Published in final edited form as:
Cognition. 2009 February ; 110(2): 160–170. doi:10.1016/j.cognition.2008.11.010.

Development of infants' attention to faces during the first year

Michael C. Frank[1], **Edward Vul**[1], and **Scott P. Johnson**[2]

1 *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

2 *Department of Psychology, University of California, Los Angeles*

Abstract

In simple tests of preference, infants as young as newborns prefer faces and face-like stimuli over distractors. Little is known, however, about the development of attention to faces in complex scenes. We recorded eye-movements of 3-, 6-, and 9-month-old infants and adults during free-viewing of clips from *A Charlie Brown Christmas* (an animated film). The tendency to look at faces increased with age. Using novel computational tools, we found that 3-month-olds were less consistent (across individuals) in where they looked than were older infants. Moreover, younger infants' fixations were best predicted by low-level image salience, rather than the locations of faces. Between 3 and 9 months of age, infants gradually focus their attention on faces. We discuss several possible interpretations of this shift in terms of social development, cross-modal integration, and attentional/executive control.

A lot can be learned about our social world from the faces of others. Faces provide information about age, race, gender, physical health, emotional state, and focus of attention, giving observers a window into the mental states of other human beings. During the first year after birth, infants begin to extract a large amount of information from faces: they begin to recognize identities (Pascalis, De Haan, Nelson, & De Schonen, 1998), recognize and prefer faces from their own race (Kelly et al., 2005), detect affect (Cohn & Tronick, 1983; Tronick, 1989), and follow gaze (Corkum & Moore, 1998; Scaife & Bruner, 1975). However, these sophisticated abilities are of little use if infants don't look at faces to begin with. Put another way, to extract information from faces, infants must first attend to them.

Although there is a large literature on the origins and development of infants' face representations during infancy, far less research has examined the behavior of infants outside of controlled laboratory settings. In particular, both the extent to which infants attend to faces when other objects are present—as in most real-world situations—and the extent to which this behavior changes across development are still largely unknown. The reasons for this gap in the literature may be both methodological and theoretical. Methodologically, standard looking-time paradigms used in infant research typically produce only qualitative evidence and do not make sense in older populations; hence it is difficult to design experimental paradigms whose results can be compared across wide age ranges. Theoretically, many researchers have been interested primarily in the question of innateness: whether human infants are predisposed to treat human faces as "special" relative to other objects and whether the representations underlying these judgments are qualitatively similar to those used by adults.

Address for correspondence: Michael C. Frank, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave., Room 46-3037D, Cambridge, MA 02139, tel: (617) 452-2474, email: mcfrank@mit.edu.

Our goal in the current study is to address the resulting question—how does infants' attention to faces change across development—by characterizing the development of infants' attention to faces in complex, noisy settings. To motivate our work, we begin by briefly reviewing two literatures: first, face perception in infancy and second, the abilities of adults to detect faces in complex scenes.

A large body of evidence suggests that newborn infants have a generalized bias to attend to faces and face-like stimuli (Cassia, Turati, & Simion, 2004; Farroni et al., 2005; Johnson, Dziurawiec, Ellis, & Morton, 1991; Simion, Macchi Cassia, Turati, & Valenza, 2001). This bias may result from the application of specific face-recognition mechanisms (Farroni et al., 2005; Johnson et al., 1991; Morton & Johnson, 1991) or more general preferences (Cassia et al., 2004; Nelson, 2001; Simion et al., 2001). However, experiments in this literature do not establish either the degree of this preference—what portion of the time infants will typically look at face—or its robustness—to what extent infants prefer to look at faces in noisy, real-world situations.

Work on the later development of face processing has examined the selectivity of infants' representations of faces. Much of this work has supported the *perceptual narrowing* view first proposed in studies of infants' phonetic development (Kuhl, 2000; Kuhl, Tsao, & Liu, 2003). On this view, infants' face representations become specific to those types of faces they see most often as they lose the ability to discriminate the faces of other races and other species during the period from 3 to 9 months (Kelly et al., 2007; Kelly et al., 2005; Pascalis, de Haan, & Nelson, 2002; Pascalis et al., 2005). For instance, 6-month-olds discriminated pairs of monkey faces as well as pairs of human faces, but 9-month-olds provided no evidence of discriminating pairs of monkey faces (Pascalis et al., 2002); the ability to make these discriminations was preserved via repeated exposure to monkey faces (Pascalis et al., 2005). Similarly, baby monkeys reared with no exposure to faces of any species maintained an ability to discriminate both monkey and human faces, but upon selective exposure to monkey faces, discrimination of human faces suffered, and vice versa (Sugita, 2008).

Other research has investigated whether infants' face representations exhibit the same behavioral signatures as adult face processing. For instance, 3-month-olds (but not 1-month-olds) were found to show prototype effects, suggesting that by 3 months, infants represent faces within a "face-space" that shows some similarities to the perceptual space of adult face representations (de Haan, Johnson, Maurer, & Perrett, 2001). Event-related potential (ERP) research has examined differential responses to inverted as opposed to upright faces, a signature of face-specific processing in adults, and suggests that there is an inversion effect in some ERP components by 6 months (de Haan, Pascalis, & Johnson, 2002; Halit, Csibra, Volein, & Johnson, 2004; Halit, de Haan, & Johnson, 2003). All of these results shed light on the format of infants' representations of faces, but do not address whether infants choose to (or are able to) attend to faces in the real world.

In adults, the question of attention to faces has primarily been investigated via visual search tasks. This literature suggests that faces are relatively easy to identify, even in crowded displays (Hershler & Hochstein, 2005; Lewis & Edmonds, 2003, 2005; VanRullen, 2006). For example, Lewis and Edmonds (2005) found that search for faces in grids of scrambled non-face stimuli was efficient, with shallow search slopes (search latencies that did not increase much as the number of distractors increased), suggesting that this search relied on a "parallel," pre-attentive component. In follow-up experiments, they found that inverting the luminance of faces (which makes face-identification quite challenging) increased not only search latencies, but also search slopes. However, VanRullen (2006) showed that pop-out effects could also be found for cars when distractor stimuli were properly controlled, suggesting again that pop-out effects for faces are driven by their lower-level features (such as phase information) rather than their social

relevance. Regardless of whether the relevant cues are low-level perceptual cues or higher-level, face-specific cues, if infants represent faces in a qualitatively similar scheme as adults, infants should show rapid and effortless detection of faces even in displays with multiple distractors. However, because one cannot give explicit verbal instructions to infants, it is impossible to compare how infants and adults perform in explicit visual search tasks.

In the current study, we measured the extent to which faces within complex, dynamic scenes draw attention (as measured by eye-movements) in a task with no explicit instructions. We presented 3-, 6-, and 9-month-old infants and adults a series of 4-second clips from an animated children's cartoon (*A Charlie Brown Christmas*; see Figure 1, top panels) and used a corneal reflection eye-tracker to measure where observers were looking during the video clips.

By using an implicit free-viewing task instead of an explicit search task, we could eliminate reliance on interpretive assumptions linking looking times after habituation/familiarization to preference (Haith, 1998) and directly compare the distribution of attention across a range of age groups, including adults. Nevertheless, comparing the viewing behavior of young infants to that of adults can be problematic: developments in visual acuity (Mayer & Dobson, 1982) might account for changes in infants' tendencies to fixate on faces. To control for this possibility, we showed a separate group of adult participants a version of our stimuli that had been blurred to simulate the contrast sensitivity function of a 3-month-old.

The genesis of this study comes from a previous experiment (Johnson, Davidow, Hall-Haro, & Frank, 2008) in which the *Charlie Brown* cartoon was used as an engaging distractor stimulus. Because children were so drawn to the *Charlie Brown* stimulus, we were able to gather a large amount of data on their fixations patterns, and our anecdotal observation of the youngest infants' fixations suggested that they were distributed far more broadly over the movie than the fixations of older observers. One important goal of the current study is to quantify this observation and examines its developmental time-course. Despite the schematic, cartoon nature of the faces in *Charlie Brown*, our stimulus provides a visual and linguistic environment that is rich in social content, enjoyable and motivating to our infant participants, and far more complex than those used in previous face perception experiments. Thus, infants' preferences in viewing this stimulus will give some insight into their attention to faces in real-world scenes. Of course, no laboratory task perfectly captures the structure of real experience, and we will discuss how the details of our stimulus (a cartoon movie on a small screen) limit our ability to generalize these results.

Methods

Participants

The final sample comprised 14 3-month-olds (*M* age = 81 days), 14 6-month-olds (*M* age = 185 days), and 12 9-month-olds (*M* age = 270 days), and two control groups of adults, one in the same condition as the infants ($n = 16$) and one in an acuity control condition ($n = 16$) described subsequently. Data from participants in a previous study (Johnson et al., 2008) who met inclusion criteria for this study were included in the 3-month-old group.

Inclusion criteria

Infants were included in the final sample if their point of gaze (POG) was recorded for at least 50% of the movies and if their average point-of-gaze across the 24 viewings of the central fixation point was less than 2 degrees from the center (the fixation point had a radius of approximately one degree). Fourteen additional 3-month-olds, 9 6-month-olds, and 14 9-month-olds completed the study but were excluded on the basis of these restrictions. We smoothed eye-movement data using a bilateral filtering algorithm (Durand & Dorsey, 2002).

This algorithm eliminates jitter in fixation (in some cases caused by inconsistent binocular calibrations, primarily in the 3-month-old group) while preserving saccades. We also interpolated across losses of point-of-gaze lasting less than 100 ms with cubic splines (polynomial smoothing functions). Results without smoothing and interpolation did not differ qualitatively from the results reported here.

Stimulus and presentation

Our stimulus was a 4-minute section of *A Charlie Brown Christmas* (Melendez, 1965), broken into 24 separate 10-s segments, each consisting of a movie (4 s), a central fixation point (2 s), and a pair of random-dot kinematograms (4 s) included for an unrelated experiment. Here we report only the responses to the cartoon; responses to the kinematograms are reported elsewhere (S.P. Johnson et al., 2008). Stimuli were presented on a 43 cm flat panel monitor. At the viewing distance of 60 cm, the 18.5 × 12.4 cm stimulus subtended 17.5 × 11.8 degrees visual angle. The soundtrack from the movie played throughout, creating a sense of continuity. Central-fixation points were dynamic, infant-friendly toys accompanied by characteristic sounds. Acuity control stimuli were constructed by blurring the movies with a Gaussian filter (with standard deviation of approximately 0.5 degrees) to simulate the acuity limit and contrast sensitivity function of a 3-month old (Norcia, Tyler, & Hamer, 1990). A Tobii ET-17 binocular corneal-reflection eye-tracker was used for stimulus presentation and data collection.

Results

Our dataset consisted of the record of each participant's point of gaze as he or she viewed each movie. Our first analysis assessed changes in the proportion of time each group spent looking at the faces of the characters in the movies (Figure 2). We conducted a one-way ANOVA with mean proportion looking to faces as the dependent variable and age-group as the factor. We found that changes in attention to faces across development were highly significant ($F(3,52) = 46.61$, $p < .0001$), with all pair-wise differences significant as well (all ps $< .01$ in Tukey post-hoc tests).

A particularly striking feature of these data was the wide distribution of 3-month-olds' fixations. Three-month-olds did not look at faces as much as other groups, but viewing the replays of their gaze did not reveal another obvious focus of interest such as bodies or hands.[1] We applied two analysis techniques to further characterize the changes in distribution of gaze across groups. Each is adapted from accepted statistical methods, but to our knowledge neither has previously been applied in precisely this form to eye-movement records, and so we discuss each analysis in some detail.

Entropy analysis

In our first analysis, we characterized the distribution of each group's fixations via a kernel density estimate which we obtained by smoothing the fixations of each age group on each movie (Hastie, Tibshirani, & Friedman, 2001). This technique produced a three-dimensional probability distribution reflecting the probability of fixation at each X- and Y-value over time (Figure 3 shows two-dimensional versions of these distributions produced by averaging across time). We then used Shannon entropy to characterize the predictability of these distributions (MacKay, 2003). Entropy is measured in bits of information and all our analyses are conducted over the entropy (in bits) of the distribution of fixations. Higher entropy (a larger number of bits) reflects more uncertainty about samples from a distribution; in the context of a distribution

[1]Sample videos of the kernel density estimates that we used for three-month-olds and adults in the entropy analysis, the blurred stimulus from the adult control, and the face and salience maps we used in the salience map analysis can all be found online at http://tedlab.mit.edu/~mcfrank/papers.html.

over eye-movements, higher entropy can be interpreted as greater uncertainty in predicting where new individuals in a group would look. This method provides a principled statistical method for characterizing the spread of each group's fixations while taking account of the temporal structure of the eye-movement data (see Appendix A for more details).

As Figure 3 reveals, we saw a pronounced narrowing in the spread of fixations across age, reflecting a tighter distribution of attention. Reflecting this narrowing, we found decreases in the average number of bits needed to characterize the distribution of attention; a one-way ANOVA with bits of entropy on each movie as the dependent measure and age group as the factor was statistically significant ($F(3,96) = 39.79$, $p < .0001$)[2]. Means and standard deviations are shown in Table 1. Tukey post-hoc tests between age groups showed statistically significant decreases in entropy between 3 and 6 months ($t(48) = -2.29$, $p < .05$) and between 9-month-olds and adults ($t(48) = -2.53$, $p < .05$). This decrease in entropy reflects the greater consistency of fixation targets across individuals in the older infant and adult groups.

Salience map analysis

In our second analysis we investigated the causes of the decrease in entropy we observed by identifying the particular image properties that drew the attention of our participants. We examined two possible hypotheses: first, that infants' fixations would be drawn primarily by faces, and second, that infants' fixations would be directed primarily to low-level perceptual salience (the attractiveness of basic perceptual features such as color, luminance, and motion). We defined two predictive models based on the contents of the movies that participants saw (Figure 1): the face model assumed that fixations would be directed primarily to movie regions that contained faces, whereas the perceptual salience model (Itti & Koch, 2001;Koch & Ullman, 1985) assumed that fixations would be directed to image regions proportional to their salience. Faces were hand-coded[3], while salience was computed as a linear combination of motion (temporal luminance contrast) and spatial luminance contrast (usually from edges such as object or surface boundaries in the image). Intuitively, this analysis allowed us to measure the extent to which perceptual saliency and social semantic content predicted eye movements for each group. (For more details of this analysis, see Appendix B.)

We first asked whether there were significant changes in model fit (how well the models predicted behavior) across the three infant groups and the adults. To quantify the predictive success of these two models, we calculated the likelihood of each infant's fixations on a particular movie under each model. The probability of a sequence of independent fixations is expressed as the product of the probabilities of each individual fixation. However, simply multiplying these probabilities would make the data of infants with more fixations (because of fewer blinks or losses of track) less probable than the data of infants with fewer fixations and more lost data, simply because the product of more probabilities will be on average smaller. To avoid this bias, we took the mean of each infant's fixation likelihoods for each clip. The arithmetic mean would under-weight low probability events, therefore we computed the geometric mean of infants' fixation likelihoods within movie clips.

These mean likelihoods of fixation are plotted in Figure 4. Because these model fits were normally distributed for each subject across movies (and hence not subject to the same problem with low-probability events as individual fixations), we calculated the arithmetic mean model

[2]All statistical tests in the entropy analysis are conducted on group data for each movie clip (as in a "by items" analysis), since density estimates are computed across all infants in a group at once, rather than for each infant (producing one entropy value for each group on each movie).

[3]Because of the extensive anthropomorphization of Snoopy (the animated dog that appears in the upper-left panel of Figure 1)—he uses tools, reads a script, and stands upright—and because his face is a target of fixations across all age groups, we included Snoopy's face in the face model. However, in the section on control analyses below, we consider only the movie clips which do not include him; this analysis confirms that the effects we report are not driven by fixations on Snoopy's face over human faces.

fit for each participant by averaging across the different movie clips. We then computed an ANOVA over these data. We used mean likelihood for each participant as the dependent measure, with age (3 months/6 months/9 months/adults) and model (faces/salience) as factors. We found main effects of both age ($F(3,104) = 49.23$, $p < .0001$) and model ($F(1,104) = 26.56$, $p < .0001$), as well as a significant interaction of the two ($F(3,104) = 16.38$, $p < .0001$), These results were similar when the same analysis was conducted using mean likelihood for each movie clip, averaging across participants (as in a conventional "by items" analysis, rather than a "by participants" analysis). Likelihoods for both models increased across development (likely reflecting the decrease in entropy—hence an increase in predictability—that we observed across development). To follow up this analysis we conducted planned paired *t*-tests, comparing the fit of the two models for each age group. At 3 months, fixations were best predicted not by the locations of faces but by the perceptual salience of various regions of the image ($t(13) = -3.23$, $p = .006$); at 6 months neither model predicted better than the other ($t(13) = 1.37$, $p = .19$); and at 9 months the face model predicted better than the low-level salience model ($t(11) = 2.69$, $p = .02$). Thus, younger infants' fixations (though predicted less well by any model than older infants' fixations) were better predicted by low-level image features, while older infants' fixations were best predicted by the location of faces.

Control Analyses

Because of the complexity of the data we gathered, as well as the rich but relatively uncontrolled nature of our stimuli, there are a number of possible confounds to the results reported above. In the following sub-sections we discuss subsidiary analyses which examine these potential confounds, including controls for the acuity of younger infants, random fixation by younger infants, irrelevant cinematic features, and the short length of the movie clips.

Did changes in visual acuity cause the developmental changes we observed?

Results from the adult control with blurred movies suggested that the development of visual acuity does not account for the interaction of age and salience model we observed. Eye movements of adults who viewed the *Charlie Brown* clips blurred to match the contrast sensitivity of a 3-month-old were equally well predicted by the face model as those of adults who saw the unmodified movies ($t(30) = .63$, $p = .54$). In addition, there was no difference in entropy of fixations between adults in the blurred and un-blurred conditions ($t(48) = -.75$, $p = .46$) and no difference in dwell time on faces ($t(30) = 1.50$, $p = .15$). These results suggest that low acuity was not alone responsible for the greater spread of younger infants' fixations.

Do younger infants fixate more randomly?

To better understand the fit of our models to the youngest infants' fixations, we conducted an additional analysis. Roughly speaking, there are two major differences between the salience model and the face model: first, the salience model predicts eye-movements to *different* targets than the face model, and second, it predicts a *broader spread* of fixations (because salient targets are distributed more broadly than faces). We were concerned that 3-month-olds might be less predictable on account of being more random; thus the better fit of the low-level model to their data might be driven exclusively by the models' breadth, with some random fixations obscuring otherwise face-directed looking.

Thus, we need to test directly our claim that the saliency model better predicted 3-month-olds' fixations because it predicted greater attention to different targets than the face model—targets identified by their salience, not simply by virtue of their not being faces. To test this, we split the video clips into two different categories: those where the face and salience models predicted looking to relatively similar regions of the movie, and those movies where the two models predicted looking to relatively dissimilar regions of the movie. We split the movies in this way

using the Kullback-Leibler (KL) divergence between the two models (MacKay, 2003). KL divergence is an information-theoretic measure related to entropy (and similarly measured in bits of information). The KL divergence between two distributions measures the differences between them by quantifying the amount of uncertainty in one distribution given the other.

We used the KL divergence measure to ask how different the face and salience models were for each video. A high divergence between the two models for a particular movie clip indicated that there were many perceptually salient regions of the movie other than faces. In contrast, a low divergence between the two models indicated that faces were among the most salient stimuli. In low-divergence clips, the predictions of the two models were confounded—faces were the most salient stimuli. Therefore fixations on faces could be driven by perceptual salience and fixations on salient regions could be driven by the fact that the salient regions were faces; in this situation there is no way to distinguish the contributions of salience and faces. In the high-divergence movies, however, the two models make different predictions: non-face regions are most salient, so these clips are much more diagnostic of whether fixations are being directed at faces, or salient regions. Using this logic, we could assess whether the saliency model better predicted 3-month-olds' fixations because it predicted that attention would be drawn to *different* regions of the movies than faces.

We examined the fit of the two models to 3-month-olds' fixations on both the high and low divergence movies (Figure 5) using an ANOVA with infants' mean likelihood of fixation as dependent variable and divergence (high/low) and model (face/salience) as factors. We found significant effects both of model ($F(1,52) = 12.76$, $p < .001$) and divergence between models ($F(1,52) = 17.46$, $p = .0001$) and a trend towards an interaction of the two ($F(1,52) = 3.16$, $p = .08$). This results suggests that infants were not simply looking at faces like adults with the addition of random fixation noise – if they were, the breadth of the salience model would account for random fixations equally well in both sets of movie clips (and there should be no main effect of divergence and no trend towards an interaction). Instead, fixations in movies where the non-face regions are more salient were better predicted by the perceptual salience model than by the face model. We conclude, therefore, that the eye-movements of 3-month-olds were being driven by salience, rather than being solely the product of a highly noisy face preference.

Did irrelevant cinematic features of the cartoons cause the differences we observed?

Another possible explanation for the scattered fixations of younger infants invokes the specific properties of our cartoon stimuli. Perhaps the presence of moving backgrounds, cuts, camera movements, and Snoopy (the animated dog present in some clips) confused 3- and 6-month-olds, leading to the greater scatter in their fixations. To test this hypothesis, we identified 10 of the 25 movie clips in our study that contained no moving backgrounds, cuts, camera movements, or animals. We conducted an age by model ANOVA individually for each of these 10 clips, again using mean likelihood of fixation as dependent variable and including age (3, 6, and 9 months) and model (faces/salience) as factors. The same trend reported in our original salience map analysis was present in each of these 10 movies; the face model was more predictive of 9-month-olds' attention whereas the salience model was more predictive of 3-month-olds' attention. The presence of the observed trend (as opposed to its reverse) occurring by chance in all 10 clips is highly unlikely (sign test $p = .002$). Even given the very limited data-set used in this control analysis (analyzing fixations on each video clip individually), the crossover interaction we observed reached significance in 5 of the 10 video clips. Thus, even when viewing animated faces talking on a static background, therefore, 3-month-olds' fixations were guided less by faces than those of 9-month-olds.

Did the short length of our clips cause developmental differences in fixation?

Our movie clips were each 4 s in duration. Perhaps younger infants were simply unable to orient to faces as quickly as older infants. To test this explanation, we conducted a split-half analysis by testing our models separately on the first and second halves of each clip. We introduced order as a factor into our analyses, using a 3-way group (3-/6-/9-month-old/adult) × model (faces/salience) × half (first/second half) ANOVA. We found main effects of all three factors ($F(3,192) = 16.92$, $p < .0001$; $F(1,192) = 116.50$, $p < .0001$; $F(1, 192) = 71.45$, $p < .0001$), as well as significant two-way interactions of group and model as reported in the first analysis ($F[3,192] = 8.02$, $p < .0001$) and a significant two-way interaction of model and order ($F(1,192) = 48.60$, $p < .0001$), but no interaction of group and order ($F(3,192) = 0.01$, $p = .99$) and no evidence of a three-way interaction of group, model, and order ($F(3,192) = 0.48$, $p = 69$). The results of this analysis suggest that all groups look more at faces during the second half of the clip. However, the lack of a group by order interaction or a three-way interaction strongly suggest that the length of the clips did not differentially affect younger infants (for example, by pushing them more towards non-face aspects of the clips during the first half).

General Discussion

Anecdotally, many parents remember the first day when their infant was interested in looking at their faces rather than the window-blinds or the ceiling fan (regions of high contrast and motion). While the impression of a sharp transition may be mistaken, our results suggest that there is some general truth to the impression of a shift between looking at perceptual salience and looking at faces. In the current study we found evidence of this shift through observation of the fixation patterns of infants and adults as they viewed clips from an animated movie: schematic yet highly engaging social stimuli. We observed a dramatic increase in looking toward faces across development. We probed the origins of this increase through two novel analyses: first, an analysis of the entropy of fixations which found that consistency of fixation targets across individuals increased over development; second, a salience map analysis which suggested that 3-month-olds' fixations were better predicted by a low-level model of image salience than by a model that predicted looking to faces and hence were likely scattered more broadly because they were attending to a range of salient objects throughout the scene. The developmental pattern we observed was present even in parts of the cartoon that had no cuts or camera motion and was not generated by changes in acuity across development or by the short length of the clips. Thus, whatever predispositions to look at faces that infants may bring with them into the world, these predispositions did not yield attention to faces in our stimulus as consistently in younger infants as in older ones.

Three developmental changes might contribute in various degrees to the pattern of results we observed: the development of a preference to look at faces as a social information source, the development of sensitivity to the intermodal coordination between faces and speech, and the development of attentional/inhibitory mechanisms allowing for the suppression of salient background stimuli in favor of faces. We hope that disentangling the relative contributions of these three factors will be a focus of further work.

The first possible factor in the greater looking to faces that we observed is a direct increase in infants' preference for looking at faces. Faces are highly relevant information sources for the kind of social inferences that children are beginning to make robustly during the latter part of their first year (e.g., Carpenter, Nagell, & Tomasello, 1998; Corkum & Moore, 1998). It may simply be the case that infants gradually become aware of the relevance of faces as a source of social information. This kind of developmental trend could arise either through the action of gradual reinforcement (e.g., Triesch et al., 2006), through some kind of biased learning (e.g., Morton & Johnson, 1991), or through a maturational process (e.g., Baron-Cohen, 1995). Thus

the pattern of data we observed could be attributed simply to infants' motivation to look at faces.

However, a second possible factor is the development of a greater sensitivity to intermodal regularities. Because our videos were accompanied by dialogue, perhaps older infants were better equipped to notice the correspondence between the dialogue and the moving mouths on the faces of the characters in the video clips, leading to an increase in looking at the faces. Certainly, intermodal redundancy is a very powerful cue for learning and can allow for the extraction of cross-modally salient stimulus regularities (Bahrick & Lickliter, 2000). In the specific case of the intermodal relationship between speech and faces, Dodd (1979) reported matching between voices and faces on the basis of lip/voice synchrony in 2-month-olds, and Kuhl & Meltzoff (1982) gave evidence that this synchrony relied on the spectral information present in vowels. Thus, some intermodal speech-to-face matching ability is present relatively early in development. However, further experimental work will be necessary to determine whether changes in the perception of intermodal correspondence between faces and speech in natural settings could account for our results.

The clearest test of intermodal synchrony as the dominant factor would be a replication of our findings using an uninformative, unsynchronized soundtrack. We conducted a control study with 16 adults by replacing the coordinated dialogue in our movie clips with unsynchronized classical music. This manipulation decreased looking to faces, but only to the level of 9-month-olds, suggesting that a semantically and temporally congruous audio-track accounts for the majority of the difference between 9-month-olds and adults, but that additional factors develop prior to 9 months of age. However, results from adult studies can provide at best indirect evidence; studies with infants are necessary to test the intermodal synchrony hypothesis directly.

A third possible factor in older infants' and adults' greater looking to faces might be developmental changes in the mechanisms governing attentional control. Under this hypothesis, all infants might share the same preference for faces, but differ in their abilities to orient to them consistently and to suppress the effects of distracting background information. Certainly, there is a large body of evidence suggesting major changes in these abilities across the first year (e.g., Amso & Johnson, 2005, in press; Butcher, Kalverboer, & Geuze, 2000; Johnson, Posner, & Rothbart, 1991), but it is not known if limitations in orienting can account or the developmental pattern we observed. Current experiments are testing this possibility.

Limitations of our stimuli

A common concern in interpreting the results of any experiment is that the stimuli may not be representative of the real world. This concern is especially justified in the current study, where we used short clips from an animated movie as stimuli. To what extent can we generalize our conclusions in this study to the behavior of infants outside the lab? Some aspects of the differences between animated movies and real visual experience might serve to increase the trends we observed. If anything, the real world is noisier, more cluttered, less centrally organized, and less face-dominated, than the schematic world of Charlie Brown. These factors might contribute to an even stronger bias in younger infants to look to non-face targets. On the other hand, the schematic nature of the faces in our cartoons might decrease young infants' ability to recognize them as faces. While this concern is plausible, we believe that the current literature does not support it. Schematic faces have repeatedly been shown to attract the attention of very young infants (Farroni et al., 2005; Morton & Johnson, 1991). Moreover, as infants mature, their face representations become more detailed (Cohen & Strauss, 1979; Haith, Bergman, & Moore, 1977): a mismatch between our schematic cartoon faces and the details of real-world faces would predict the opposite developmental trend from the one we observed —a decrease, rather than an increase, in looking to schematic faces with age. Thus, although

caution is warranted in interpreting the generality of our results, and future experiments will need to be conducted with more life-like stimuli, we believe that the current study is an important first step in measuring the looking behavior of infants in noisy, cluttered environments.

Methodological innovations

Although our goal in the current work was to investigate the development of infants' looking to faces, the methods we employed may have broad applications to other questions in development. Because all of our analyses rely on free-viewing with no explicit instructions, they are applicable to groups from early infancy to adulthood, avoiding some of the issues of changing preferences that characterize methods like habituation and violation of expectation (Hunter & Ames, 1988; Aslin, 2007). In addition, our use of eye-tracking combined with multiple discrete movie clips allows us to make more precise measurements of individual infants by evaluating their performance across multiple trials. Though much future work will be needed to validate our measures in individual infants, the type of design we employed holds the promise of creating measures that are reliable estimates of preferences for individuals. In turn, reliable measures of individual infants' preferences could be used in studies that use individual differences to test detailed developmental claims (see Johnson et al., 2008, for a recent example) as well as opening the door to clinical applications that require high precision in individual patients.

Conclusions

In early infancy, a weak bias for faces may suffice to spur learning about conspecifics across a variety of real world contexts. This same weak bias—whatever its origins—is likely to account for infant performance in face perception experiments employing face stimuli in isolation. But our results suggest that this initial bias, whether domain-general or face-specific, is only a small part of the story. To understand how infants come to appreciate the importance of faces, future work must focus not only on the origins of early face preferences but also on the mechanisms underlying the later development of attention to faces.

Acknowledgements

The authors wish to thank Nancy Kanwisher, Mary Potter, Joshua Tenenbaum, Jennifer Yoon, and three anonymous reviewers for their helpful comments. We are especially grateful to the infant participants and their parents. The research was funded by grants from the National Science Foundation (BCS-0418103), the National Institutes of Health (R01-HD40432 and R01-HD48733) and the McDonnell Foundation (412478-G/5-29333). The first author was supported by a Jacob Javits Graduate Fellowship.

References

Cassia VM, Turati C, Simion F. Can a Nonspecific Bias Toward Top-Heavy Patterns Explain Newborns' Face Preference? Psychological Science 2004;15:379. [PubMed: 15147490]

Cohn JF, Tronick EZ. Three-Month-Old Infants' Reaction to Simulated Maternal Depression. Child Development 1983;54:185–193. [PubMed: 6831986]

Corkum V, Moore C. The origins of joint visual attention in infants. Developmental Psychology 1998;34:28–38. [PubMed: 9471002]

de Haan M, Johnson MH, Maurer D, Perrett DI. Recognition of individual faces and average face prototypes by 1-and 3-month-old infants. Cognitive Development 2001;16:659–678.

de Haan M, Pascalis O, Johnson MH. Specialization of Neural Mechanisms Underlying Face Recognition in Human Infants. Journal of Cognitive Neuroscience 2002;14:199–209. [PubMed: 11970786]

Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. ACM Transactions on Graphics (TOG) 2002;21:257–266.

Farroni T, Johnson MH, Menon E, Zulian L, Faraguna D, Csibra G. Newborns' preference for face-relevant stimuli: Effects of contrast polarity. Proceedings of the National Academy of Sciences (USA) 2005;102:17245–17250.

Haith MM. Who put the cog in infant cognition? Is rich interpretation too costly? Infant Behavior and Development 1998;21:167–179.

Halit H, Csibra G, Volein A, Johnson MH. Face-sensitive cortical processing in early infancy. Journal of Child Psychology and Psychiatry 2004;45:1228–1234. [PubMed: 15335343]

Halit H, de Haan M, Johnson MH. Cortical specialisation for face processing: face-sensitive event-related potential components in 3-and 12-month-old infants. Neuroimage 2003;19:1180–1193. [PubMed: 12880843]

Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction. Springer; 2001.

Hershler O, Hochstein S. At first sight: A high-level pop out effect for faces. Vision Research 2005;45:1707–1724. [PubMed: 15792845]

Itti L, Koch C. Computational modeling of visual attention. Nature Reviews Neuroscience 2001;2:194–203.

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

Johnson SP, Davidow J, Hall-Haro C, Frank MC. Perceptual completion originates in information acquisition. Developmental Psychology 2008;44:1214–1224. [PubMed: 18793055]

Kelly DJ, Quinn PC, Slater AM, Lee K, Ge L, Pascalis O. The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. Psychological Science 2007;18:1084–1089. [PubMed: 18031416]

Kelly DJ, Quinn PC, Slater AM, Lee K, Gibson A, Smith M, et al. Three-month-olds, but not newborns, prefer own-race faces. Developmental Science 2005;8:31–36.

Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 1985;4:219–227. [PubMed: 3836989]

Kuhl, PK. A new view of language acquisition. Vol. 97. National Acad Sciences; 2000. p. 11850-11857.

Kuhl PK, Tsao FM, Liu HM. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. Proceedings of the National Academy of Sciences 2003;100:9096–9101.

Lewis MB, Edmonds AJ. Face detection: Mapping human performance. Perception 2003;32:903–920. [PubMed: 14580138]

Lewis MB, Edmonds AJ. Searching for faces in scrambled scenes. Visual Cognition 2005;12:1309–1336.

MacLeod DIA, Boynton RM. Chromaticity diagram showing cone excitation by stimuli of equal luminance. J Opt Soc Am 1979;69:1183–1186. [PubMed: 490231]

Mayer DL, Dobson V. Visual acuity development in infants and young children, as assessed by operant preferential looking. Vision Res 1982;22:1141–1151. [PubMed: 7147725]

Melendez, B. A Charlie Brown Christmas. Melendez, B.; Mendelson, L., editors. USA: CBS; 1965.

Morton J, Johnson MH. CONSPEC and CONLERN: A Two-Process Theory of Infant Face Recognition. Psychological Review 1991;98:164–181. [PubMed: 2047512]

Nelson CA. The development and neural bases of face recognition. Infant and Child Development 2001;10:3–18.

Norcia AM, Tyler CW, Hamer RD. Development of contrast sensitivity in the human infant. Vision Research 1990;30:1475–1486. [PubMed: 2247957]

Pascalis O, de Haan M, Nelson CA. Is Face Processing Species-Specific During the First Year of Life? Science 2002;296:1321–1323. [PubMed: 12016317]

Pascalis O, De Haan M, Nelson CA, De Schonen S. Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. Journal of Experimental Psychology: Learning, Memory, and Cognition 1998;24:249–260.

Pascalis O, Scott LS, Kelly DJ, Shannon RW, Nicholson E, Coleman M, et al. Plasticity of face processing in infancy. Proceedings of the National Academy of Sciences 2005;102:5297–5300.

Peterson MS, Kramer AF, Irwin DE. Covert Shifts of Attention Precede Involuntary Eye Movements. Perception and Psychophysics 2004;66:398–405. [PubMed: 15283065]

Scaife M, Bruner JS. The capacity for joint visual attention in the infant. Nature 1975;253:265–266. [PubMed: 1113842]

Simion F, Macchi Cassia V, Turati C, Valenza E. The origins of face perception: specific versus non-specific mechanisms. Infant and Child Development 2001;10:59–65.

Sugita Y. Face perception in monkeys reared with no exposure to faces. Proceedings of the National Academy of Sciences 2008;105:394.

Tronick EZ. Emotions and emotional communication in infants. American Psychologist 1989;44:112–119. [PubMed: 2653124]

VanRullen R. On second glance: Still no high-level pop-out effect for faces. Vision Research 2006;46:3017–3027. [PubMed: 16125749]

Appendix A: Entropy analysis details

To measure the spread of infants' fixations, we plotted the fixations of infants from each group as points within a three-dimensional probability distribution (X position, Y position, and time). We then smoothed this distribution by convolving it with a Gaussian kernel (Hastie, Tibshirani, & Friedman, 2001) which spread probability mass around discrete fixation points to reflect both our uncertainty about the exact eye position, as well as the assumption of a continuous distribution of attention in the region surrounding where fixations were observed. The kernel we chose was isotropic in space with a standard deviation of 100 pixels but truncated in time to extend only into the past with a standard deviation of 165 ms (reflecting that fixations may only be driven by movie content preceding the fixation, and that the focus of covert attention shifts to a particular spatial location before eye-movements are directed there (Peterson, Kramer, & Irwin, 2004). We then calculated the Shannon entropy (H) of this distribution (which we notate *A*) by

$$H(A) = -\sum_{x,y,t} p(A_{x,y,t}) \log_2 p(A_{x,y,t})$$

where $p(A_{x,y,t})$ reflects the smoothed estimate of the probability of a fixation at a point in *A* given by *x*, *y*, and *t*.

Appendix B: Salience map analysis details

We constructed the face-based model by hand-coding the bounding ellipse of the faces present in each frame of the movies. We then smoothed the resulting frame maps with a Gaussian kernel (ensuring that fixations immediately outside the face itself were given some probability of having been on the face but misdirected by tracker error). We additionally added a small uniform probability (1% of total) to the face map to give fixations outside of the face nonzero probabilities.

To construct the salience model, we first computed spatial contrast maps for luminance, orientation, and color, a motion/flicker map (temporal luminance contrast), and a color saturation map (for similar approaches to computing perceptual saliency from images, see Itti, 2006). All spatial contrast maps were computed by convolving difference-of-Gaussians (Enroth-Cugell & Robson, 1966) filters with feature maps. The scale (standard deviations) of the positive and negative Gaussians were roughly 0.5 and 0.3 degrees. For luminance contrast, the feature map was the sum of the three color channels. For orientation contrast, the feature maps were differences in orientation energy (difference of output between full-wave rectified orthogonal gabors). Two such orientation energy maps were defined: 0–90 degrees and 45–

135 degrees, then the spatial contrast from each of these maps was summed. This orientation contrast is a specific type of second-order contrast which often defines surface boundaries (Sutter, Sperling, & Chubb, 1995), and thus may plausibly drive attention. Color contrast was calculated by converting RGB values from the movies into a physiologically-based color space (Yrb, MacLeod & Boynton, 1979), and then computing spatial contrast individually for the r and b dimensions for RG-contrast and BY-contrast, respectively. Motion and flicker were approximated as temporal luminance contrast (the square of frame-to-frame pixel-wise luminance change). Saturation maps were created by transforming color values to a hue-saturation-luminance space and isolating the saturation value. It should be noted that each of these analyses was applied across the entire frame, regardless of whether a particular pixel was in a face or not—thus, salient features predict some looking to faces (since faces usually differ from the background), simply less than the face model.

Each of these features was computed over each frame of each Charlie Brown movie to create a set of three-dimensional feature-based maps. We used each feature individually to predict fixations. Across groups, motion and luminance contrast were most predictive, with orientation highly confounded with luminance and the three color terms less predictive (though of the three, red-green contrast was best). Accordingly, we created a single salience map by summing luminance and motion information with equal weights. This saliency map was converted into a prediction of the distribution of attention by assuming that people would fixate regions of space proportionally to how salient they were relative to everything else in the scene; thus we simply normalized the saliency maps to integrate to 1 at each frame to generate model predictions. We added the same small uniform noise term as in the face model.

We also created a chance model (shown in Figure 4) by assuming that infants' fixations would be uniform across the movie; we experimented with another chance model that preferred central fixations but found only minor differences in prediction.

In order to evaluate the fit of the models to the fixations of different groups, we computed the probability of fixation at a particular location for each recorded timestep by each participant for each model. We then took the geometric mean across time in each clip to create a likelihood value for each participant in each movie; we used these values as the basis of our further analyses. This analysis can be thought of as a maximum-likelihood comparison of models; or equivalently, as Bayesian model selection (Gelman et al., 2004) with a uniform prior.

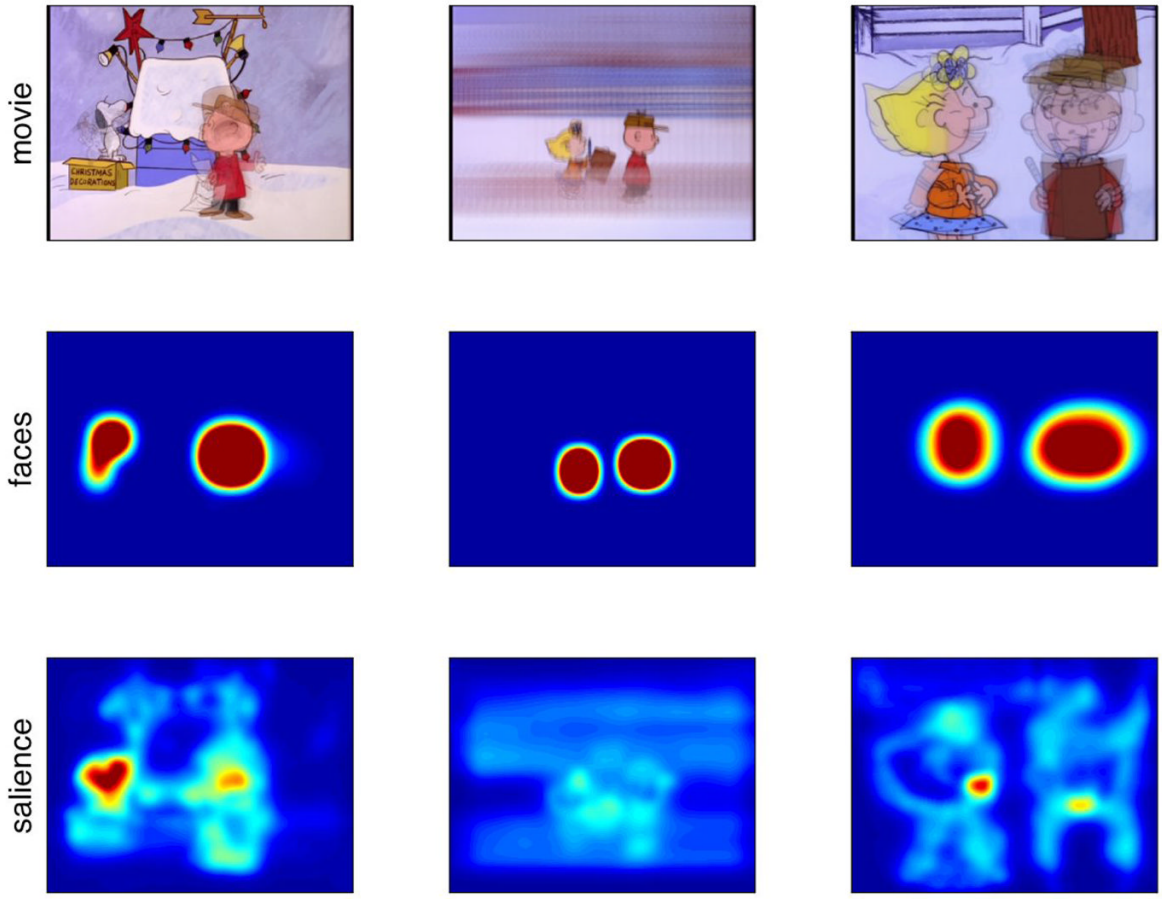

Figure 1.
Example stimuli and models (averaged across time) for three different 4-s clips from *A Charlie Brown Christmas*. The first row depicts time-averaged stimuli. The second row shows the assignment of predictive probability in the face model for each of these movie clips. The third row shows the assignment of probability for the low-level salience model. Warmer colors indicate higher probability; probabilities are scaled equally across all images.

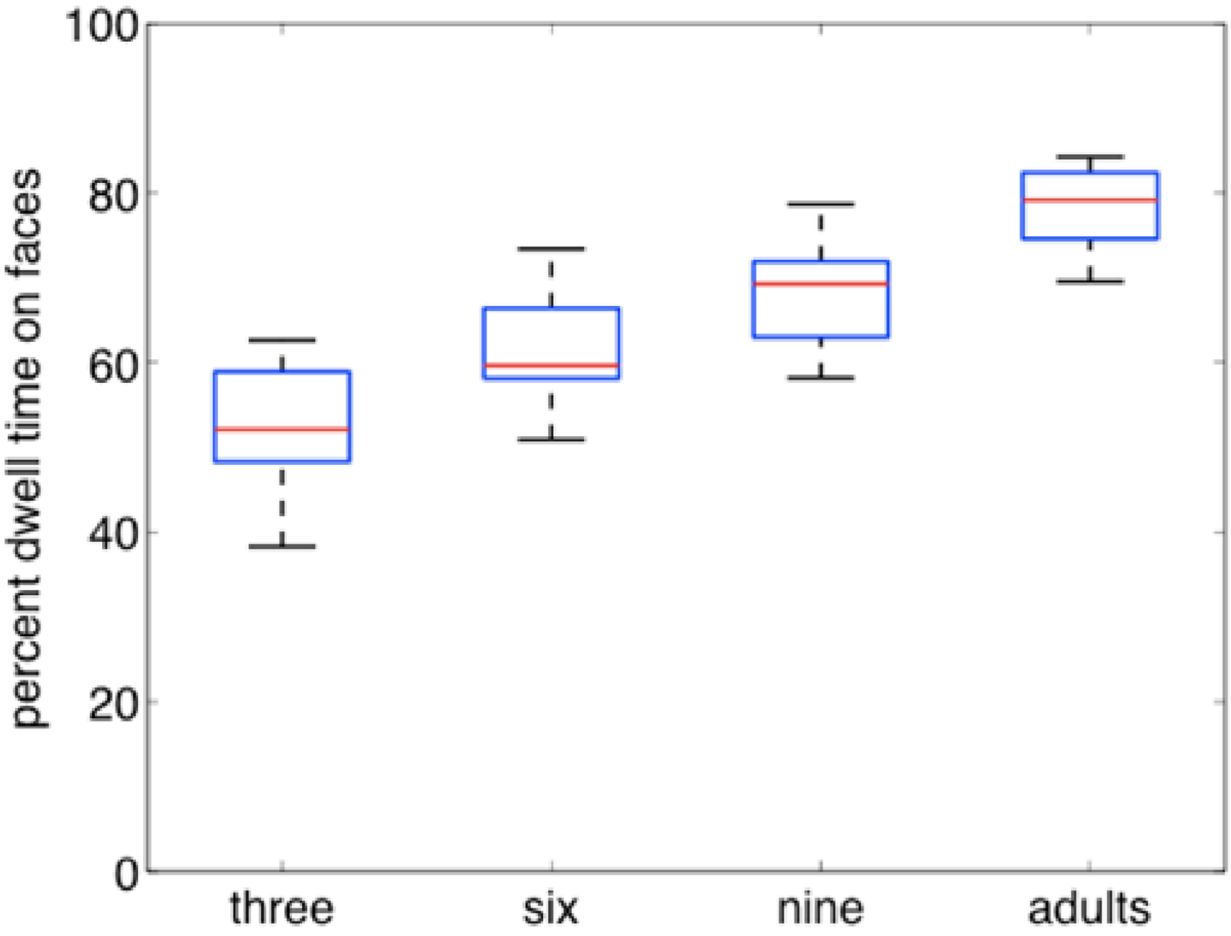

Figure 2.
Box-and-whisker plot of percent dwell time to faces by group. Boxes represent median, lower and upper quartiles; whiskers represent minimum and maximum values.

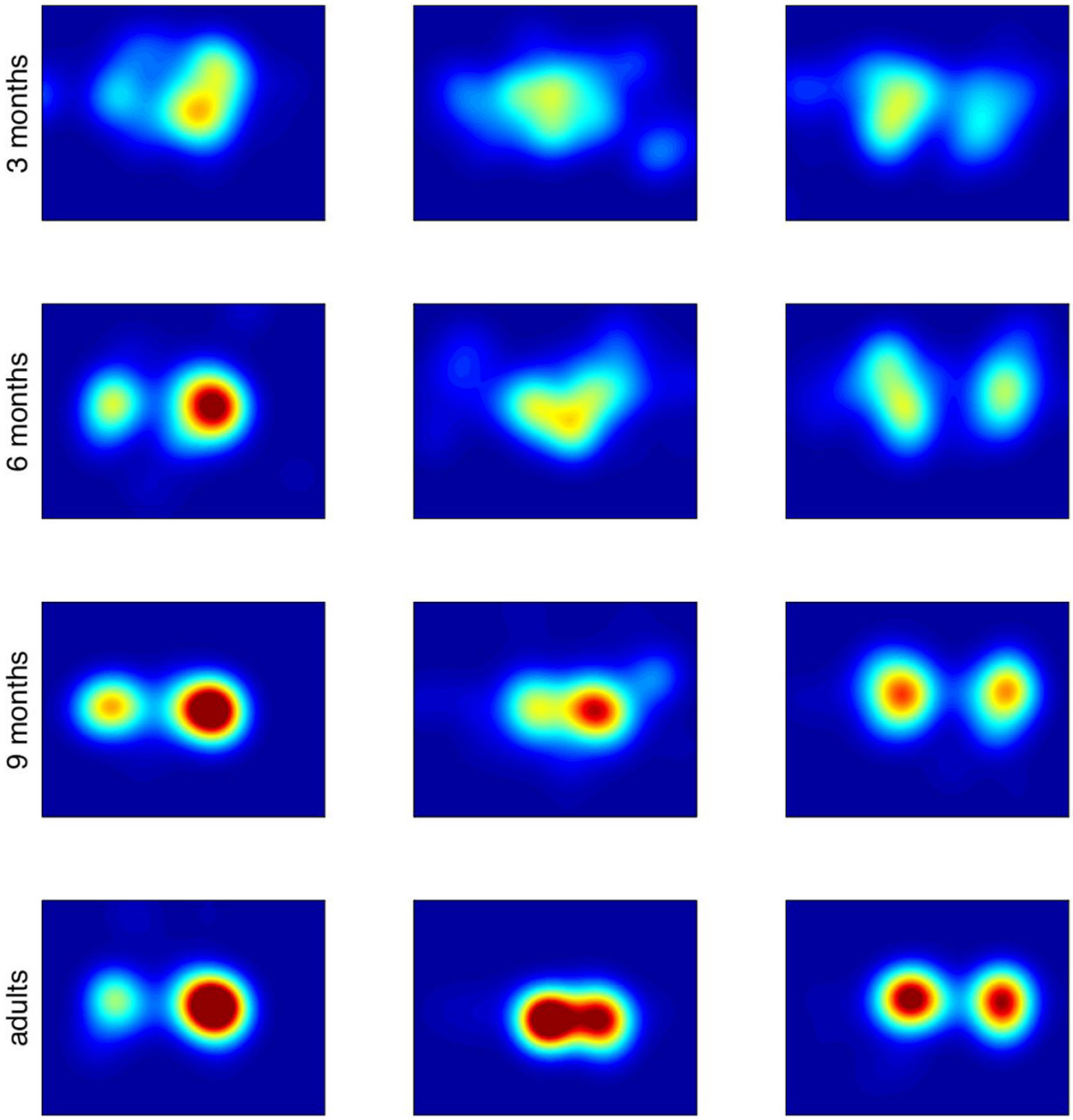

Figure 3.
Example kernel density estimates (averaged across time) corresponding to the clips shown in Figure 1. Rows are data from (top to bottom) 3-month-olds, 6-month-olds, 9-month-olds, and adults. Warmer colors indicate higher probability; probabilities are scaled equally across all images.

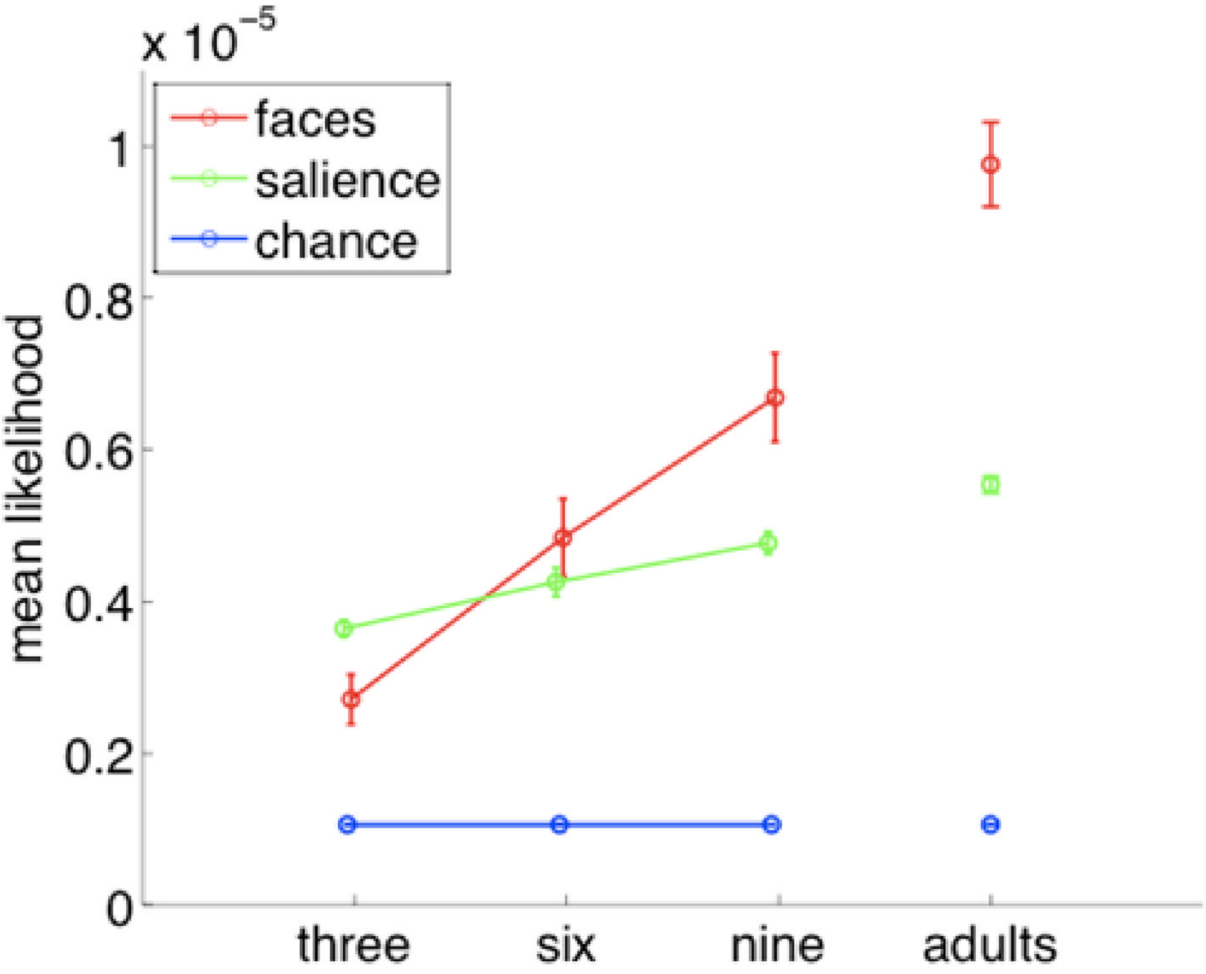

Figure 4.

Fit of face, low-level salience, and chance models to experimental data. Higher values indicate greater mean fixation probability for infants of that group. Error bars indicate standard error of the mean across participants. Because the likelihood of an infant fixating any given pixel is *a priori* very small, these probabilities are very low in absolute terms. Nonetheless, when examined relative to the two different models of fixation, they show reliable changes across groups.

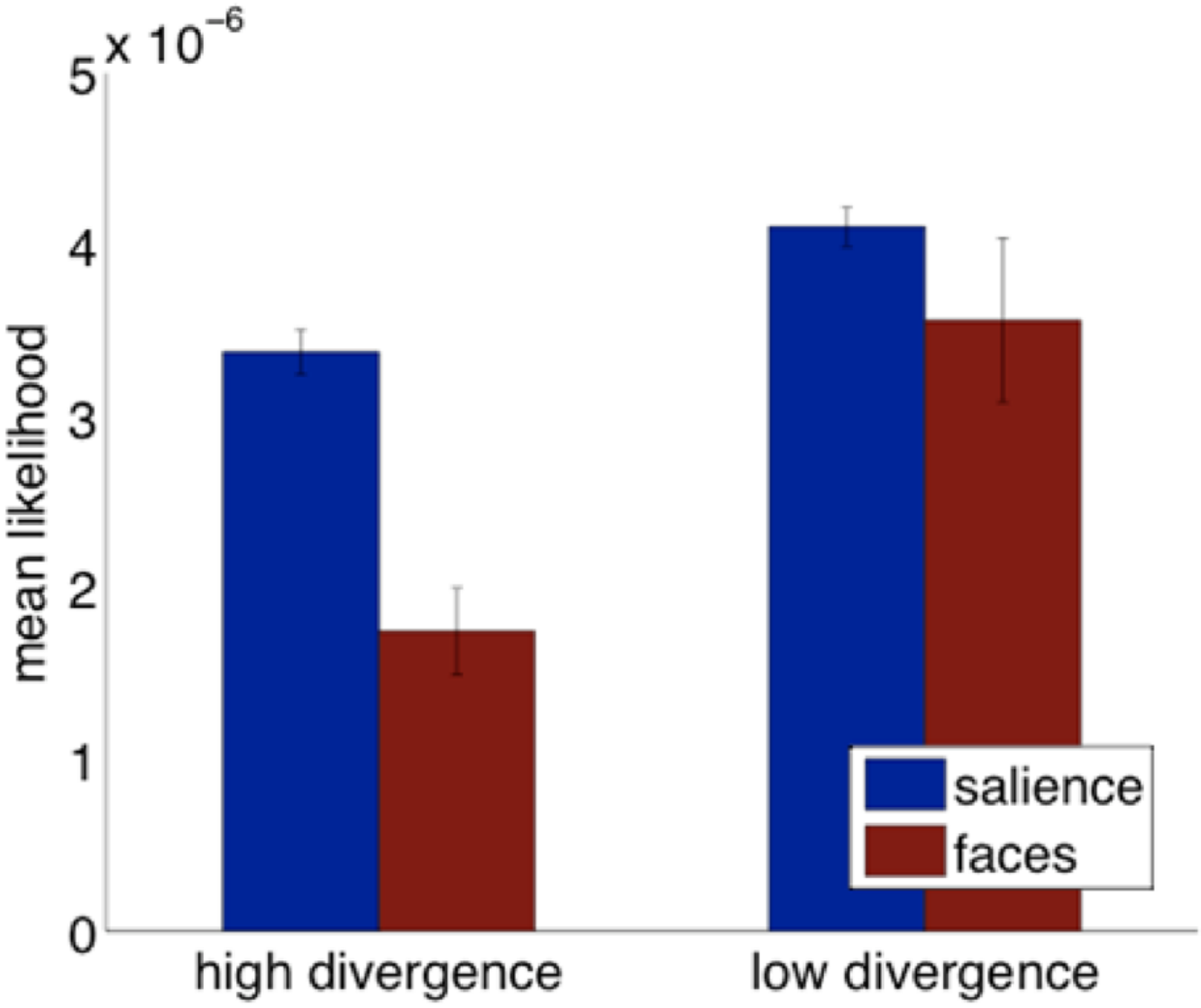

Figure 5.
Mean likelihood for three-month-olds under face and salience models across high and low KL-divergence movies.

Table 1

Mean entropy (standard deviation) across movies and age groups.

3-month-olds	6-month-olds	9-month-olds	adults
15.12 (.24)	14.84 (.30)	14.68 (.25)	14.32 (.27)