden competencies that are critical to and play out in later learning. For example, the precision of visual discrimination of dot arrays predicts later mathematics achievement (Halberda, Mazzocco, & Feigenson, 2008) and the shape bias in toddlers predicts the ability to learn letters (Augustine, Jones, Smith, & Longfield, 2015) and objects that comprise a category (Smith, 2005, 2009). The early layer representations formed in one task will be reused and in principle can influence – both negatively and positively – the solutions that are found in learning other tasks. The computational value of ordered training sets for such hierarchically layered learning systems is not yet well understood but is critical to a complete understanding of visual development.

20.4 Everyday Visual Tasks

Experiments, computational models, and machine learning see the problem of learning as one of classification and discrimination: Is this the face of person A or person B? Is this a cup, or a bowl, or a vase? Both models and experimental data suggest that the best training for this kind of learning consists of a uniform frequency distribution of many examples from many categories (e.g., Foody, McCulloch, & Yates, 1995; Perry, Samuelson Malloy, & Schiffer, 2010). However, an extensive literature (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Montag, Jones, & Smith, 2018; Salakhutinov, Torralba, & Tenenbaum, 2011) shows that everyday learning environments are characterized by *skewed* frequency distributions in which a very few types (the mother's face, the sippy cups) are very frequent, but most types (all the different faces encountered at a grocery store, other cups in the cupboard) are encountered quite rarely. For example, in the visual world of infants dense with faces, just three individual faces account for over 95% of all face images (Jayaraman et al., 2015). This number declines only slightly to 80% for 1-year-old infants. From this highly selective and nonuniform sampling of faces, infants learn to recognize and discriminate faces in general. How does this work?

Distributions shaped like the one illustrated in Figure 20.5 emerge because our experiences – tied to our physical bodies and physical locations – are constrained by space and time: We do not discretely jump between contexts like randomized slides in an experiment but transition continuously between visual moments in a smoothly changing physical location. Given these physical constraints of space and time, the faces of family members and caregivers will account for most face experiences, the infant's favorite sippy cup will account for most cup experiences; the family dog will account for most dog experiences. Training expertise with a single object may well be the optimal start for human category learning. In an elegant series of experiments, Oakes and colleagues (Hurley & Oakes, 2015; Kovach-Lesh, Horst & Oakes, 2008; Kovack-Lesh, Oakes & McMurray, 2012) tested the ability of infants raised with and without a dog (or other pets) to recognize and discriminate different animal categories. The extensive visual experiences of infants with their own family dog was associated with advanced recognition of dogs in particular and advanced discrimination and categorization of *other* animals in general. The extensive experience with the proximal images of the dog under various conditions – up close, profile, in clutter, while running, in poor light, from seeing a shadow of the tail – has enabled recognition of the distal object (dog). Extremely skewed distributions with many different visual experiences of one or a very few instances may be optimal for solving this core problem.

Infants learn object names and object categories before they can actually say those names, before their first birthday, (Bergelson & Swingley, 2012). How they do so is not clear as the everyday world is visually cluttered and viewing conditions are often far from optimal. The statistics of egocentric images

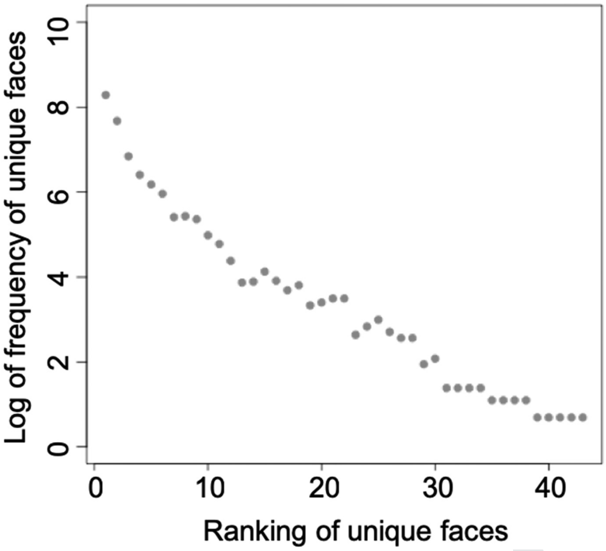

Figure 20.5. *The (log) frequency distribution of individual people's faces in infant head-camera images, ranked by appearance.*

of 8- to 10-month-old infants may hold the solution. Infants this age can sit steadily, even crawl, and play with objects, but their manual skills are quite limited compared to older infants. Neither faces nor hands acting on objects are statistically dominant in their images; instead, images contain a mixture of the various body parts of nearby people (Fausey et al., 2016; Jayaraman et al., 2017). Overall, these scenes are highly cluttered (Figure 20.6), containing many objects (Clerkin et al., 2017). However, a small number of object categories are pervasively present, with most object types being quite rare. Across similar contexts, such as mealtimes, very few objects (spoons, bottles) are repetitively present and many more objects (jugs, salad tongs, ketchup bottles) were present in only a very few scenes (Clerkin et al., 2017). The pervasive repetition of instances from a select set of categories could help infants to find and attend to those objects – even in clutter – and provide a foundation for linking those objects to their names.

20.5 Relevance to Developmental Neuroscience

Human perception, behavior, and cognition is the product of dynamic neural activity; that activity changes the microarchitecture at the local neural level and changes the changing brain connectivity as "both" cause and consequence of developmental changes in the brain and in behavior (Edelman, 1987; James, 2010; Lloyd-Fox, Wu, Richards, Elwell, & Johnson, 2013; Menon, 2013; Power, Fair, Schlaggar, & Petersen, 2010; Sporns & Edelman, 1993). An

Figure 20.6. *Cluttered head-camera image of a 9-month-old infant, containing many different objects and body parts.*

expansion of studies examining age-related changes in these networks (James, 2010; Menon, 2013; Power et al., 2010) and their consequences for local change has set the stage for new research approaches directed to understanding the processes through which one dynamic system, that is the brain at one age, turns into the dynamic system that is the brain at a later age. Because the neural activity at any moment in the brain is a product of its intrinsic properties (themselves a product of a developmental history), the bodily behavior generated by the brain, and the sensory input, studying brain development cannot be separated from the study of behavior and experience. For example, when 4- and 5-year-old children are first learning to write the letters of the alphabet, they connect in new ways the neural processes underlying motor behavior, motor planning, and the visual system. The manual act of writing letters – which recruits this functional neural network – appears essential not just to forming this network but also to tuning the specialized regions for letter recognition within the ventral pathway in the visual cortex (James, 2010; James & Atwood, 2009). The path to specialized and localized visual processing of letters involves the functional connectivity of that system to motor planning and to behavior (James, 2010; James & Atwood, 2009). These advances should move us beyond timeworn debates about innate versus experience-dependent (and experience malleable) processes. This debate has, for example, been a core driving issue in the development of face perception. Are the specialized visual processes for human face perception evident in adults principally determined by genetic processes or are they the product of massive visual experience? The argument on the innate side points to the biological importance of faces and the existence of specialized circuitry for face processing (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Goren, Sarty, & Wu, 1975; Johnson et al., 1991; Kanwisher, 2000; McKone, Crookes, Jeffery, & Dilks, 2012). A specific distributed network of brain areas is preferentially involved in face processing, including the "fusiform face area" within the inferotemporal cortex (Adolphs, 2002; Haxby, Hoffman, & Gobbini, 2002; Iidaka et al., 2001; Kanwisher, McDermott, & Chun, 1997). The development of this specialized region has multiple intrinsic biases supporting it, including the optics of newborn infant eyes that strongly favor inputs with low-level visual properties of up-close frontal views of faces as well as the neediness of young babies and the emotional attachment of parents that make up-close frontal views plentiful. But experience of faces is nonetheless essential. Recent studies of nonhuman primates show that without face experiences, infant monkeys do not develop their species-specific and specialized regions for face processing (Arcaro, Schade, Vincent, Ponce, & Livingstone, 2017), a finding that has similarities to the results of infants born with cataracts. Moreover, the existence of specialized areas for face processing, although perhaps more strongly biased by infant optics and parent behaviors, may emerge nonetheless from the same general principles as specialized visual regions for unnatural object categories such as letters of the alphabet and buildings (Aguirre, Zarahn, & D'esposito, 1998; Cohen & Dehaene, 2004).

Disruptions in early sensory-motor coordination have been implicated as early diagnostic markers along several atypical developmental pathways (D'Souza, D'Souza, & Karmiloff-Smith, 2017; Hinton, et. al., 2013; Iverson, 2010) and may be perturbations in the developmental trajectory with far-reaching consequences, including language learning (D'Souza et al., 2017; Iverson, 2010), social interactions (Leonard & Hill, 2014), attention (Ravizza, Solomon, Ivry, & Carter, 2013), behavioral control, and school achievement (Cameron, Cottone, Murrah, & Grissmer, 2016). Focused behavioral training with respect to these early perturbations provides a principled means for optimizing developmental outcomes of individuals with different developmental disorders. However, doing so requires an understanding of the role of behavior, not just as a phenotype but also as part of the causal pathway in brain development. Considerable and growing research suggests that disparities in children's developmental environments have serious consequences for brain development and for social, emotional, and cognitive behaviors. If we are to fully understand human visual development, and human brain development more generally, we need to situate the developing brain within the developing organism and the extended brain–behavior–experience network (Byrge, Sporns, & Smith, 2014). Human variation emerges as each individual organism travels along its unique developmental path, a path surely started – and influenced throughout the lifetime – by genetic components, but pushed forward by the developing child's own interactions with the world.

20.6 Social and Cultural Contexts

Infants do not develop alone but do so in the company of adults. Infants' egocentric views – and the data on which visual learning depends – is strongly influenced by those adults. Long before infants hold objects on their own, as early as when infants are 3 months of age, parents hold, bring, and show objects to their infants (Clark & Estigarribia, 2011; Libertus & Needham, 2011; Zukow, 1990). One recent study (Burling & Yoshida, 2018) used head-mounted eye trackers to directly measure visual object information as infants aged 5 to 24 months played with objects with a parent. The parent's hand actions created the views for younger infants, with the views created by the infant's own hand actions increasing with the age of the infant. The analyses showed similarities between parent-generated views for younger infants and the views that were self-generated by the older infants. Held objects – and the targets of infant gaze regardless of who was holding – were large in the image, close to the perceiver, and segregated from the background. Other studies show that parent behavior linked to the infant's own behavior. For example, parents are more likely to jointly attend to and talk about objects when infants hold objects or carry the object while walking (Karasik, Tamis-LeMonda, & Adolph, 2011; Yu & Smith, 2017).

What parents do in social contexts with their infants, however, depends on their own history, culture, and geography. There are marked cultural differences in the physical and visual properties of environments (e.g., Dolgin, 2015), in parenting practices (e.g., Lansford et al., 2018; Prevoo & Tamis-LeMonda, 2017), and in motor development (e.g., Karasik, Tamis-LeMonda, Adolph, & Bornstein, 2015). These differences could, in principle, affect visual development. One relevant line of research concerns scene processing in Western adults (residing in North America and Europe) and Eastern adults (residing in China, Japan, and Korea), cultures that in wealth, education, the character of work and family have many similarities. Nonetheless, a large number of studies have documented fundamental differences in visual scene processing by a variety of measures, including recognition measures (Ishii, Tsukasaki, & Kitayama, 2009; Masuda & Nisbett, 2001, 2006), eye tracking (Chua, Boland, & Nisbett, 2005; Kelly, Miellet, & Caldara, 2010; Masuda et al., 2008), and brain imaging (Goh et al., 2013; Han & Northoff, 2008; Hedden, Ketay, Aron, Markus, & Gabrieli, 2008; Masuda, Russell, Chen, Hioki, & Caplan, 2014). In aggregate, the findings suggest that Western perceivers are more selective, more focused on local elements in scenes, and less affected by visual context than Eastern perceivers. In contrast, Eastern perceivers are more holistic and more sensitive to the relational structure among elements in a scene (e.g., Chua et al., 2005; Hedden et al., 2008; Kitayama, Duffy, Kawamura, & Larsen, 2003; Masuda et al., 2008; Masuda & Nisbett, 2001, 2006; Miyamoto, Yoshikawa, & Kitayama, 2011; Nisbett & Masuda, 2003; Nisbett & Miyamoto, 2005; Nisbett, Peng, Choi, & Norenzayan, 2001). These cultural differences in visual processing have also been reported in children (Duffy, Toriyama, Itakura, & Kitayama, 2009; Imada, Carlson, & Itakura, 2013; Moriguchi, Evans, Hiraki, Itakura, & Lee, 2012; Senzaki, Masuda, & Nand, 2014), including those as young as 3 years of age (Kuwabara & Smith, 2012, 2016; Kuwabara, Son, & Smith, 2011). Much like object processing, adults from Eastern societies tend to process faces more holistically by centrally fixating around the nose when learning and recognizing faces, while their Western counterparts are more analytical and fixate around salient parts of the face like the eyes and mouth (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Kelly, et al., 2011). Understanding the nature and origins of these cultural differences – and thus the adaptive nature of human visual prowess – requires the direct study of *visual* environments across different cultural contexts.

To fully understand human vision, we need to understand the full variety of infant experiences, not just those in typical Western societies that make up for only an eighth of the world population (Henrich, Heine, & Norenzayan, 2010a, 2010b).

20.7 Policy Implications

Descartes in 1628 gave us the basic approach to science. In studying any phenomenon, reduce it to its essential components and dissect away everything else. This analytic approach is motivated by the belief that complicated systems are most fruitfully investigated, and causes most precisely determined, at the lowest possible level. The goal is to find components that are simple enough to fully analyze, explain, and control. The spectacular success of this methodology in modern biology is undeniable. It has led to unprecedented knowledge of the molecular components of life. It has not led to a corresponding understanding of how large collections of such components operate *as systems* (Bechtel & Richardson, 1993; Keller, 2007; Kitson et al., 2018). This failure is evident in the challenge of effective translation of basic science. Mechanisms clearly and cleanly specified in laboratories often fail to work in real life (Collins, 2011; Lenfant, 2003; Norman, 2010). For developmental psychologists, Descartes' tenet has meant moving the phenomena out of everyday life and into the laboratory, using logically clean and well-controlled experiments that manipulate hypothesized variables to determine individual causes. But as in other biological sciences, our ability to take those findings into the world in all its complexity and variability is limited (Henrich et al., 2010a, 2010b).

The study of "egocentric vision," capturing the visual structure of the learning environment, like the sister efforts in language trying to capture the language input with day-long recordings (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019), is a start to trying to understand real-world environments. These studies are relatively new, and there is much we do not know, we still have a long way to go and still lack the right analytic tools for understanding the properties of the environment at scale. Consider the case of current research on the disparities in children's language learning environments. Researchers are now using wearable audio recorders that can capture all the words a child hears in a day. They use algorithmic analyses that operate on the raw recorded sounds to estimate the number of words in child-directed speech. These studies (as well as earlier ones using more traditional speech sampling and transcription) reveal that the average preschool child hears about 20,000–38,000 total words a day (Fernald & Weisleder, 2015; Hart & Risley, 1995; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; Weisleder & Fernald, 2013). But there is extreme variability, some children hear as few as 2,000 child-directed words a day and some as many as 50,000 words a day (Fernald & Weisleder, 2015; Hart & Risley, 1995; Weisleder & Fernald, 2013). These differences in amount of talk to individual children are strongly predictive of the child's vocabulary size and early school achievement (Dickinson, Golinkoff, & Hirsh-Pasek, 2010; Hart & Risley, 1995; Hoff, 2003; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Rowe, 2012; Walker, Greenwood, Hart, & Carta, 1994) and are also highly associated with the socioeconomic standing of the families (Hoff, 2003; Hurtado, Marchman, & Fernald, 2008; Huttenlocher et al., 2010; Weisleder & Fernald, 2013). By some estimates (Hart & Risley 1995) there is a 30-million-word gap in the cumulative number of words directed to children from poorer versus richer families. As a result, there is now a considerable public health effort directed to increasing parent talk to young children (Leffel & Suskind, 2013; Reese, Sparks, & Leyva, 2010; Roberts & Kaiser, 2011; and public health initiatives such as Providence Talks, First 5 California, and Too Small to Fail, among many others). Talking to children seems a likely positive factor in child development but doubts about the basic idea have been raised (Sperry, Sperry, & Miller, 2018), contested (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2018), and there many open questions (Montag et al., 2018; Romeo et al., 2018). What exactly is the pathway to learning? Is it really just more words? Does the context of talk matter? Or, perhaps, is the amount of parent talk correlated with some other environmental factor that is more critical? To answer these questions, we need to know how to measure and compare developmental environments beyond just counting words. This is a much more complex problem than it might seem because the scale of everyday experience is huge; for example, the average child who hears 20,000 child-directed words a day hears over 7 million words in a year (Bergelson et al., 2019; Hart & Risley, 1995; Shneidman et al., 2013). The problem is also complicated by the fact that the frequency distribution of words in produced language (like the frequency distribution of objects in the world) is not normal and thus the usual assumptions about sampling from normal distributions and statistical inference do not apply (Clerkin et al., 2017; Montag et al., 2018; Salakhutinov et al., 2011). Usual assumptions about attention, memory, and learning pathways may also not apply because what we know about these processes is based on small-scale experiments and uniform distributions of learning items (Clerkin et al., 2017). These issues – in the study of language, of visual environments, of social environments – are becoming urgent as both the data indicating consequential disparities in environments advances and as new methods emerge to capture environmental regularities at scale advance (Gilkerson & Richards, 2008; Roy et al., 2006; VanDam et al., 2016) as do large data sets of many children in common contexts like Many Babies (Frank et al., 2017), Homeview (Jayaraman & Smith, 2017), and Play & Learning Across a Year (Adolph, Tamis-LeMonda, Gilmore, & Soska, 2018).

20.8 Conclusion

Development is a personal journey, albeit one that is taken in the supportive company of others. But it is the personal vantage point of the learner, selective and localized, and their own path that constitute the tasks, behaviors, and experiences that determine and build the competencies of the individual. Although we most definitely need fine-grained laboratory experiments of the development of behavior, cognition, and neural processes, we also need to study the developmental structure of environments and ideally do so at the level of the individual. Here we considered visual environments and visual cognition. But more generally, the developmental study of the environments may also yield a deeper understanding of individual differences in early cognitive development. The source of differences – and interventions to support healthy development in all children – may emerge in part from the developmental structure of the data for learning, data that are determined by the immediate surroundings of the learner and their developing behaviors in those surroundings.

References

Adolph, K., Tamis-LeMonda, C., Gilmore, R. O., & Soska, K. (2018). Play & learning across a year (PLAY) project summit (2018-06-29 Philadelphia). *Databrary*. Retrieved from http://doi.org/10.17910/B7.724.

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron, 21*, 373–383.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401.

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience*, *20*(10), 1404.

Augustine, E., Jones, S. S., Smith, L. B., & Longfield, E. (2015). Relations among early object recognition skills: Objects and letters. *Journal of Cognition and Development*, *16*(2), 221–235.

Bambach, S., Crandall, D. J., & Yu, C. (2015). Viewpoint integration for hand-based recognition of social interactions from a first-person view. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, November, 351–354.

Bechtel, W., & Richardson, R. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Bertenthal, B. I., Campos, J. J., & Barrett, K. C. (1984). Self-produced locomotion. In B. I. Bertenthal, J. J. Campos, & K. C. Barrett (Eds.), *Continuities and discontinuities in development* (pp. 175–210). New York, NY: Springer.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PloS One*, *3*(8), e3022.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, *51*(13), 1588–1609.

Brandone, A. C. (2015). Infants' social and motor experience and the emerging understanding of intentional actions. *Developmental Psychology*, *51*(4), 512.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108.

Burling, J. M., & Yoshida, H. (2018). Visual constancies amidst changes in handled objects for 5- to 24-month-old infants. *Child Development*, *90*(2), 452–461.

Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*(4), 1005–1021.

Bushnell, W. (2003). Newborn face recognition. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood* (pp. 41–53) New York, NY: Nova Science.

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain–body–behavior networks. *Trends in Cognitive Sciences*, *18*(8), 395–403.

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate

IT cortex for core visual object recognition. *PLoS Computational Biology*, *10*(12), e1003963.

Cameron, C. E., Cottone, E. A., Murrah, W. M., & Grissmer, D. W. (2016). How are motor skills linked to children's school performance and academic achievement? *Child Development Perspectives*, *10*(2), 93–98.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, *1*(2–3), 253–274.

Castelhano, M. S., & Witherspoon, R. L. (2016). How you use it matters: Object function guides attention during visual search in scenes. *Psychological Science*, *27*(5), 606–621.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, *102*(35), 12629–12633.

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture*, *11*(1), 1–23.

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B*, *372*(1711), 20160055.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *Neuroimage*, 22(1), 466–476.

Collins, F. S. (2011). Reengineering translational science: The time is right. *Science Translational Medicine*, *3*(90). doi: 10.1126/scitranslmed.3002747.

D'Souza, D. E., D'Souza, H., & Karmiloff-Smith, A. (2017). Precursors to language development in typically and atypically developing infants and toddlers: The importance of embracing complexity. *Journal of Child Language*, *44*(3), 591–627.

DiCarlo, J. J., & Cox, D. D. J. (2007). Untangling invariant object recognition. *Trends in Cognitive Science*, *11*(8), 333–341.

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*(4), 305–310.

Dobson, V., Teller, D. Y., & Belgum, J. J. (1978). Visual acuity in human infants assessed with stationary stripes and phase-alternated checkerboards. *Vision Research*, *18*(9), 1233–1238.

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*(3), 296–312.

Dolgin, E. (2015). The myopia boom. *Nature*, *519*(7543), 276.

Duffy, S., Toriyama, R., Itakura, S., & Kitayama, S. (2009). Development of cultural strategies of attention in North American and Japanese children. *Journal of Experimental Child Psychology*, *102*, 351–359.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York, NY: Basic Books.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878.

Fantz, R. L. (1963). Pattern vision in newborn infants. *Science*, *140*(3564), 296–297.

Farroni, T., Johnson, M. H., Brockbank, M., & Simion, F. (2000). Infants' use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition*, *7*(6), 705–718.

Farroni, T., Pividori, D., Simion, F., Massaccesi, S., & Johnson, M. H. (2004). Eye gaze cueing of attention in newborns. *Infancy*, *5*(1), 39–60.

Fathi, A., Ren, X., & Rehg, J. M. (2011). *Learning to recognize objects in egocentric activities*. Paper presented at the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107.

Fernald, A., & Weisleder, A. (2015). Twenty years after "meaningful differences," it's time to reframe the "deficit" debate about the importance of children's early language experience. *Human Development*, *58*(1), 1.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769.

Foody, G. M., McCulloch, M. B., & Yates, W. B. (1995). The effect of training set size and composition on artificial neural network classification. *International Journal of Remote Sensing*, *16*, 1707–1723

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931.

Franchak, J. M., & Adolph, K. E. J. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*(24), 2766–2774.

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science*, *21*(4), e12626.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … Lew-Williams, C. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring the face recognition mechanisms. *Vision Research*, *37*(12), 1673–1682.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192.

Gerson, S. A., Meyer, M., Hunnius, S., & Bekkering, H. (2017). Unravelling the contributions of motor experience and conceptual knowledge in action perception: A training study. *Scientific Reports*, *7*, 46761.

Gilkerson, J., & Richards, J. A. (2008). *The LENA natural language study*. Boulder, CO: LENA Foundation.

Gilmore, R. O., Baker, T. J., & Grobman, K. (2004). Stability in young infants' discrimination of optic flow. *Developmental Psychology*, *40*(2), 259.

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*(4), 878–894.

Goh, J. O. S., Hebrank, A. C., Sutton, B. P., Chee, M. W. L., Sim, S. K. Y., & Park, D. C. (2013). Culture-related differences in default network during visuo-spatial judgments. *Social Cognitive and Affective Neuroscience*, *8*, 134–142.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5), 234–241.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2018). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, *90*(3), 985–992.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549.

Hadad, B. S., Maurer, D., & Lewis, T. L. (2011). Long trajectory for the development of sensitivity to global and biological motion. *Developmental Science*, *14*(6), 1330–1339.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665.

Han, S., & Northoff, G. (2008). Reading direction and culture. *Nature Reviews Neuroscience*, *9*(12), 965.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, *4*(6), 223–233.

(2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*(1), 59–67.

Hedden, T., Ketay, S., Aron, A., Markus, H. R., & Gabrieli, J. D. E. (2008). Cultural influences on neural substrates of attentional control. *Psychological Science*, *19*, 12–17.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*(2–3), 111–135.

(2010b). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Higgins, C. I., Campos, J. J., & Kermoian, R. (1996). Effect of self-produced locomotion on infant postural compensation to optic flow. *Developmental Psychology*, *32*(5), 836.

Hinton, R., Budimirovic, D. B., Marschik, P. B., Talisa, V. B., Einspieler, C., Gipson, T., & Johnston, M. V. (2013). Parental reports on early language and motor milestones in fragile X syndrome with and without autism spectrum disorders. *Developmental Neurorehabilitation*, *16*(1), 58–66.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hurley, K. B., & Oakes, L. M. (2015). Experience and distribution of attention: Pet exposure and infants' scanning of animal images. *Journal of Cognition and Development*, *16*(1), 11–30.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365.

Iidaka, T., Omori, M., Murata, T., Kosaka, H., Yonekura, Y., Okada, T., & Sadato, N. (2001). Neural interaction of the amygdala with the prefrontal and temporal cortices in the processing of facial expressions as revealed by fMRI. *Journal of Cognitive Neuroscience*, *13*(8), 1035–1047.

Im, H. Y., Park, W. J., & Chong, S. C. (2015). Ensemble statistics as units of selection. *Journal of Cognitive Psychology*, *27*(1), 114–127.

Imada, T., Carlson, S. M., & Itakura, S. (2013). East–West cultural differences in context sensitivity are evident in early childhood. *Developmental Science*, *16*, 198–208.

Ishii, K., Tsukasaki, T., & Kitayama, S. (2009). Culture and visual perception: Does perceptual inference depend on culture? *Japanese Psychological Research*, *51*(2), 103–109.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261.

James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental Science*, *13*(2), 279–288.

James, K. H., & Atwood, T. P. (2009). The role of sensorimotor learning in the perception of letter-like forms: Tracking the causes of neural specialization for letters. *Cognitive Neuropsychology*, *26*(1), 91–110.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PloS one*, *10*(5), e0123780.

(2017). Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology*, *53*(1), 38.

Jayaraman, S., & Smith, L. B. (2017). *The homeview project*. Retrieved from www.iub.edu/~cogdev/homeview.html.

(2018). Faces in early visual environments are persistent not just frequent. *Vision Research*, *157*, 213–221.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19.

Johnson M. H., & Morton J. (1991) *Biology and cognitive development: The case of face recognition*. Oxford: Blackwell.

Kanwisher, N. J. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759.

Kanwisher, N. J., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development*, *82*(4), 1199–1209.

Karasik, L. B., Tamis-LeMonda, C. S., Adolph, K. E., & Bornstein, M. H. (2015). Places and postures: A cross-cultural comparison of sitting in 5-month-olds. *Journal of Cross-Cultural Psychology*, *46*(8), 1023–1038.

Keller, E. F. (2007). The disappearance of function from "self-organizing systems." In F. Boogerd, F. Bruggeman, J. H. Hofmeyre, & H. V. Westerhoff (Eds.), *Systems biology* (pp. 303–317). Amsterdam: Elsevier.

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*(5), 1176–1184.

Kelly, D. J., Miellet, S., & Caldara, R. (2010). Culture shapes eye movements for visually homogeneous objects. *Frontiers in Psychology*, *1*, 6.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science, 14*, 201–206.

Kitson, A., Brook, A., Harvey, G., Jordan, Z., Marshall, R., O’Shea, R., & Wilson, D. (2018). Using complexity and network concepts to inform healthcare knowledge translation. *International Journal of Health Policy and Management*, *7*(3), 231.

Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorization. *Infancy*, *13*(4), 285–307.

Kovack-Lesh, K. A., Oakes, L. M., & McMurray, B. (2012). Contributions of attentional style and previous experience to 4-month-old infants’ categorization. *Infancy*, *17*(3), 324–338.

Krogh-Jespersen, S., & Woodward, A. L. (2018). Reaching the goal: Active experience facilitates 8-month-old infants’ prospective analysis of goal-based actions. *Journal of Experimental Child Psychology*, *171*, 31–45.

Kuwabara, M., & Smith, L. B. (2016) Cultural differences in visual object recognition in 3-year-old children. *Journal of Experimental Child Psychology*, *147*, 22–38.

Kuwabara, M., & Smith, L. B. (2012). Cross-cultural differences in cognitive development: Attention to relations and objects. *Journal of Experimental Child Psychology*, *113*, 20–35.

Kuwabara, M., Son, J. Y., & Smith, L. B. (2011). Attention to context: U.S. and Japanese children’s emotional judgments. *Journal of Cognition and Development*, *12*, 502–517.

Lansford, J. E., Godwin, J., Al-Hassan, S. M., Bacchini, D., Bornstein, M. H., Chang, L., & Malone, P. S. (2018). Longitudinal associations between parenting and youth adjustment in twelve cultural groups: Cultural normativeness of parenting as a moderator. *Developmental Psychology*, *54*(2), 362.

Leffel, K., & Suskind, D. (2013, November). Parent-directed approaches to enrich the early language environments of children living in poverty. *Seminars in Speech and Language*, *34*(4), 267–278

Lenfant, C. (2003). Clinical research to clinical practice: Lost in translation? *New England Journal of Medicine*, *349*(9), 868–874.

Leonard, H. C., & Hill, E. L. (2014). The impact of motor development on typical and atypical social cognition and language: A systematic review. *Child and Adolescent Mental Health*, *19*(3), 163–170.

Libertus, K., & Needham, A. (2011). Reaching experience increases face preference in 3-month-old infants. *Developmental Science*, *14*, 1355–1364.

Lloyd-Fox, S., Wu, R., Richards, J. E., Elwell, C. E., & Johnson, M. H. (2013). Cortical activation to action perception is associated with action production abilities in young infants. *Cerebral Cortex*, *25*(2), 289–297.

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*(9), 1443–1457.

Macchi, C. V., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science*, *15*(6), 379–383.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & van de Veerdonk, E. (2008). Placing the face in context: cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*(3), 365.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*(5), 922.

(2006). Culture and change blindness. *Cognitive Science*, *30*(2), 381–399.

Masuda, T., Russell, M. J., Chen, Y. Y., Hioki, K., & Caplan, J. B. (2014). N400 incongruity effect in an episodic memory task reveals different strategies for handling irrelevant contextual information for Japanese than European Canadian. *Cognitive Neuroscience*, *5*, 17–25.

Maurer, D. J. (2017). Critical periods re-examined: Evidence from children treated for dense cataracts. *Cognitive Development*, *42*, 27–36.

Maurer, D. J., & Lewis, T. L. (2001a). Visual acuity: The role of visual input in inducing postnatal change. *Clinical Neuroscience Research*, *1*(4), 239–247.

Maurer, D., & Lewis, T. (2001b). Visual acuity and spatial contrast sensitivity: Normal development and underlying mechanisms. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 237–250). Cambridge, MA: MIT Press.

Maurer, D., Mondloch, C. J., & Lewis, T. L. (2007). Sleeper effects. *Developmental Science*, *10*(1), 40–47.

McKone, E., Crookes, K., Jeffery, L., & Dilks, D. D. (2012). A critical review of the development of face recognition: Experience is less important than previously believed. *Cognitive Neuropsychology*, *29*(1–2), 174–212.

Menon, V. (2013). Developmental pathways to functional brain networks: Emerging principles. *Trends in Cognitive Sciences*, *17*(12), 627–640.

Miyamoto, Y., Yoshikawa, S., & Kitayama, S. (2011). Feature and configuration in face processing: Japanese are more configural than Americans. *Cognitive Science*, *35*, 563–574.

Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive Science*, *42*, 375–412.

Moriguchi, Y., Evans, A. D., Hiraki, K., Itakura, S., & Lee, K. (2012). Cultural differences in the development of cognitive shifting: East–West comparison. *Journal of Experimental Child Psychology*, *111*, 156–163.

Moulson, M. C., Westerlund, A., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2009). The effects of early experience on face recognition: An event-related potential study of institutionalized children in Romania. *Child Development*, *80*(4), 1039–1056.

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences*, *100*(19), 11163–11170.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*, 467–473.

Nisbett, R. E., Peng, K. P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*, 291–310.

Nishimura, M., Scherf, S., & Behrmann, M. (2009). Development of object recognition in humans. *F1000 Biology Reports*, 1, 56.

Norman, D. A. (2010). The research-practice gap: The need for translational developers. *Interactions*, *17*(4), 9–12.

Oakes, L. M. (2017). Plasticity may change inputs as well as processes, structures, and responses. *Cognitive Development*, *42*, 4–14.

Oruç, İ., & Barton, J. (2010). Critical frequencies in the perception of letters, faces, and novel shapes: Evidence for limited scale invariance for faces. *Journal of Vision*, *10*(12), 20-20.

Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the links among face processing, language processing, and narrowing during development. *Child Development Perspectives*, *8*(2), 65–70.

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902.

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The development of human functional brain networks. *Neuron*, *67*(5), 735–748.

Prevoo, M. J., & Tamis-LeMonda, C. S. (2017). Parenting and globalization in Western countries: Explaining differences in parent–child interactions. *Current Opinion in Psychology*, *15*, 33–39.

Ravizza, S. M., Solomon, M., Ivry, R. B., & Carter, C. S. (2013). Restricted and repetitive behaviors in autism spectrum disorders: The relationship of attention and motor deficits. *Development and Psychopathology*, *25*(3), 773–784.

Reese, E., Sparks, A., & Leyva, D. (2010). A review of parent interventions for preschool children's language and emergent literacy. *Journal of Early Childhood Literacy*, *10*(1), 97–117.

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology*, *20*(3), 180–199.

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., … Levit, M. (2006, September). The human speechome project. In C. Nehaniv, P. Vogt, Y. Sugita, & E. Tuci (Eds.), *Symbol grounding and beyond: Third International Workshop on Emergence and Evolution of Linguistic Communication* (pp. 192–196). Berlin: Springer.

Salakhutdinov, R., Torralba, A., & Tenenbaum, J. (2011). Learning to share visual appearance for multiclass object detection. *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition,* 1481–1488. https://doi.org/10.1109/CVPR.2011.5995720

Scherf, K. S., & Scott, L. S. (2012). Connecting developmental trajectories: Biases in face processing from infancy to adulthood. *Developmental Psycholobiology*, *54*(6), 643–663.

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*(4), 197–201.

Senzaki, S., Masuda, T., & Nand, K. (2014). Holistic versus analytic expressions in artworks: Cross-cultural differences and similarities in drawings and collages by Canadian and Japanese school-aged children. *Journal of Cross-Cultural Psychology*, *45*, 1297–1316.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*, 672–686.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, *13*(2), 144–149.

Smith, L. B. (2005). Action alters shape categories. *Cognitive Science*, *29*(4), 665–679.

(2009). From fragments to geometric shape: Changes in visual object recognition between 18 and 24 months. *Current Directions in Psychological Science*, *18*(5), 290–294.

Smith, L. B., Yu, C., & Pereira, A. F. J. D. S. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17.

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. *Journal of Cognition and Development*, *16*(3), 407–419.

Sommerville, J. A., Upshaw, M. B., & Loucks, J. (2012). The nature of goal-directed action representations in infancy. *Advances in Child Development and Behavior, 43*, 351–387.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*, B1–B11.

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*(1), 129.

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2018). Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child Development*, 90(4), 1303–1318.

Sporns, O., & Edelman, G. M. (1993). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Development*, *64*(4), 960–981.

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences*, *10*(10), 471–476.

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, *56*(2), 249–261.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., de Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language*, *37*(2), 128–142.

Vida, M. D., & Maurer, D. (2012). Gradual improvement in fine-grained sensitivity to triadic gaze after 6 years of age. *Journal of Experimental Child Psychology*, *111*(2), 299–318.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, *65*(2), 606–621.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84.

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

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356.

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

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye–hand coordination. *PloS One*, *8*(11), e79659.

(2017). From infant hands to parent eyes: Hand–eye coordination predicts joint attention. *Child Development*, *88*(6), 2060–2078.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966.

Zukow, P. G. (1990). Socio-perceptual bases for the emergence of language: An alternative to innatist approaches. *Developmental Psychobiology*, *23*, 705–726. https://doi.org/10.1002/dev.420230711