ARTICLE IN PRESS

Vision Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Faces in early visual environments are persistent not just frequent

Swapnaa Jayaraman*, Linda B. Smith

Indiana University, 1101 E. 10th st., Bloomington, IN 47404, United States

ARTICLE INFO

Number of reviewers = 2

Keywords:
Face
Infant
Development
Duration
Environment
Experience

ABSTRACT

The regularities in very young infants' visual worlds likely have out-sized effects on the development of the visual system because they comprise the first-in experience that tunes, maintains, and specifies the neural substrate from low-level to higher-level representations and therefore constitute the starting point for all other visual learning. Recent evidence from studies using head cameras suggests that the frequency of faces available in early infant visual environments declines over the first year and a half of life. The primary question for the present paper concerns the temporal structure of face experiences: Is frequency the key exposure dimension distinguishing younger and older infants' face experiences, or is it the duration for which faces remain in view? Our corpus of head-camera images collected as infants went about their daily activities consisted of over a million individually coded frames sampled at 0.2 Hz from 232 h of infant-perspective scenes, recorded from 51 infants aged 1 month to 15 months. The major finding from this corpus is that very young infants (1–3 months) not only have more frequent face experiences but also more temporally persistent ones. The repetitions of the same very few face identities presenting up-close and frontal views are exaggerated in more persistent runs of the same face, and these persistent runs are more frequent for the youngest infants. The implications of early experiences consisting of extended repeated exposures of up-close frontal views for visual learning are discussed.

1. Introduction

The first three months of life appear especially important for the development of human face processing. At birth infants show a bias to look at high-contrast low-spatial frequency face-like patterns (Fantz, 1963; Johnson, Dziurawiec, Ellis, & Morton, 1991; Macchi, Turati, & Simion, 2004). This bias may be related to the ocular structure of the eye in early infancy which limits the infant's abilities to coordinate the two eyes and brings objects into focus (Dobson, Teller, & Belgum, 1978; Maurer & Lewis, 2001a, 2001b; Oruc & Barton, 2010). Despite these limitations, or perhaps in part because of them, infants during these first 3 months of post-natal life come to preferentially look at, recognize and discriminate faces that are similar to their caretakers in race and gender (Bushnell, 2003; Pascalis et al., 2014; Scott, Pascalis, & Nelson, 2007).

Infants who are deprived of early face experiences show permanent deficits in configural face processing, a signature property of mature human face processing (Maurer, Le Grand, & Mondloch, 2002). Infants born with congenital cataracts that were removed as early as when the infants were 2 to 6 months of age do not develop configural face processing (Ellemberg, Lewis, Liu, & Maurer, 1999; Gwiazda, Bauer, Thorn, & Held, 1997; Maurer, Ellemberg, & Lewis, 2006; Maurer & Lewis, 2001a, 2001b) even though they show typical visual development by many other measures. Configural face processing begins to emerge, not in infancy, but in childhood (Carey & Diamond, 1994; de Heering, Houthuys, & Rossion, 2007; De Heering, Rossion, & Maurer, 2012; Pellicano & Rhodes, 2003; Schwarzer, 2002; Tanaka, Kay, Grinnell, Stansfield, & Szechter, 1998). It has been hypothesized that very early visual experiences preserve and/or establish the neural substrate for this later development (Maurer, Mondloch, & Lewis, 2007). Infants raised in orphanages with a selective deficit in experiences of caretaker faces also show atypical patterns of face processing even after they had been placed in regular social environments (Moulson, Westerlund, Fox, Zeanah, & Nelson, 2009; Parker, Nelson, & Group, 2005). These results suggest that infants' early exposure to the specific faces in their visual environment may be a defining step on the developmental pathway to mature face processing.

In principle, the implied sensitive period for early face experiences would be determined by internal changes in ocular structure and neural plasticity. Changes in ocular structure will change the quality of visual input that is received from the environment – through differing abilities in vergence, accommodation and contrast sensitivity – and passed on to the brain. Changes in neural plasticity will determine how the brain processes the visual information received from the ocular system. Due to such significant changes in the visual system throughout infancy, the faces in the environments of infants of different age groups could well

---

* Corresponding author.
E-mail addresses: swapnaa@indiana.edu (S. Jayaraman), smith4@indiana.edu (L.B. Smith).

https://doi.org/10.1016/j.visres.2018.05.005
Received 15 October 2017; Received in revised form 19 May 2018; Accepted 23 May 2018

Please cite this article as: Jayaraman, S., Vision Research (2018), https://doi.org/10.1016/j.visres.2018.05.005

be the same in terms of the specific identities of faces and the frequencies of those faces in the visual environment, for there to be a difference in visual experiences during and after the sensitive period. Contrary to this hypothesis, emerging research using head cameras to capture infant egocentric views in everyday contexts suggest that the visual environment with respect to faces changes throughout infancy (Fausey, Jayaraman, & Smith, 2016; Jayaraman, Fausey, & Smith, 2015, 2017; Sugden, Mohamed-Ali, & Moulson, 2014). In particular, faces appear to be much more frequent in the experiences of infants younger than 3 months of age than in the experiences of older infants. For example, one in-home head camera study (Jayaraman et al., 2015) found that faces were in view for about 15 min out of each awake hour for infants under 3 months of age but were in view only 6 min out of every hour by the time infants approached their first birthday (see also Fausey et al., 2016). The relative rarity of faces in the first-person views of older infants has been replicated by many laboratories in many different contexts (Frank, Simmons, Yurovsky, & Pusiol, 2013; Pereira, Smith, & Yu, 2014; Yu & Smith, 2012). In brief, the visual input about faces may be different for younger and older infants.

Visual learning depends on both the visual processes that filter and change with experience, and the properties of the visual input itself. Thus differences in the input – perhaps in conjunction with or perhaps as a product of the distinct biases and properties of the young visual system – could themselves be a key factor in the outsized importance of early face experiences (Smith, Jayaraman, Clerkin, & Yu, 2018). Accordingly, in this study we focus on the properties of visual input itself and how visual face experiences may differ for younger and older children with the goal of a more detailed description of the potentially unique properties of early face experiences, beyond mere frequency. We focus on "persistence", a tendency for a face belonging to an individual to remain in view of the infant for a contiguous stretch of time: how persistence changes with age and how persistence relates to other visual properties of the faces that infants see. We do so for two reasons: First, given the limits and biases in young infants' visual systems and their limited motor abilities, their face experiences may principally result from when caregivers put their own faces in front of infants in the contexts of caring for them or face-to-face play (Bremner, 2017; Fausey et al., 2016). This context may yield not just many experiences of faces that are close and visually large but also experiences that are extended in time and that involve very few individual faces. Second, and as considered in more detail in the Discussion section, visual experiences that extend in time may engage learning mechanisms central to both early and later visual face processing (Berkes & Wiskott, 2005).

2. Method

2.1. Overview and rationale for the approach

We analyzed the frequency, persistence in time, proximity, identity, and pose of faces in a large corpus of head camera images collected by infants as they went about their daily lives (Jayaraman & Smith, 2017). The corpus contains images from infants 1 month to 15 months of age. Although our central questions concern the potentially distinct properties of face experiences of very young infants (under 3 months old), we analyzed all the available images across the age range to generate a more complete developmental picture of how face experiences change during infancy. It is only by considering a broad age range that we can identify how and in what ways the visual face experiences of very young infants may be unique.

The extant evidence indicates the first three months of post-natal life as one of rapid learning about faces and a potentially sensitive period for the development of face processing. Accordingly, the first defined age group consists of infants aged 1–3 months. Our youngest infant was 3 weeks of age and contributed to only 15 min of data (our data collection procedure was too demanding for parents of newborns). Subsequent age groups were defined as shown in Table 1. The first 4 groups span 2 month periods; the oldest two groups span 3 month periods. We used larger age-bins for the two oldest groups (9–12 months and 13–15 months) because of the fewer data sets in the corpus at these ages and because we expected minimal changes in face experiences at these oldest groups.

Table 1
Description of corpus: data contributed by individual infants grouped for analyses.

Age group	Infants in group	Recording time (hours)	Coded frames
1–3 months	8	33.26	23,945
3–5 months	9	39.57	28,492
5–7 months	10	49.59	35,703
7–9 months	9	46.58	33,541
9–12 months	9	37.20	26,787
12–15 months	6	26.49	19,073
Total	51	232.7	1,005,246

The main analyses of age-related effects are based on the defined age groups and not on individuals. Age-group analyses are the statistically superior approach given the nature of the data. Each infant contributed on average about 4 h of head camera video. Even when down-sampled (see below), the number of data points from individual participants are considerable. However, these datasets have two important attributes to consider. First, the visual entities in the world – like words in language – are not normally nor uniformly distributed but are extremely right-skewed and bursty (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Jayaraman et al., 2015; Salakhutdinov, Torralba, & Tenenbaum, 2011). Thus, small, context-bound samples can be misleading about the properties of the whole distribution (Montag, Jones, & Smith, 2018). Second, and consequently, the current sampling from each infant – despite an average of 4 h of video obtained at random intervals in time – is sparse for the purposes of generalizing the experiences of one individual infant for that age. Our goal is to describe the properties of experience that characterize periods of development (rather than to characterize an individual infant). Accordingly, the statistical solution is to amass large samples of data across contexts and individuals as estimation of the distribution of faces typical for each period of development. Note that major advances in understanding language acquisition have emerged from a similar approach of analyzing the statistical properties of large corpora of infant-directed speech assembled from many different children (MacWhinney, 2000).

2.2. Ethics statement

All experimental protocols and consent materials were approved by the Indiana University Institutional Review Board. Parents of all participating infants provided written informed consent prior to the experiment. The present research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.3. Collection of the corpus

We collected the corpus of scenes from a cross-sectional sample of 51 infants (25 female) who were all between the ages of 1 and 15 months. 24 of these infants contributed images to the head camera analyses in Jayaraman et al. (2017), Jayaraman et al. (2015), and Fausey et al. (2016). To collect the infant perspective scenes, we used a commercially available, wearable camera (Looxcie) that was easy for parents to operate, safe (did not heat up), and very lightweight (22 g). We secured the camera using elastic loops stitched on a hat that was custom fit to the infant so that when the hat was securely placed on the infant, the lens was centered above the nose and did not move (see Jayaraman et al. (2015) for details). We gave parents the hats fitted with cameras and instructed them on how to use them at an initial meeting. We asked parents to collect videos throughout the daily activities of their infant, and told them that we were interested in visual development and that they were free to record whenever it suited their family's schedule. Parents recorded videos at various times of day and in multiple locations (home, playground, store, church, etc.) while infants were engaged in a variety of activities.

The diagonal field of view (FOV) of the camera was 75 degrees, vertical FOV was 42 degrees, and horizontal FOV was 69 degrees with a 2″ to infinity depth of focus. The camera recorded at 30 Hz, and the battery life of each camera was one to three continuous hours, rechargeable at any time. We gave participating families two to four cameras and asked them to record up to 6 h of video. Video was stored on the camera until parents had completed their recording and then we transferred them to laboratory computers for storage and processing. Head cameras measure the scene in front of the viewer; they do not provide momentary gaze information and in principle, gaze could be outside of the head-camera field. However, head mounted eye-tracking studies show that under active viewing conditions, human observers including infants typically turn both heads and eyes in the same direction, align heads and eyes within 500 ms of a directional shift, and maintain head and eye alignment when sustaining attention (Ballard et al., 1992; Bambach, Crandall, & Yu, 2013; Bloch & Carchon, 1992; Daniel & Lee, 1990; Pereira et al., 2014; Ruff & Lawson, 1990; Schmitow & Stenberg, 2015; Schmitow, Stenberg, Billard, & Von Hofsten, 2013; Yoshida & Smith, 2008). The result is that the distribution of gaze in active viewing in space (not on screens) is centered in the head camera image. Thus, in a large corpus of images recorded during active viewing, the likelihood of gaze falling within the head-camera image is over 97% (Bambach et al., 2013; Bambach, Smith, Crandall, & Yu, 2016; Foulsham, Walker, & Kingstone, 2011; Li, Fathi, & Rehg, 2013).

2.4. The corpus

The mean length of video collected per infant was 4.6 h (SD = 1.4) yielding 232.7 h of head camera video (recorded at 30 Hz) and over 25 million frames. We sampled 1 frame every 5 s for analysis (Fig. 1), generating over 1 million analyzed frames. We pooled data contributed by individual infants into six groups by age (see Table 1) and performed all developmental analyses on these grouped data.

2.5. Image coding

Trained human coders coded the images through a series of steps as follows.

2.5.1. The presence of a face

We organized sampled images into sets of 100 for coding by naïve coders who were paid for their work. We trained coders on a set of nine instruction images. Four different coders answered a single yes/no question about each image: "do you see a face or a part of a face in this image?". An image was deemed to contain a face if at least three out of four coders agreed. For 95.8% of the frames, at least three coders agreed that the frame did (or did not) contain a face or a face part. Coders further analyzed each image that contained a face or a face part for attributes described below. In cases where there were two or three faces per image (18.6% of face images), coders coded each face separately for all attributes. Images that contained four or more faces were not coded any further.

2.5.2. Face attributes: Identity

To code unique identities, we combined the sampled frames to form a video – a stream of faces in time. Trained coders kept track of the unique faces in each subject's recording and assigned a unique identifier to each face seen in the videos, in order of appearance. Once the entire video for a subject was coded, we ordered the unique identifiers by rank (from most to least frequent appearance) and assigned an alphanumeric code to each. Identifiers were not assigned for the faces in frames in which the face was occluded, blurry, too small or in a crowd of 4 or more faces and thus difficult to identify. Identifiers for faces in media (product packaging, television, books) were also not assigned. We included all these faces in the total face count but not the unique identities count. The proportion of unidentified faces (for all the reasons listed above) was very small, less than 0.01 of all face frames. For analyses that required identity information for a group of infants, we matched faces from all infants from the group by rank and assigned the same identifier. For example, for the most frequently seen identities in the datasets of subject 1 and subject 2 within a group, we assigned the same identifier A, the second most frequent face B and so on. A second coder independently reanalyzed 20% of the frames, and coder agreement was 98%.

2.5.3. Face attributes: Distance

To estimate the distance of each face from the infant wearing the head camera, trained coders matched faces (and parts of faces) to size templates that were generated by determining head camera image size for an average adult female face at 1-foot increments from the head camera. Coder agreement exceeded 98%.

2.5.4. Face attributes: Pose

We asked coders to determine the pose of each face and face part in the images. Coders assigned different codes for faces showing a frontal pose displaying both eyes and ears, an angled pose displaying both eyes and one ear, a profile pose displaying one eye and ear, a profile pose with one ear and neither of the eyes visible, and a top/bottom pose showing only the forehead and nose tip or neck and nostrils. As with all our face attribute coding, faces that weren't fully in view of the camera were coded as though the parts that were cut off were also in view. Coder agreement was 93%.

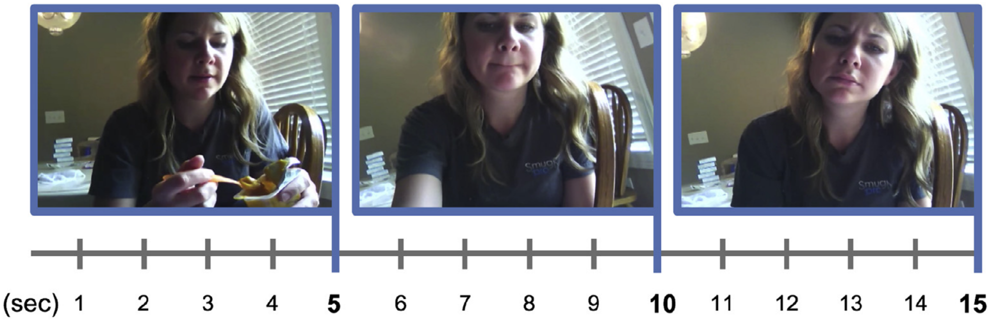

Fig. 1. Down-sampled frames from head camera video at 1 frame every 5 s.

2.6. Face run lengths

We determined run lengths of contiguous images containing the same face from the identity coding and the stream – ordered in time – of the down-sampled images. This stream provided a snapshot of the input every five seconds. We use run length of the same face as a proxy measure of the persistence of the same individual faces in the infant views. For a set of contiguous face images to be considered as a run, the identity of the face must be the same across the entire run. By this definition, eight contiguous face images displaying person A in the first three images and person B in the next five images were coded as two separate runs of 3 and 5. A run of three contiguous images of the same face suggests that face is available for viewing for 15 s, albeit with potential missed gaps. The lengths of these contiguous face images, or "face run lengths", thus provide a rough but informative measure of the developmental changes in the temporal properties of faces in infant experiences.

3. Results

3.1. Frequency of faces

Fig. 2 shows the proportion of face frames in the sampled images by age group. A linear mixed effects analysis of the relationship between infant age group and proportion of faces in view of infants was performed using R (R Core Team, 2014) and lme4 (Bates, Maechler, Bolker, & Walker, 2014). Age groups were modeled as fixed effects and individual subjects as random effects. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the effect of age group against the model without the effect of age group, confirming that the proportion of faces in view significantly declined with age group [χ^2 (1) = 18, $p < 0.01$]. An analysis of variance (ANOVA) on the proportion of face images in view also yielded significant variation between age groups [F(4,45) = 3.8, $p < 0.01$]. A post hoc Tukey test showed that the youngest age group (1–3 months) experienced significantly higher proportions than the 7–9 month group and the 12–15 month group at $p < 0.05$.

The proportions in Fig. 2 provide an estimate of the amount of face time available to infants. The 1–3 month olds experienced faces nearly 18 min out of every recorded hour, whereas the 12–15 month olds only got about 10 min per hour of recorded face time, findings similar to those reported earlier (Jayaraman et al., 2015). One new question for this study is whether this greater frequency reflects the greater persistence of faces in time, measured as face run lengths, for younger than older infants.

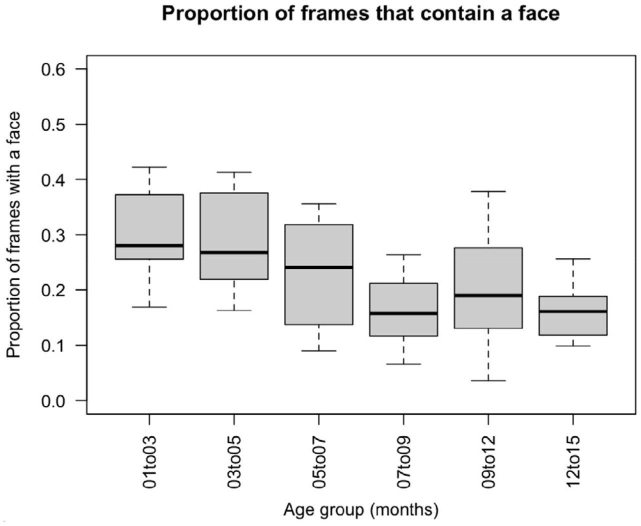

Fig. 2. The proportion of frames that contain a face as a function of the age group.

3.2. Face run lengths

Fig. 3a shows the frequency distribution of all face run lengths obtained. The observed distribution plotted in orange seems to closely follow a power function, specifically a Zipfian function (Newman, 2005; Zipf, 1949), as evidenced by the estimated values plotted in grey. The approximate straight-line form of the distribution in logarithmic scales (Fig. 3b) implies that the distribution follows a power law. To prove linearity, an ordinary least squares regression with the squared (non-linear) rank term in the regression model was computed. The coefficient of this term was statistically insignificant, suggesting that the effects were linear in the rank term ($R^2 = 0.90$, F(2,49) = 230, $p < 0.0001$). This face runs distribution joins the ranks of several natural phenomena in which power-law distributions occur. For all infants, most run-lengths are very short but there is a very long tail of much longer run lengths.

Fig. 4 shows the distribution of run length as a proportion of total runs for each age group. All run lengths of 4 and above are combined in the main figure and their full distributions are plotted in the inset. The youngest infants (1–3 month olds) experienced the highest proportions of long face runs (about 17% of all runs are 4 or longer) and this proportion is different [χ^2 (5) = 292.5, $p < 0.0001$] from those that all other ages experience (ranging from 5 to 10%). Perhaps consequently, the youngest infants also experience the lowest proportion of short run lengths (a run length of 1 is 60% of total runs) compared to the 70–77% that their older counterparts experience [χ^2 (5) = 257.4, $p < 0.0001$]. Therefore, the previously reported changes in face frequencies during the first year does reflect, at least in part, a transition from more temporally persistent faces in the infants' field of view to much briefer face events.

3.3. Face attributes

What are the visual attributes that define these face experiences and how are they distributed across these age groups? Fig. 5 shows, for each age group, the proportion of all face images that contain faces, with respect to the three previously reported dominant attributes of infant face experiences (Jayaraman et al., 2015): the face belongs to one of the top 3 most frequently encountered individuals in the infant's environment (a), the face displays both eyes (b), and the face appears within 2 feet of the infant (c).

A linear mixed effects analysis of the effects of infant age group modeled as fixed effects and subjects modeled as random effects on the proportion of faces showing these three attributes was performed. The proportion of faces that belonged to the top three most frequently encountered faces did not change with age [χ^2 (5) = 7.3, p = 0.2], nor did the proportion of faces that presented both eyes [χ^2 (5) = 2.19, p = 0.82]. There appeared to be a decline in the proportion of up-close faces with age, however, not one that was statistically significant [χ^2 (5) = 10.5, p = 0.06]. Separate analyses of variance (ANOVA) on the proportion of total face images exhibiting each of the three attributes confirmed that age group did not have an effect on the proportion of times when a face belonged to one of the 3 most frequently occurring faces [F(4,45) = 1.39, p = 0.25], when a face presented both eyes [F(4,45) = 0.4, p = 0.85], or when a face was within 2 feet from the infant [F(4,45) = 2.25, p = 0.065]. Overall, these results (from a larger corpus than the original report, Jayaraman et al., 2015) suggest at best small changes in these three properties of faces. Across the age groups examined, faces decline in frequency and in persistence but are generally close frontal views of a small number of people for all infants. This finding in and of itself suggests that persistence may be the critical stimulus property unique to young infants.

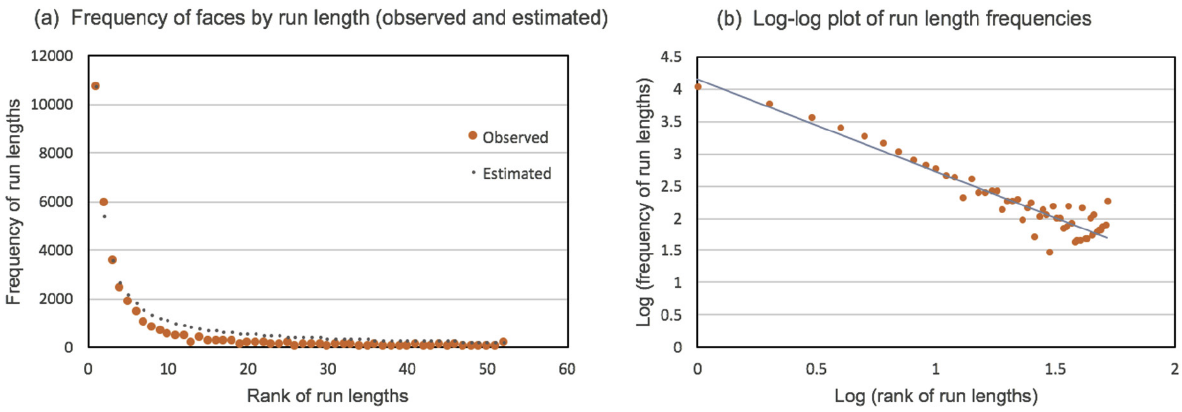

Fig. 3. Distribution of the frequency of observed face run lengths by rank in orange and the estimated Zipfian function in grey (a). On the right is the same observed data but plotted on logarithmic scales (b).

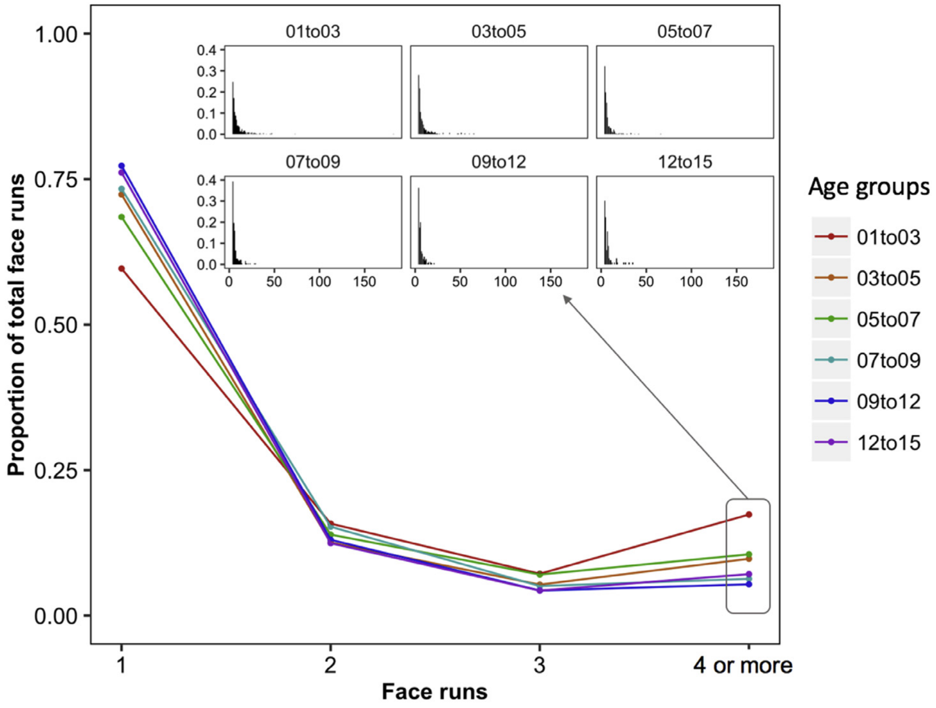

Fig. 4. Proportion of face run lengths by age group.

However, the results show that these properties also vary as function of run length. In Fig. 6, the three measured face attributes are plotted as the proportion of total runs for each run length (4 and above runs are combined) for each of the age groups. To answer the question of whether faces in longer run lengths are more likely to exhibit each of the three basic face attributes, binary logistic regression models using run length and age group as predictor variables were used, and the corresponding marginal effects were calculated. The dependent variable in each of the models is whether the faces exhibit one of the 3 attributes: the face belongs to one of the 3 most frequently seen individuals, the face displays both eyes, and the face is within 2 feet from the infant. The presence of these properties is denoted by a 1 and the absence by a 0. Since the dependent variables are discrete, the ordinary least squares regression can technically be used to fit a linear probability model. However, since the linear probability model is heteroskedastic and may predict probability values beyond the (0,1) range, logistic regression models using run length and age group as predictor variables were used. As the number of runs increases, the faces in these runs are significantly more likely to exhibit each of the three basic properties. The likelihood of a face belonging to one of the top 3 identities increases by 1.33 times with every increase in the number of runs ($p < 0.001$), and decreases by 0.12 times with every increase in age group ($p < 0.001$). Similarly, the likelihood of a face displaying both eyes increases by 1.19 times with an increase in run length ($p < 0.001$), but does not change with age ($p = 0.49$). Finally, the likelihood of a face being within 2 feet of the infant increases by 1.18 times with every increase in run length ($p < 0.001$) and decreases by 0.07 times with every increase in age group ($p < 0.001$).

In sum, long face runs as opposed to shorted ones at all ages are likely to be runs of the one of three most frequent faces, to be frontal views and to be close. To capture how these properties – top 3 individuals, frontal views, and proximity – may be amplified for the younger infants by the long tail of more persistent faces, we calculated a binary measure “face quality” that signifies that a face carries all 3 attributes: belongs to one of the most frequently appearing individuals, presents both eyes, and is within 2 feet from the infant. Each image in our corpus was given a binary code: 1 if the face possessed all 3 attributes and 0 if one or more of the 3 attributes was lacking. To determine

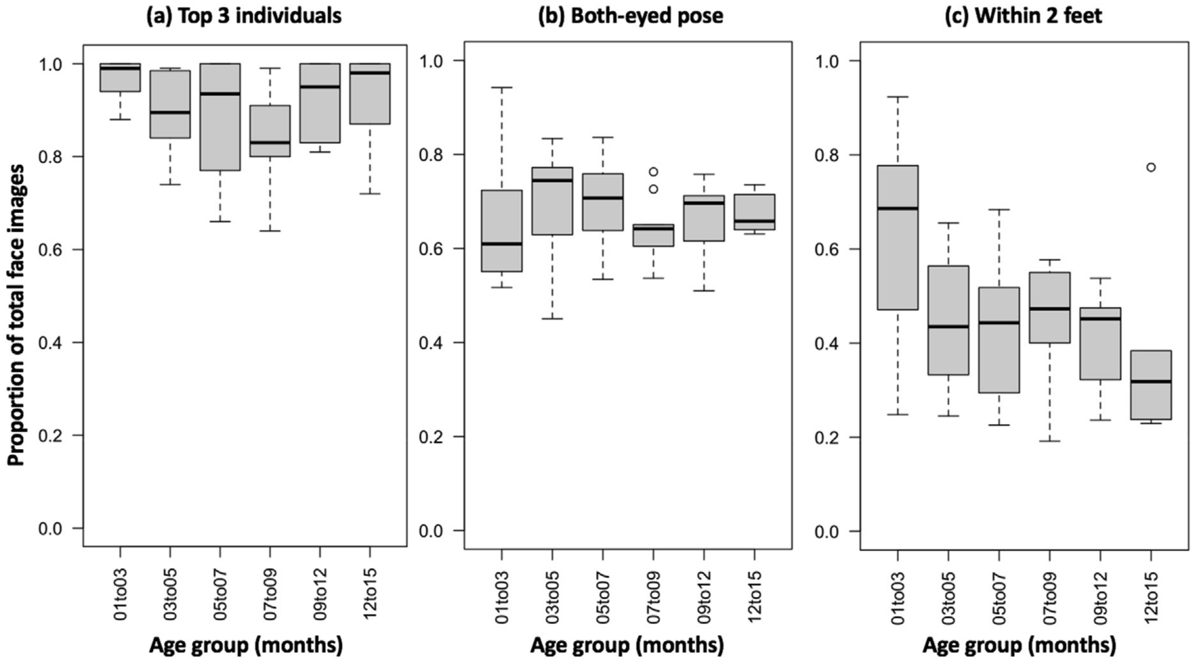

Fig. 5. Proportion of total face images that contain the top 3 most frequently appearing individuals (a), display both eyes (b), and appear within 2 feet of the infant (c) by age group.

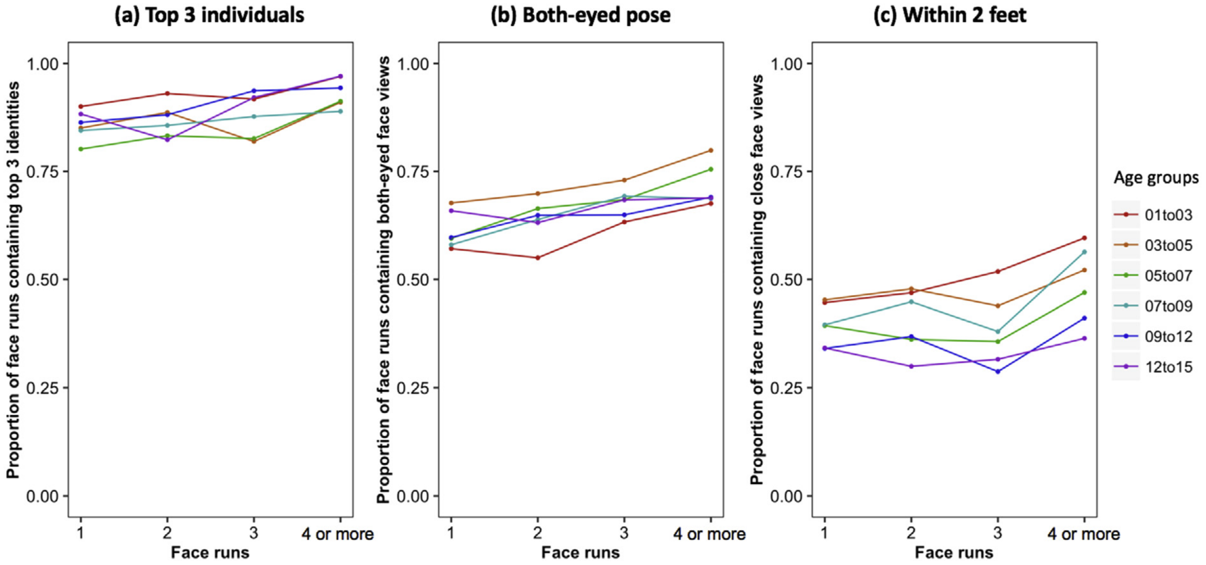

Fig. 6. Proportion of face runs that contain the top 3 most frequently appearing individuals (a), display both eyes (b), and appear within 2 feet of the infant (c) by age group.

if the likelihood of face quality changes with a change in run length or age group, a binary logistic regression model using run length and age group as predictor variables was run. As the number of runs increases, faces are significantly more likely to be of high quality. The likelihood increases by 1.2 times with every increase in the number of runs ($p < 0.001$), and decreases by 0.07 times with every increase in age group ($p < 0.001$). Fig. 7 shows the quality of faces in the long (4 or more) face runs for each age group. The proportions of images with high quality faces, shown in white at the bottom of the stacked bars, reliably decline with age [χ^2 (5) = 19.79, $p < 0.01$]. As is evident, the youngest age of infants (1–3 month olds) who see more faces per unit time than the other age groups, also see more faces in longer runs, and more of those longer runs consist of frontal view of a close and highly familiar face.

These conclusions were confirmed by pair-wise comparisons of the age groups which show that the youngest two age groups (1–3 and 3–5 month olds) exhibit significantly higher proportions of quality in these long face exposures than all other age groups ($p < 0.05$). These relatively frequently occurring long face bouts are characterized by faces with three critical properties: close and frontal views that present information about major face properties and that are of just a very few people.

4. Discussion

The regularities in very young infants' visual worlds may have outsized effects on the development of the visual system because they comprise the first-in experience that specifies, tunes, and maintains the neural substrate from low-level to higher-level representations and therefore constitute the starting point for all other visual learning. Thus,

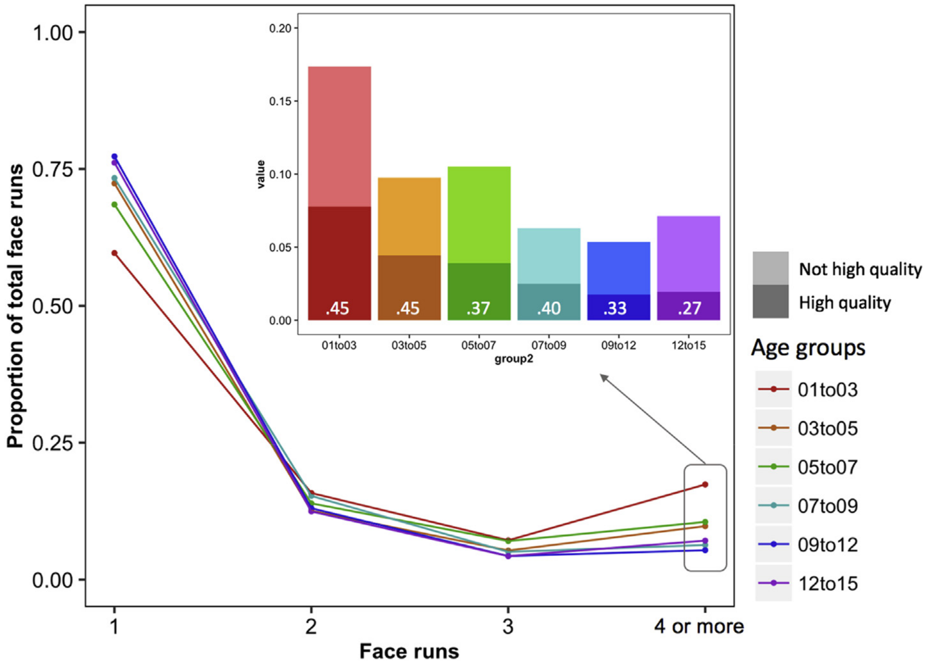

Fig. 7. Quality of faces in the long (4 or more) face runs by age group. The proportions of images with high quality faces are shown in white at the bottom of the stacked bars.

the regularities in early visual experiences are critical whether they are unique to infancy or occur during other periods of life. The present findings indicate that one of the key differences between very young infants' experiences of faces and those of older infants is the persistence of face experiences in time. Very young infants do not just have more frequent face experiences but also more temporally enduring ones. Older infants, on the other hand, tend to have less frequent and briefer encounters of faces. The age-related changes in other properties of face views are not marked; up-close frontal views of a very few individual people constitute most face experiences from the first month of life to the fifteenth. However, repetitions of the same very few identities, up-close and frontal are exaggerated in more persistent runs of the same face, and these persistent runs – as well as these qualities among the faces in these persistent runs – are more frequent for the youngest infants. If early visual face learning depends on extended repeated exposures of up-close frontal views, then very young perceivers may have the optimal face input for setting up the face processing system.

Why might persistence in time – not just frequency – be a critical stimulus factor in early input? First, the retinal information relevant to perceiving any visual entity, including faces, varies continuously in lighting, color, pose, location, and occlusion as the result of even small movements by the perceiver as well as by events in the physical world. Considerable evidence indicates that perceivers discover the relevant invariances by tracking stable and slow-changing information in time (Földiák, 2008; Wiskott & Sejnowski, 2002). There is now growing evidence for this idea at the neural level in receptive field formation (Dähne, Wilbert, & Wiskott, 2014) and in human, animal and machine learning of higher perceptual invariances in object recognition (Franzius, Wilbert, & Wiskott, 2011; Li & DiCarlo, 2008; Wiskott & Sejnowski, 2002; Wood & Wood, 2016). More generally, persistence (and/or proximity) in time supports the integration of varying input into a single coherent representation. Second, real world faces present meaningful dynamic information – smiles, eye-gaze shifts, eye-brow lifts and nods. Because these dynamic cues evolve in time, temporally longer exposures to dynamic face information may support learning. The evidence indicates that this learning does begin very early (Farroni, Johnson, Brockbank, & Simion, 2000; Hood, Willen, & Driver, 1998; Vecera & Johnson, 1995; Xiao, Quinn, Wheeler, Pascalis, & Lee, 2014; Zhu, Zhang, Luo, Dilks, & Liu, 2011). Finally, persistent dynamic faces are likely to attract and hold the infant's attention enabling deeper processing and faster learning.

The whole suite of properties that characterize most of infants' seen faces and that dominate more enduring face experiences may work together to build the early substrate that enables young infants to recognize familiar faces and prepares the system for later developments in face processing. Dense persistent visual experiences with *a few individuals* may be critical for extracting subtle face properties and emotional cues, an idea that fits findings on the role of familiarity in young infants' recognition of emotional cues (Kahana-Kalman & Walker-Andrews, 2001). Persistent close frontal views match young infants early visual skills and biases and may hold attention for learning. Experiences of a few up-close familiar individuals may also support learning by providing a stable backdrop against which to discover and extract meaning-laden facial gestures amidst the flux of other irrelevant retinal light patterns. Indeed, research on infant (and adult) sensitivity to eye gaze depends on stable frontal views of faces (Corkum & Moore, 1998; Doherty, Anderson, & Howieson, 2009; Farroni et al., 2000). These hypotheses – generated by measuring the ego-centric visual environments of infants in their everyday lives – need to be tested in experimental studies. Clearly, there is much more that we need to know about the detailed dynamic events that comprise the longer run-lengths reported here.

An additional question about the input and its potential special characteristics for young infants concerns the multimodal nature of these early persistent dynamic experiences, and particularly the sounds that parents make when their faces are in front of their infant. Some laboratory research suggests that infants younger than 3 months may not adeptly use multi-modal synchronies to recognize faces (Bahrick, Gogate, & Ruiz, 2002; Bahrick, Todd, Castellanos, & Sorondo, 2016), even though they are sensitive to coordination of their caregivers' typical facial expressions and sounds (Izard et al., 1995; Kahana-Kalman & Walker-Andrews, 2001; Walker-Andrews, 1997). New evidence suggests that multimodal – auditory-visual – experiences in early infants may have long term consequences. Adults who were born with cataracts removed by 4 months of age show deficits in the processing of sight-sound synchronies as well as in configural face processing (Bremner, 2017; Chen, Lewis, Shore, & Maurer, 2017). Bremner (2017) proposes that both deficits may derive from the lack of the same early evolutionarily-expected experience: the up-close faces of caretakers as they talk to, coo, and smile at their very young infants.

The run-lengths of the same face is extremely right-skewed for all infants in the present study. The discussion has concentrated on the role of the heavy tail – heavier in the youngest infants – of longer runs of the same face. What is the role of the much more frequent briefer face events in infants' visual fields? One possibility is that the rarer more enduring face experiences do the heavy-lifting in learning and the briefer encounters play a limited role. However, the entire right-skewed distribution – with many brief encounters and some very long encounters – of the faces of a very few people may be relevant to early visual learning about faces. The frequent brief encounters of the individuals most often encountered in an infant's life may create what has been called a "desirable difficulty" (Bjork & Bjork, 2011), activating representations honed by more persistent experiences and forcing the learning system to rapidly identify individual faces. These are clearly important open questions for future research.

Because research on ego-centric vision and the natural statistics of point-of-view environments is in its early days, there is much that we do not know and strong conclusions and hypotheses in any direction are probably not warranted. However, we do know that during the first three months of life, infants visual systems instantiate the statistical regularities of the specific faces in infant environments (Scott et al., 2007) and that visual experiences in those early months are relevant to later developments in face processing (Maurer et al., 2007). A reasonable assumption is that the visual mechanisms that accomplish this are well suited to the natural distributional statistics of faces in infant lives and in the visual regularities presented by those faces. For these reasons, the study of the everyday visual environments of infants is essential to understanding how and why human face processing has the properties that it does. The full explanation will also require recognizing that infants are not stationary learners. Over the first two years of life, the bodies and sensory-motor abilities of infants change markedly creating a series of different visual environments with very different properties. Very young infants have limited acuity and can do very little with their bodies (Braddick & Atkinson, 2011). As a consequence, much of what they see depends on what caretakers put in front and close to the infant's face (Jayaraman et al., 2015). An older crawling baby can see much farther and can move to a seen distant object for a closer view; however, when crawling, the infant sees only the floor and actually has to stop crawling to sit up to see social partners (Kretch, Franchak, & Adolph, 2014). Each new sensory-motor achievement both opens gates to new regularities and closes gates to previously available regularities. The evidence from congenital cataract patients suggests that first three months of life may be a sensitive period for visual face processing. Such sensitive periods may arise from changes in the neural plasticity of the relevant systems (Knudsen, 2004). However, they may also result from developmental changes in the egocentric environments. Because the personal view of an infant depends not just on what is in the world but on the infants own sensory-motor abilities, the gate on the up-close faces of caretakers may close as a consequence of the infants own increasing motoric autonomy (Fausey et al., 2016; Smith et al., 2018).

Acknowledgments

The research was supported by the National Science Foundation grant BCS-15233982 and by Indiana University through the Emerging Area Initiative titled Learning: Brains, Machines and Children. We thank Char Wozniak, Ariel La, and Elizabeth Hart for managing the data collection and preparation.

References

Bahrick, L. E., Gogate, L. J., & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: The salience of actions over faces in dynamic events. *Child Development, 73*(6), 1629–1643.

Bahrick, L. E., Todd, J. T., Castellanos, I., & Sorondo, B. M. (2016). Enhanced attention to speaking faces versus other event types emerges gradually across infancy. *Developmental Psychology, 52*(11), 1705.

Ballard, D. H., Hayhoe, M. M., Li, F., Whitehead, S. D., Frisby, J., Taylor, J., & Fisher, R. (1992). Hand-eye coordination during sequential tasks [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences, 337*(1281), 331–339.

Bambach, S., Crandall, D. J., & Yu, C. (2013). Understanding embodied visual attention in child-parent interaction. Paper presented at the Development and Learning and Epigenetic Robotics (ICDL), 2013 IEEE Third Joint International Conference on.

Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016). Objects in the center: How the infant's body constrains infant scenes. Paper presented at the Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2016 Joint IEEE International Conference on.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using Eigen and S4. *R Package Version, 1*(7), 1–23.

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision, 5*(6), 9.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society, 2*, 59–68.

Bloch, H., & Carchon, I. (1992). On the onset of eye-head coordination in infants. *Behavioural Brain Research, 49*(1), 85–90.

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

Bremner, A. J. (2017). Multisensory development: Calibrating a coherent sensory milieu in early life. *Current Biology, 27*(8), R305–R307.

Bushnell, W. (2003). The development of face processing in infancy and early childhood. In Olivier Pascalis, & Alan Slater (Vol. Eds.), *The development of face processing in infancy and early childhood: Current perspectives: vol. 41*, (pp. 41–53). Nova Science Publishers, Inc.1-59033-696-8.

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

Chen, Y.-C., Lewis, T. L., Shore, D. I., & Maurer, D. (2017). Early binocular input is critical for development of audiovisual but not visuotactile simultaneity perception. *Current Biology, 27*(4), 583–589.

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.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology, 34*(1), 28.

Dähne, S., Wilbert, N., & Wiskott, L. (2014). Slow feature analysis on retinal waves leads to V1 complex cells. *PLoS Computational Biology, 10*(5), e1003564.

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

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96*(1), 57–70.

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

Dobson, V., Teller, D. Y., & Belgum, 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 3years. *Journal of Experimental Child Psychology, 104*(3), 296–312.

Ellemberg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research, 39*(14), 2325–2333.

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.

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.

Földiák, P. (2008). Learning invariance from transformation sequences. *Learning, 3*(2).

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.

Frank, M. C., Simmons, K., Yurovsky, D., & Pusiol, G. (2013). Developmental and postural changes in children's visual access to faces. Paper presented at the Proceedings of the 35th annual meeting of the Cognitive Science Society.

Franzius, M., Wilbert, N., & Wiskott, L. (2011). Invariant object recognition and pose estimation with slow feature analysis. *Neural Computation, 23*(9), 2289–2323.

Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1997). Development of spatial contrast sensitivity from infancy to adulthood: Psychophysical data. *Optometry & Vision Science, 74*(10), 785–789.

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*(2), 131–134.

Izard, C. E., Fantauzzo, C. A., Castle, J. M., Haynes, O. M., Rayias, M. F., & Putnam, P. H. (1995). The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Developmental Psychology, 31*(6), 997.

Jayaraman, S., & Smith, L. B. (2017). The Homeview Project. Retrieved from < http://www.iub.edu/~cogdev/homeview.html/ > – Homeview.

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.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2017). Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology, 53*(1), 38.

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.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72*(2), 352–369.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*(8), 1412–1425.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development, 85*(4), 1503–1518.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 321*(5895), 1502–1507.

Li, Y., Fathi, A., & Rehg, J. (2013). Learning to predict gaze in egocentric video. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

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.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database*MIT Press.

Maurer, D., Ellemberg, D., & Lewis, T. L. (2006). Repeated measurements of contrast sensitivity reveal limits to visual plasticity after early binocular deprivation in humans. *Neuropsychologia, 44*(11), 2104–2112.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.

Maurer, D., & Lewis, T. (2001a). Visual acuity and spatial contrast sensitivity: Normal development and underlying mechanisms. *Handbook of Developmental Cognitive Neuroscience,* 237–250.

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

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

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

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.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*(5), 323–351. http://dx.doi.org/10.1080/00107510500052444.

Oruc, I., & Barton, J. J. S. (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.

Parker, S. W., Nelson, C. A., & Group, B. E. I. P. C. (2005). An event-related potential study of the impact of institutional rearing on face recognition. *Development and Psychopathology, 17*(3), 621–639.

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.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science, 14*(6), 618–622.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review, 21*(1), 178–185.

R Core Team (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology, 26*(1), 85.

Salakhutdinov, R., Torralba, A., & Tenenbaum, J. (2011). Learning to share visual appearance for multiclass object detection. Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on.

Schmitow, C., & Stenberg, G. (2015). What aspects of others' behaviors do infants attend to in live situations? *Infant Behavior and Development, 40*, 173–182.

Schmitow, C., Stenberg, G., Billard, A., & Von Hofsten, C. (2013). Using a head-mounted camera to infer attention direction. *International Journal of Behavioral Development, 37*(5), 468–474.

Schwarzer, G. (2002). Processing of facial and non-facial visual stimuli in 2–5-year-old children. *Infant and Child Development, 11*(3), 253–269.

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.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*. http://dx.doi.org/10.1016/j.tics.2018.02.004.

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.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5*(4), 479–496.

Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition, 2*(1), 59–87.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*(3), 437.

Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*(4), 715–770.

Wood, J. N., & Wood, S. M. (2016). The development of newborn object recognition in fast and slow visual worlds. Paper presented at the Proc. R. Soc. B..

Xiao, N. G., Quinn, P. C., Wheeler, A., Pascalis, O., & Lee, K. (2014). Natural, but not artificial, facial movements elicit the left visual field bias in infant face scanning. *Neuropsychologia, 62*, 175–183.

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. (2012). Embodied attention and word learning by toddlers. *Cognition, 125*(2), 244–262.

Zhu, Q., Zhang, J., Luo, Y. L., Dilks, D. D., & Liu, J. (2011). Resting-state neural activity across face-selective cortical regions is behaviorally relevant. *Journal of Neuroscience, 31*(28), 10323–10330.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Reading, Mass: Addison-wesley.