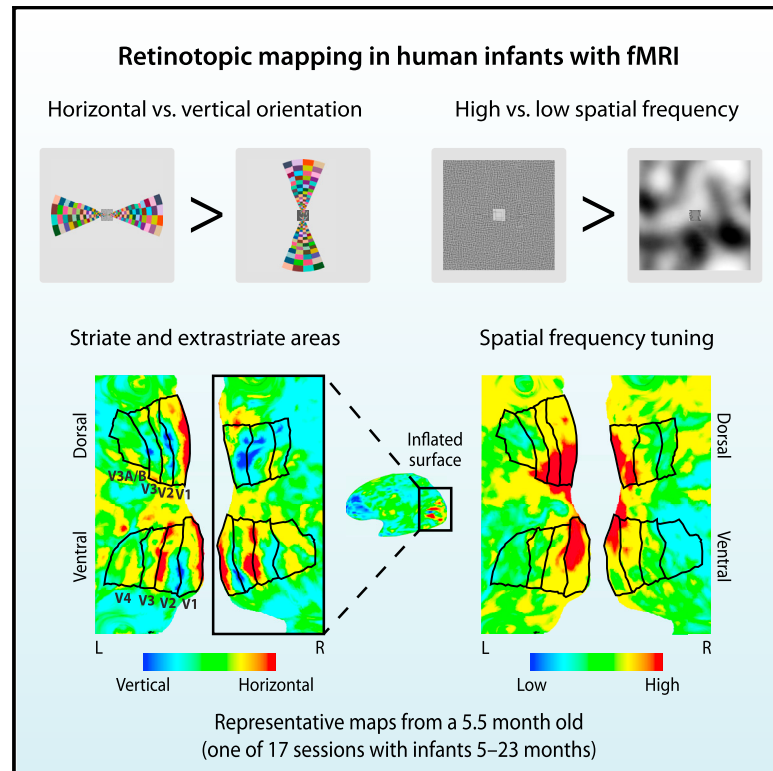


Retinotopic organization of visual cortex in human infants

Graphical abstract



Authors

Cameron T. Ellis, Tristan S. Yates,
Lena J. Skalaban, Vikranth R. Bejjanki,
Michael J. Arcaro,
Nicholas B. Turk-Browne

Correspondence

cameron.ellis@yale.edu

In brief

Ellis et al. use fMRI to study how the visual system is organized in human infants. They find evidence for multiple visual maps throughout visual cortex, indicating an earlier maturity of extrastriate cortex than previously appreciated. This nascent organization may be critical for scaffolding subsequent developmental changes in perception.

Highlights

- Retinotopic organization of striate and extrastriate visual areas in infancy
- Spatial frequency sensitivity varies within and across infant visual areas
- Weak developmental change from 5 to 23 months, suggesting early maturation
- Possible to identify and interrogate functional regions of interest in human infants



Article

Retinotopic organization of visual cortex in human infants

Cameron T. Ellis,^{1,5,*} Tristan S. Yates,¹ Lena J. Skalaban,¹ Vikranth R. Bejjanki,² Michael J. Arcaro,³ and Nicholas B. Turk-Browne^{1,4}

¹Department of Psychology, Yale University, New Haven, CT 06511, USA

²Department of Psychology, Hamilton College, Clinton, NY 13323, USA

³Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

⁴Wu Tsai Institute, Yale University, New Haven, CT 06511, USA

⁵Lead contact

*Correspondence: cameron.ellis@yale.edu

<https://doi.org/10.1016/j.neuron.2021.06.004>

SUMMARY

Vision develops rapidly during infancy, yet how visual cortex is organized during this period is unclear. In particular, it is unknown whether functional maps that organize the mature adult visual cortex are present in the infant striate and extrastriate cortex. Here, we test the functional maturity of infant visual cortex by performing retinotopic mapping with functional magnetic resonance imaging (fMRI). Infants aged 5–23 months had retinotopic maps, with alternating preferences for vertical and horizontal meridians indicating the boundaries of visual areas V1 to V4 and an orthogonal gradient of preferences from high to low spatial frequencies. The presence of multiple visual maps throughout visual cortex in infants indicates a greater maturity of extrastriate cortex than previously appreciated. The areas showed subtle age-related fine-tuning, suggesting that early maturation undergoes continued refinement. This early maturation of area boundaries and tuning may scaffold subsequent developmental changes.

INTRODUCTION

Vision is the dominant sense in humans but develops slowly throughout childhood and even into adolescence (Braddick and Atkinson, 2011; Lewis and Maurer, 2005; Kiorpes, 2016). How is infant visual cortex organized and how does this organization change over early development? Arguably the most fundamental organizing principle of the mature mammalian visual cortex is that it contains orderly, topographic representations of visual space (i.e., maps; Kaas, 1997; White and Fitzpatrick, 2007). A retinotopic map represents eye-centered spatial relationships in which adjacent neurons along the cortical surface receive input from adjacent points along the surface of the retina. The existence of multiple retinotopic maps indicates that the visual cortex is organized into areas (i.e., arealization). There is also a spatial frequency map, in which different visual areas and parts of areas respond to different frequencies of periodic stimulation on the retina. Given the centrality of these maps to our understanding of vision, it has long been considered a fundamental challenge to research on visual development that they have not yet been described in human infants (Braddick and Atkinson, 2011).

What we know about the development of visual maps comes from animal research. The organization of the striate visual cortex (V1) is established prenatally (Crowley and Katz, 1999; Farley

et al., 2007) by molecular signaling mechanisms (McLaughlin and O'Leary, 2005) guiding topography-preserving projections from the lateral geniculate nucleus (LGN) of the thalamus. Though studies have found mature, adult-like receptive field properties of single neurons in V1 and, to a lesser extent, secondary visual cortex (V2) (Chino et al., 1997; Kiorpes and Movshon, 2014; Zhang et al., 2005; Zheng et al., 2007), laminar organization (Bourne and Rosa, 2006) and population-level activity appears immature for the first several months, particularly in extrastriate cortex (Arcaro and Livingstone, 2017; Distler et al., 1996; Van Grootel et al., 2017). Indeed, there appears to be a sequence across the visual hierarchy, suggesting that striate areas mature before extrastriate areas (Bourne and Rosa, 2006; Condé et al., 1996; Distler et al., 1996; Gomez et al., 2018; Kiorpes, 2016; Zhang et al., 2005), referred to as hierarchical maturation. Immaturity in extrastriate areas is particularly critical for scaffolding behavioral changes in vision (Kiorpes and Movshon, 2014; Van Grootel et al., 2017).

However, the maturity of extrastriate cortex in human infants remains mysterious. Perhaps the strongest indication that the striate and extrastriate cortices are developed in human infants comes from macaques, where functional connectivity in visual cortex reflects boundaries between areas throughout the visual cortex (Arcaro and Livingstone, 2017). However, this evidence is indirect because it relied on functional connectivity rather



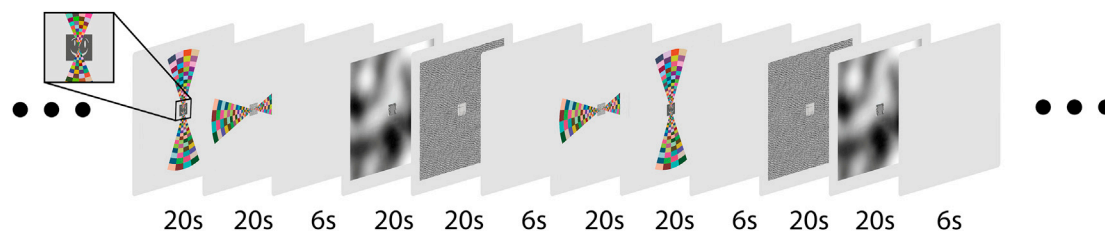


Figure 1. Experimental stimuli for retinotopic mapping

Participants viewed wide-field (40° visual angle) meridian or spatial frequency stimuli on the ceiling of the scanner bore in an alternating block design. Each block contained two phases in a counterbalanced order: vertical and horizontal for meridian mapping and high and low for spatial frequency mapping. The two phases lasted 20 s each and appeared back to back with no break, followed by 6 s of rest. A small, animated movie (1.5°), enlarged in the inset, of monochromatic shapes was played at center to encourage fixation.

than on more standard mapping stimuli. Moreover, the timeline of maturation is unclear because infancy is more protracted in humans than macaques (Boothe et al., 1985; Kiorpes, 2016). Even at maturity, there are similarities (Wandell et al., 2007) and differences (Angelucci and Rosa, 2015; Fize et al., 2003; Gattass et al., 2015; Lyon and Connolly, 2012; Orban et al., 2004) across primate species, particularly in extrastriate areas. Current evidence from human infants supports, albeit indirectly, the idea that extrastriate cortex is immature in infancy (Kovács, 2000; Siu and Murphy, 2018). Visual behaviors thought to rely primarily on V1, such as orientation discrimination and spatial frequency discrimination, are present in rudimentary form near birth (Baker et al., 2011; Banks et al., 1985; Braddick et al., 1986) and improve rapidly throughout the first year (Norcia et al., 1990). More complex visual behaviors that are thought to depend on V2–V4 and interconnectivity between visual areas (Burkhalter et al., 1993; Kiorpes and Bassin, 2003; Zhang et al., 2005), such as contour integration (Baker et al., 2008; Kovács et al., 1999), develop up to a year later. Indeed, receptive field properties in high-level extrastriate cortex (e.g., face-selective areas) continue to develop during childhood (Gomez et al., 2018). Together, these findings suggest that extrastriate cortex undergoes substantial functional and anatomical maturation during human infancy; however, the timeline of this development remains unknown.

We measure the organization of visual cortex in human infants using retinotopic mapping with functional magnetic resonance imaging (fMRI)—the gold standard for defining visual areas V1–V4 in older children and adults (Conner et al., 2004; Wandell et al., 2007). Indeed, no prior study has used experimental stimuli to define the boundaries and properties of V1–V4 in infant primates and the youngest evidence of retinotopy in humans so far has come from 5 year olds (Conner et al., 2004; Gomez et al., 2018). The reason why this was not previously performed in infant primates is that fMRI studies in awake infants (human or non-human) present many challenges. Some of these challenges are general, such as head motion and fussiness, whereas others are specifically problematic for retinotopic mapping, such as an inability to instruct or enforce eye fixation (Braddick and Atkinson, 2011; Ellis and Turk-Browne, 2018). Moreover, the organization of infant visual cortex may not be detectable at the macroscopic level accessible to fMRI (Chapman et al., 1999). That said, there is some reason

for optimism: fMRI has revealed that infant visual cortex is functionally interconnected (Gao et al., 2017), shows evoked responses to visual inputs (Ellis et al., 2020), and responds in a localized way to motion (Biagi et al., 2015) and categories (Deen et al., 2017; Kosakowski et al., 2021).

To perform retinotopic mapping, we used a protocol that enables fMRI in awake and behaving infants (Ellis et al., 2020). In individual infants from 5 to 23 months old, we sought to define ventral and dorsal V1, V2, and V3, dorsal V3A/B, and ventral V4. In an alternating block design, we used meridian mapping (horizontal versus vertical) to identify area boundaries between quarter-field (V1–V3) and half-field (V3A/B and V4) representations (Fox et al., 1987; Schneider et al., 1993) and spatial frequency mapping (high versus low), which correlates with the eccentricity and receptive field maps in humans' and macaques' areas (Arcaro and Livingstone, 2017; Henriksson et al., 2008; Singh et al., 2000; Smith et al., 2001; Srihasam et al., 2014). We chose these stimuli because in pilot studies they were more tolerant to inconsistent fixation than traveling wave (Engel et al., 1994; Sereno et al., 1995) or population receptive field (Dumoulin and Wandell, 2008) approaches (Wandell et al., 2007). Namely, the desired stimulation was received wherever the infant fixated on the stimulus, whereas other methods are more fixation dependent. These two visual maps provided independent, potentially divergent measures of functional maturity in infant visual cortex. This allowed us to test whether striate and extrastriate cortex is organized early in infancy, as well as how it develops over our age range.

RESULTS

We showed 17 infants (5–23 months old; Table S1) blocks of stimuli (Figure 1) for meridian mapping (Fox et al., 1987; Schneider et al., 1993) and spatial frequency mapping (Arcaro and Livingstone, 2017; Henriksson et al., 2008). In meridian mapping blocks, large horizontally and vertically oriented meridian bow ties were shown for 20 s each in a counterbalanced order. In spatial frequency mapping blocks, low (0.05 cycles per visual degree) and high (1.5 cycles per visual degree) spatial frequency Gaussian random fields were shown for 20 s each in a counterbalanced order. Throughout all blocks, an animated movie of monochromatic shapes was shown at center to encourage fixation. Blocks were separated by 6 s of rest.

Behavior

Gaze was manually coded from a video recording of the infant's face during the scan. In meridian mapping blocks, participants similarly looked at the horizontal meridian ($M = 0.92$) and vertical meridian ($M = 0.92$) for almost the entirety of the block (difference confidence interval [CI] = $[-0.02, 0.02]$; $p = 0.866$). As expected—and motivating our use of task designs tolerant to eye movements—looking at the meridian did not necessarily mean fixating on the movie at center. Infants often looked left or right of center on horizontal meridians (proportion $M = 0.41$) and above or below center on vertical meridians ($M = 0.30$). There was a trend toward more off-center looking for horizontal versus vertical (difference CI = $[0.00, 0.22]$; $p = 0.050$), which may reflect the ease of saccading in azimuth compared to elevation (Aslin and Salapatek, 1975). In spatial frequency blocks, participants similarly looked at high spatial frequencies ($M = 0.90$) and low spatial frequencies ($M = 0.91$) for almost the entirety of the block (difference CI = $[-0.04, 0.01]$; $p = 0.310$). Infants infrequently looked off-center on high spatial frequencies ($M = 0.08$) and low spatial frequencies ($M = 0.05$). There was reliably more off-center looking for high versus low (difference CI = $[0.00, 0.07]$; $p = 0.033$), which may reflect a need to foveate high-frequency stimuli (given that the movie was shown at fixation). Overall, stimulation was uniform in the visual field the majority of the time, and the remaining time was still usable because of designs that ensured similar stimulation, regardless of where the stimulus was fixated. Indeed, Table S2 shows that if we censor time points from the generalized linear model (GLM) where the participant was looking away, then the statistical maps are negligibly changed.

The sample includes a wide age range of participants (5–23 months old), and so data quality may vary with age. A critical concern is that our youngest infants may have lower quality data than older infants. We evaluated this with three metrics. First, we correlated the amount of usable gaze data with participant age. Although the range of values is high and narrow (87%–95%), we found a significantly negative correlation, $r = -0.55$, $p = 0.008$, indicating that younger infants had more usable gaze data. Second, we quantified the number of time points in our data excluded because of head motion and correlated that with age. There was no significant relationship, $r = -0.16$, $p = 0.219$. Third, we correlated the number of usable events (i.e., period of stimulation for each condition) with age and found no significant relationship, $r = 0.26$, $p = 0.155$. Together, these results do not support age-related improvements in data quality within the age range of our infants.

Evidence of arealization

A GLM was used to estimate blood-oxygenation-level-dependent (BOLD) responses to the four stimulus conditions. As an initial check, infant visual cortex was robustly activated when collapsing across conditions (Figure S1A). Indeed, contrasts between horizontal and vertical meridians (Figure S1B) and between low and high spatial frequencies (Figure S1C) in each participant's volumetric space indicate strong differential responding to the conditions.

To determine whether infants have retinotopic organization, we first created surface reconstructions using iBEAT v2.0 (Li

et al., 2014, 2015, 2019; Wang et al., 2018; Figure S2). These surfaces were inflated and cut to make flatmaps. The contrast between horizontal and vertical meridians was projected onto each participant's flatmap and used for tracing visual areas. Areas were traced based on the alternations in sensitivity to horizontal and vertical meridians using a suitable protocol for adults (Wandell et al., 2007). For instance, the V1/V2 border was based on the peak in the vertical meridian representation on the gyral banks of the calcarine and the V2/V3 border was based on the next peak in the horizontal meridian representation. Figure 2A shows this contrast for an example 5.5-month-old participant. Overlaid on this surface are the manually traced areas demarcating striate and extrastriate cortex. In this participant, we traced V1, V2, and V3 in both hemispheres and in both ventral and dorsal cortex, as well as left dorsal V3A/B and bilateral ventral V4. Although there was variability across participants (Figure 2B), differential sensitivity to horizontal and vertical meridians was clear enough to identify areas in 16 out of 17 infants (average of 6.6 out of 8 areas in each hemisphere).

To verify that there were gradients of selectivity to different meridians across the visual areas (Arcaro et al., 2009), we traced lines perpendicular to the area boundaries, starting at the fundus of the calcarine sulcus and progressing anterior, and extracted the contrast values from points along those lines (dashed lines on Figure 2A). There were reliable oscillations in sensitivity to horizontal and vertical meridians across areas (Figure 2C). To quantify the evidence of boundaries, we computed the difference from zero for nodes near the borders between regions. The difference was reliably below zero (vertical selective) for the V1/V2 border, $M = -2.09$, CI = $[-2.68, -1.54]$, $p < 0.001$. For the V2/V3 border, the response was positive (horizontal selective), $M = 2.70$, CI = $[1.80, 3.65]$, $p < 0.001$. For the V3/V3AB boundary, the response was negative, $M = -1.01$, CI = $[-1.84, -0.21]$, $p = 0.013$. For the V3/V4 boundary, the difference was not reliable across participants, $M = -0.37$, CI = $[-1.20, 0.72]$, $p = 0.434$. Note that these border contrasts should be interpreted with care because the same data were used to trace the areas as were used to quantify changes across areas. Nonetheless, this analysis shows the consistency of these patterns across participants. Indeed, three participants under 6 months also showed this same pattern, providing evidence of early retinotopic organization.

Reliability of traced areas

To evaluate the reliability of our manually defined areal boundaries, we had a second mapper manually trace all sessions. This allowed us to evaluate the extent to which the two mappers independently identified retinotopic organization in the infants. The two mappers consistently found the same areas across participants, $d' = 1.87$ (Figure S3A). The extent of overlap was evaluated using the Dice coefficient and was high overall: 0.54 (range: 0.35–0.70). This varied across regions, with ventral V1 having the highest Dice ($M = 0.64$) and V4 having the lowest ($M = 0.21$; Figure S3C). The area boundaries were placed in similar locations across mappers, with a typical error between 1 and 4 mm (Figure S3D), which is close to the voxel resolution (3 mm) and smoothing kernel size (5 mm). The exception to this was V4, which had substantially larger error ($M =$

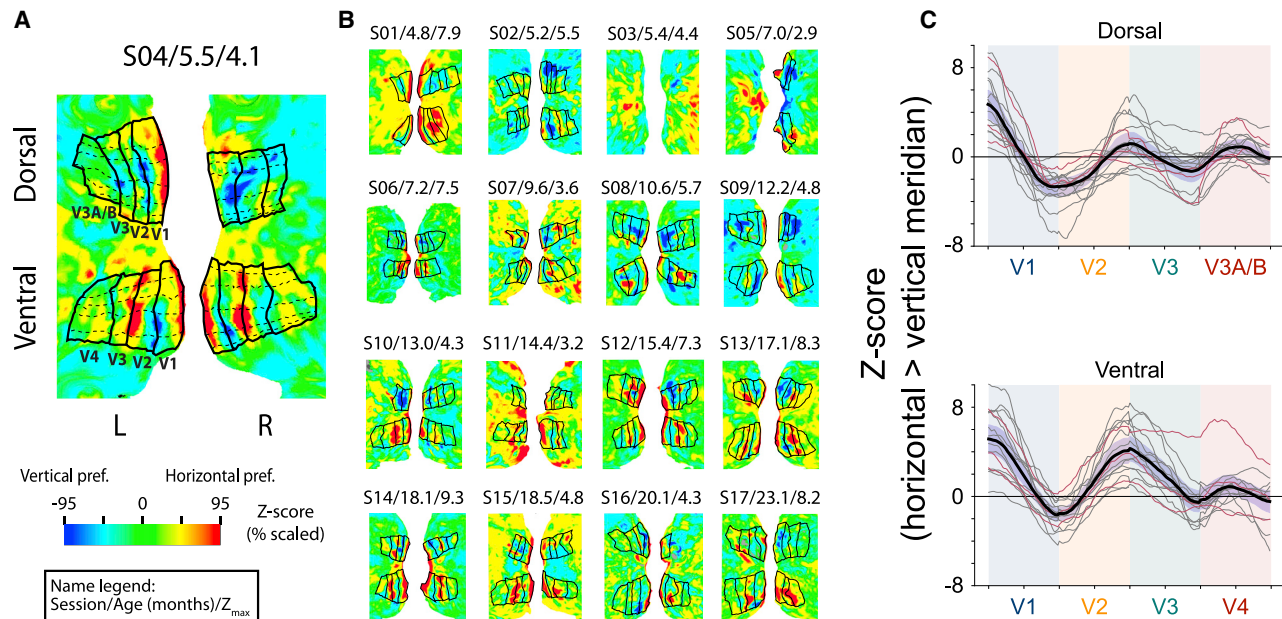


Figure 2. Arealization of infant visual cortex

(A) Example 5.5-month-old participant with areas traced on the cortical surface. Labels are shown for left V1, V2, V3, V3A/B, and V4. Colors indicate the Z-statistic value for the contrast of horizontal greater than vertical meridian. The maximum range of the color map was set to the 95th percentile of voxel Z-statistics in the occipital lobe (here, 4.1). Dashed lines drawn perpendicular to area boundaries were used to measure oscillations in sensitivity to horizontal and vertical meridians across areas. The text above each surface indicates: session label/age in months/ Z_{max} .

(B) Statistical maps for all 16 other participants, ordered youngest to oldest from top left to bottom right. Refer to [Table S1](#) for details about the participants.

(C) Contrast values of horizontal greater than vertical meridian for points on lines drawn perpendicular to the area boundaries, separately for dorsal and ventral areas. The extracted values were interpolated to a normalized size across areas, with the normalizing factor deriving from the area length, which is reported in [Table S3](#). The black line indicates the average of all participants, the gray lines indicate participants over 6 months, and the red lines indicate participants under 6 months. The purple-shaded region around the black line is the 95% confidence interval across participants estimated by bootstrapping a sampling distribution of the mean.

See also [Figure S1](#).

10.7 mm). In addition to showing that these traced areas were similar between the two mappers, the results are largely consistent when using the areas traced by this second mapper ([Figure S4](#)). The only substantial deviation is that the surface areas of V3, V3A/B, and particularly V4 were smaller for the second mapper. The size ([Conner et al., 2004; Dougherty et al., 2003](#)) and anatomical extent ([Winawer et al., 2010; Withoft et al., 2014](#)) of V4 in the second mapper better aligned with what is known from adults. Nonetheless, these analyses together provide compelling evidence that the areas identified were reliable and that the results were robust across mappers.

In addition to inter-rater reliability, we also examined whether the resulting areas were more similar within versus between infants in the subset of four participants with two or three sessions. Participant surfaces were aligned to standard space, and the similarity of the manually traced areas was computed between participants using the Dice coefficient. Similarity was higher within the same participant across sessions ($M = 0.47$) than between different participants yoked to have the same age difference ($M = 0.35$; $CI = [0.08, 0.17]$; $p < 0.001$). In fact, when a participant was compared across sessions, they were often more similar to themselves than to almost any other participant (rank $M = 2.0$; $p < 0.001$). This was true for even the youngest participant with two sessions (S02 at 5.2 months and S05 at

7.0 months). Such reliability of the functional topographies is remarkable because the surfaces were based on anatomical scans from different ages that could vary in depth of cortical folding.

Spatial frequency tuning

In traditional adult retinotopy experiments, the phase map of retinotopic space is supplemented with a functional eccentricity map. Standard eccentricity mapping requires presenting stimuli at different distances from fixation. Because we could not instruct or enforce fixation, such stimuli are not viable in infants. Instead, we used spatial frequency tuning as a proxy for eccentricity. In adult fMRI ([Arcaro and Livingstone, 2017; Henriksson et al., 2008; Singh et al., 2000; Srihasam et al., 2014](#)), high spatial frequency stimuli evoke responses at close eccentricities near the fovea and in striate areas; conversely, low spatial frequency stimuli evoke responses at farther eccentricities in the periphery and in extrastriate areas. Moreover, there is a relationship between eccentricity and receptive field size of voxels, such that voxels representing the fovea or in striate areas have smaller receptive fields than voxels in the periphery or extrastriate areas ([Smith et al., 2001](#)). The relationship between spatial frequency and receptive field size has been supported directly by electrophysiology ([Tolhurst and Thompson, 1981](#)). Note, fMRI

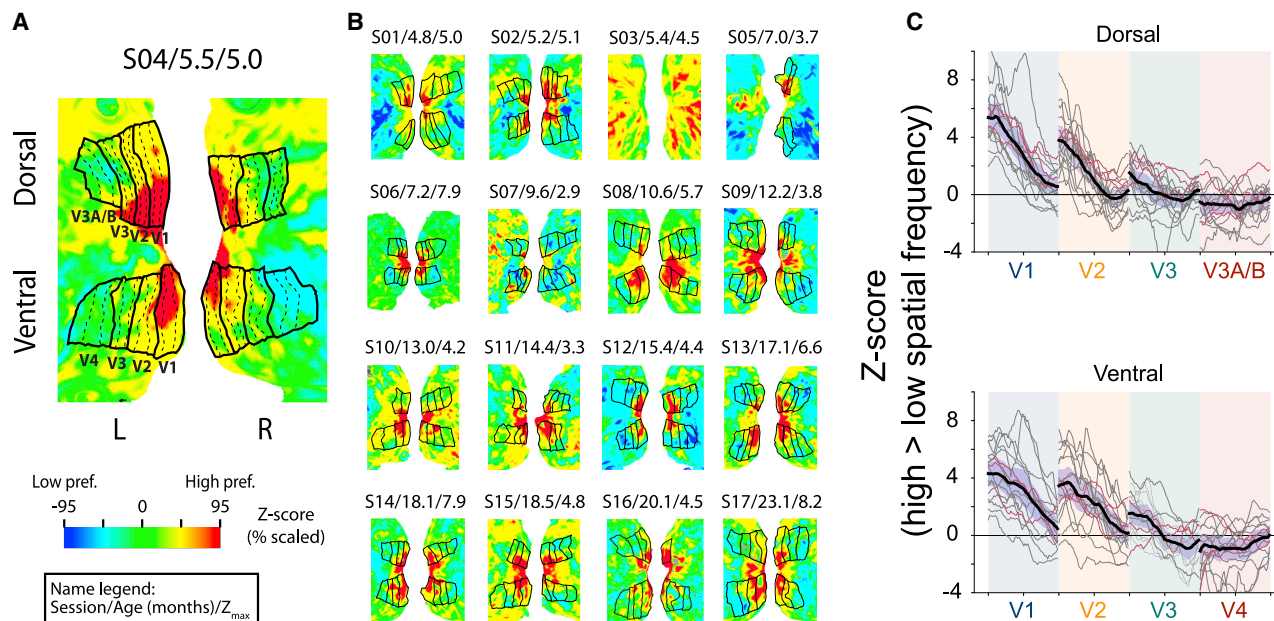


Figure 3. Spatial frequency tuning in infant visual cortex

(A) Example 5.5-month-old participant with areas traced on the cortical surface. Colors indicate the Z-statistic value for the contrast of high greater than low spatial frequency. The maximum range of the color map was set to the 95th percentile of voxel Z-statistics in the occipital lobe (here, 5.0). Dashed lines drawn parallel to area boundaries were used to measure gradients in sensitivity to high and low spatial frequencies across areas. The text above each surface indicates session label/age in months/ Z_{\max} .

(B) Statistical maps for all 16 other participants, ordered youngest to oldest from top left to bottom right. Refer to Table S1 for details about the participants.

(C) Contrast values of high greater than low spatial frequency for points on lines drawn parallel to the area boundaries, separately for dorsal and ventral areas. Each area is demarcated by a colored column. The foveal boundary of the area is on the left side of the column. The extracted values were interpolated to a normalized size across areas, with the normalizing factor deriving from the area width, which is reported in Table S4. The black line indicates the average of all participants, the gray lines indicate participants over 6 months, and the red lines indicate participants under 6 months. The purple-shaded region around the black line is the 95% confidence interval across participants estimated by bootstrapping a sampling distribution of the mean.

See also Figures S1 and S5.

is only sensitive to the selectivity preferences of the local population within voxels, which likely comprises neurons of mixed selectivity. Thus, to assess voxelwise sensitivity to eccentricity, we used a GLM to contrast BOLD activity for high versus low spatial frequency conditions. We extracted the Z-statistic map of this contrast between conditions. Figure 3A shows the spatial frequency map from a 5.5-month-old participant. There is a gradient from the fovea within each area, as well as the gradient from striate to extrastriate cortex. This pattern was observed across all participants with traced areas (Figure 3B).

To quantify the gradient in sensitivity to spatial frequency within areas, we traced lines within each area parallel to the area boundaries (dashed lines on Figure 3A). Importantly, these areas were drawn using the meridian mapping blocks, so the tracings were independent from the spatial frequency data. We measured the contrast of high greater than low spatial frequency along the lines, starting at the foveal boundary. Sensitivity to spatial frequency transitioned significantly from foveal (first quarter of the line) to peripheral (last quarter of the line) edges (Figure 3C) in dorsal and ventral V1 (CI = [3.18, 4.55]; $p < 0.001$), V2 (CI = [2.53, 4.22]; $p < 0.001$), and V3 (CI = [0.96, 2.25]; $p < 0.001$), but not V4 (CI = [-1.23, 0.05]; $p = 0.081$) or V3A/B (CI = [-0.96, 0.61]; $p = 0.695$). This confirms a gradient

in sensitivity to spatial frequency from the fovea to periphery in infant visual cortex, including in infants under 6 months. Indeed, the first versus last quarter difference did not reliably correlate with age in V1 ($r = 0.00$, $p = 0.958$), V2 ($r = 0.04$, $p = 0.793$), V4 ($r = 0.26$, $p = 0.258$), or V3A/B ($r = 0.18$, $p = 0.520$), although it was stronger in older children in V3 ($r = 0.52$, $p = 0.014$).

We also observed a gradient in sensitivity to spatial frequency across areas, with a greater overall response to high versus low spatial frequency in V1 versus V2 (CI = [0.82, 1.50]; $p < 0.001$), V2 versus V3 (CI = [1.21, 2.10]; $p < 0.001$), ventral V3 versus V4 (CI = [0.38, 1.51]; $p = 0.001$), and dorsal V3 versus V3A/B (CI = [0.52, 1.26]; $p < 0.001$). This confirms differences in spatial frequency tuning between striate and extrastriate areas in infant visual cortex, as has been observed in the adult visual cortex, although without the precise mapping that previous methods afforded (Henriksson et al., 2008). This same pattern was present in our participants under 6 months. In fact, the average difference between sensitivity to high versus low spatial frequency did not reliably correlate with age in V1 ($r = 0.35$, $p = 0.162$), V2 ($r = 0.27$, $p = 0.300$), V3 ($r = -0.12$, $p = 0.571$), V3A/B ($r = -0.39$, $p = 0.140$), or V4 ($r = -0.13$, $p = 0.683$). These findings are consistent with the possibility that gradients of spatial frequency tuning within and across areas are largely stable across infancy.

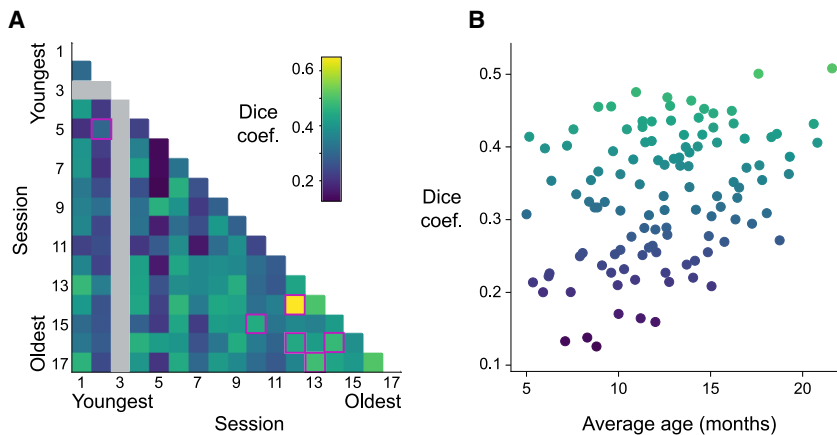


Figure 4. Overlap of retinotopic maps across infants

After alignment to standard space, the cortical locations of the manually traced visual areas were compared across pairs of sessions using the Dice coefficient.

(A) Rows and columns are ordered from youngest to oldest, with each cell depicting a session pair. The purple boxes highlight repeat sessions from the same participants. The gray strip corresponds to a participant (S03) with no traced visual areas.

(B) Correlation of Dice similarity with the average age (in months) of the infants in the two sessions being compared. Note that repeat sessions from the same infant are not visualized because they were excluded from the analysis to avoid biasing in favor of a correlation.

See also Figure S4.

We evaluated the role of high versus low spatial frequencies in this pattern of results. Figure S5 shows similar plots to Figure 3C except reporting the response to each condition separately. The high spatial frequency condition behaved more like an intermediate spatial frequency condition: although there was a gradient of decreasing response across regions, there was actually a greater response to these stimuli in the periphery of V2 and V3, rather than the fovea (Figure S5A). By contrast, the low spatial frequency condition showed strong eccentricity gradients within V1–V3 (Figure S5B). Variations in the low spatial frequency stimuli are only apparent at a large spatial scale, so the significant response in the periphery suggests that the periphery has relatively larger receptive fields. Hence, we observed gradients of spatial frequency sensitivity within and across areas *despite* the stimuli covering a partial range of spatial frequencies.

Area configuration and size

The results so far suggest that functional maps are established in infants as young as 5 months. However, this does not preclude the possibility of development in other properties of these areas, including their configuration and size. To test for such age effects, we extended the Dice similarity analysis. Rather than compare maps within individual infants across repeat sessions, here, we compare the manual tracings of areas V1–V4 in standard space across different infants (Figure 4A). We excluded repeat sessions from this analysis because they occurred more often in older infants and yielded higher Dice coefficients and thus could bias the correlation. This resulted in 114 pairs for comparison. We first tested whether maps are more similar for participants of similar ages, which would lead to a negative correlation between age difference and Dice coefficient (closer in age = higher similarity). However, this relationship was not reliable and in fact was numerically in the wrong direction ($r = 0.11$, $p = 0.454$). We then tested another potential age effect in which older infants are more similar to each other than younger infants are to each other. This would lead to a positive correlation between the average age of the two infants whose maps are being compared and their Dice coefficient (older age = higher similarity). We observed a robust positive correlation (Figure 4B; $r = 0.36$, $p = 0.033$). This correlation could be explained if we were better able to align older infants into standard space. However,

the number of manually coded alignment errors (Thompson et al., 2014) was not significantly correlated with age across sessions ($r = -0.39$, $p = 0.185$). The alignment of visual gyri and sulci between each participant and standard space also was not significantly related to age across sessions ($r = 0.14$, $p = 0.609$). When excluding outlier sessions (S02 and S05), the relationship between Dice and the difference between the participant ages was still non-significant, $r = 0.09$, $p = 0.586$. The relationship between Dice and the average age was now marginally significant, $r = 0.34$, $p = 0.065$, with the reduced significance likely reflecting the smaller sample size.

We interpret the increased similarity of older infants as reflecting convergence toward mature retinotopic organization. To evaluate this, we computed the Dice coefficient between the manual tracings of infant areas V1–V4 and a comprehensive atlas of adult visual areas in standard space (Wang et al., 2015). Overall, the similarity of infants to the standard adult atlas was high ($M = 0.39$). In fact, infants were more similar to this atlas than to other infants ($M = 0.33$; $CI = [0.03, 0.08]$; $p < 0.001$). There was an overall age effect, with similarity to the adult atlas increasing as a function of infant age (Figure 5; $r = 0.41$, $p = 0.042$). This is qualified by the fact that similarity did not increase systematically within infants across repeat sessions. Indeed, the high similarity even for the youngest infants suggests that any maturation rests on a foundation of early, adult-like arealization. The high similarity is likely also an underestimate given differences in the visual extent of the stimuli between adults (30° visual angle diameter) and infants (40°) and greater coverage of the fovea in the adult atlas. Moreover, different approaches to retinotopy were used for adults (traveling wave) and infants (meridian). The fact that the adult atlas still provided such a good guide to infant visual areas further helps validate the alignment of infant data to standard space and suggests that such atlases could plausibly be used as a starting point for regions of interest (ROIs) in future studies.

An alternative way to evaluate region maturation is to consider how the size of each area changes with age. If the visual cortex matures hierarchically, striate cortex might be mature in size by early infancy, at least relative to other areas, whereas the size of extrastriate areas would change over development. The most direct measure of size is surface area; however, this metric is

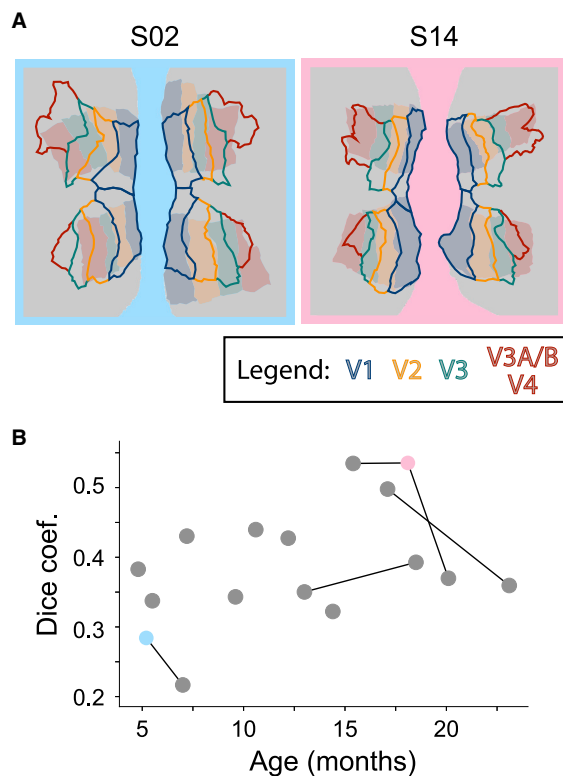


Figure 5. Overlap of infant visual areas with an adult atlas

Participants were aligned into standard space and compared to the relevant areas from an adult atlas (Wang et al., 2015).

(A) Two example infants with their manually traced visual areas in shaded fill colors and the corresponding visual areas from the standard space atlas overlaid in color outlines.

(B) Dice coefficient for each participant between visual areas from manual tracing in infants and standard atlas in adults. Lines connect the same participant across sessions. Cyan and pink dots correspond to participants S02 and S14, respectively.

See also Figure S4.

imprecise near the foveal boundary because of the video we played at fixation. We therefore used the length between area boundaries (dashed lines in Figure 2A) as our measure of size (although obtained similar results for surface area; Figure S6). Figure 6 shows the relationships across the 16 participants with traceable areas. There was a significant relationship between size and age in V1 ($r = 0.78$, $p < 0.001$) and V2 ($r = 0.50$, $p = 0.013$), and the relationship was marginal in V3 ($r = 0.54$, $p = 0.059$) and not significant in V3A/B ($r = 0.18$, $p = 0.517$) or V4 ($r = 0.23$, $p = 0.398$). These changes in V1 and V2 size could reflect the global growth of the brain over age in our sample ($r = 0.80$, $p < 0.001$). However, the relationship between size and age persisted after controlling for global volume in V1 (partial $r = 0.57$, $p = 0.024$), though not V2 ($r = 0.29$, $p = 0.208$; Table S3). Moreover, when we weigh data based on their inter-rater reliability (i.e., sessions that were similar between mappers were weighed more in the regression computation), we observe similar results: V1, $t(15) = 4.49$, $p = 0.001$; V2, $t(15) = 2.35$, $p = 0.034$; V3, $t(14) = 2.08$, $p = 0.058$; V3A/B, $t(14) = 0.58$, $p = 0.575$; and V4, $t(14) =$

0.81, $p = 0.430$. Hence, we see that only the earliest visual areas change in size across early development, counter to what would be expected if the visual cortex develops hierarchically.

DISCUSSION

We investigated the presence of retinotopic organization in human infants as young as 5 months. We present the first direct functional evidence of multiple visual maps in infants, demonstrating a maturity of extrastriate cortex that is greater than previously appreciated. We found evidence of visual areas V1–V4 even in our youngest participants, and these areas were reliable within participant across sessions. Moreover, there was a gradient in sensitivity from high to low spatial frequency within visual areas as well as a gradient across areas in the hierarchy, matching the topography of receptive fields, eccentricity, and spatial frequency tuning in adults (Arcaro and Livingstone, 2017; Henriksson et al., 2008; Smith et al., 2001; Srihasan et al., 2014; Tolhurst and Thompson, 1981). Together, these results support the existence of a hierarchical, retinotopic organization across visual cortex early in development. Area boundaries and spatial frequency maps did not consistently vary with age in our sample, although the size of V1 (but not the other areas) did increase with age controlling for global brain growth. The weak evidence for change in functional maps over this age range suggests that the development of visual cortex from 5 to 23 months may reflect fine-tuning rather than reorganization.

Our results suggest the presence of functional maps in human infants across striate and extrastriate cortex. Existence of functional maps in infant V1 has been well documented in animal models (Crowley and Katz, 1999; Farley et al., 2007; Purves and LaMantia, 1993; White and Fitzpatrick, 2007). Our results not only confirm the presence of a human infant V1 map but also show evidence of visual maps in extrastriate cortex. In particular, we find evidence of V2/V3, V3/V4, and V3/V3A/B boundaries in the infant brain. Previous connectivity-based analyses of spontaneous activity have indicated that this organization exists early in the development of the macaque visual system (Arcaro and Livingstone, 2017), yet this prior study failed to identify visual maps from visual stimulation until later in infancy. We found only moderate evidence of age-related change. It is possible that greater development would be observed outside of the age range tested here, either earlier in fetuses or neonates or later in toddlers or children. Additionally, more research is needed to explore how other functional maps in these areas develop over infancy.

Our results suggest that important architectural features of the visual system are established early in development. One consequence is that gradients in receptive field size may be available to scaffold the development of functionally selective high-level extrastriate cortex (Hasson et al., 2002; Arcaro et al., 2017; Gomez et al., 2019). In particular, this architecture may enable high-level extrastriate cortex to develop stereotyped localization of category-selective cortex (e.g., faces require high precision near fovea and buildings require precision in the periphery; Hasson et al., 2002). Our results also suggest that substantial changes in visual processing during infancy (Baker et al., 2011; Kovács et al., 1999; Lewis and Maurer, 2005; Patel et al., 2010;

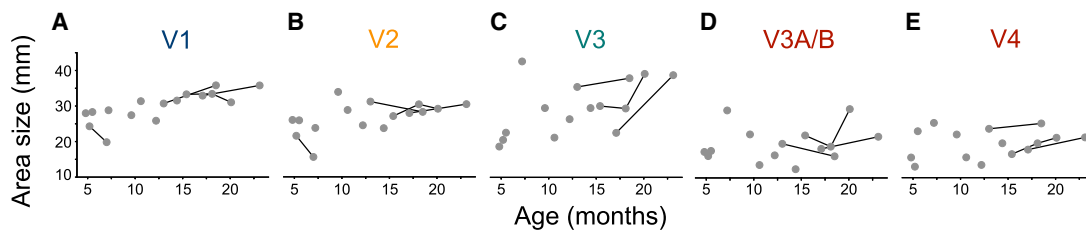


Figure 6. Relationship between the size of visual areas and infant age

Size is operationalized as the length between area boundaries averaged over hemispheres, summed over ventral and dorsal streams (for V1–V3). Lines are used to connect data from the same participant tested more than once across sessions. Raw data for each area are reported in Table S3. See also Figures S4 and S6.

Siu and Murphy, 2018) do not hinge on the emergence of retinotopic organization in striate and extrastriate visual cortex. One possibility is that improvements in visual processing over infancy and childhood could relate to improvements in downstream regions responsible for the readout of initial visual cortical processing, like high-level visual, association, or frontal cortices (Kiorpes and Movshon, 2014).

The protocol we used for drawing areas was modeled on adult retinotopy (Arcaro et al., 2009), which has been extensively validated with animal models (Fize et al., 2003; Wandell et al., 2007). The procedure sought to find reversals in sensitivity between horizontal and vertical stimuli. At the group level, we observed reliable preferences in the expected directions (e.g., the V1/V2 border was significantly selective to vertical orientations). However, there was variability at the individual participant level such that not all participants showed this pattern. An additional limitation of our study is that this protocol may not be optimized for human infants. For instance, identifying the extent of human V4 is notoriously difficult, in part because of the venous eclipse (Winawer et al., 2010) and in part because it is hard to distinguish the anterior border of V4 from surrounding areas like ventral occipital cortex (VO1) (Witthoft et al., 2014). These challenges may be exacerbated in infants because, for example, their areas are smaller and their scans are poorer quality. Indeed, in our inter-rater reliability analyses of the areal boundaries, V4 had the lowest reliability between the two mappers. Although one mapper relied purely on functional landmarks, the other mapper considered both anatomical and functional landmarks. The V4 from the latter mapper, whose results appear in the supplement, is more consistent with the expected size (Conner et al., 2004; Dougherty et al., 2003) and anatomical extent (Winawer et al., 2010; Witthoft et al., 2014) of adult V4. Nonetheless, the area boundaries were placed in similar positions across mappers and overall reliability was high, demonstrating that there is reliable evidence of arealization in the infant visual cortex.

We used a design that was common in adult retinotopy prior to the advent of modern methods. Meridian mapping has drawbacks (Wandell et al., 2007), such as limited coverage of the visual field and potentially mislabeling voxels that are only partially responsive to a stimulus. However, our goal was to demarcate visual areas (Kastner et al., 2000; Schneider et al., 1993; Shipp et al., 1995; Tootell et al., 1995), rather than estimate the selectivity of individual voxels (cf. population receptive fields; Dumoulin and Wandell, 2008), so these drawbacks are less pertinent. Indeed, the block design we used is relatively robust because

it can survive the exclusion of a small number of individual time points if the participant looks away or moves their head excessively. This is a key advantage for infant fMRI, in which data loss is expected within blocks. More critically, the traditional traveling wave approach (Engel et al., 1994) requires central fixation throughout each block. In piloting, even small eye movements in adults were deleterious for these methods. Fixation is impractical in human infants, as it cannot be instructed or enforced. Indeed, infants look away from center to the wedge in the periphery approximately one-third of the time, even with a movie at fixation. The methods we used are more robust to such eye movements because the stimulation is mostly uniform across the visual field, independent of gaze. That said, if a traveling wave approach could be adapted to these constraints, this might enable tracing of high-level extrastriate areas (Arcaro et al., 2009) that cannot be resolved with meridian mapping (Kastner et al., 2000; Shipp et al., 1995; Tootell et al., 1995).

There are also limitations to consider regarding our interpretation of spatial frequency gradients. The overlap between maps of spatial frequency, eccentricity, and receptive field size has been observed with fMRI (Arcaro and Livingstone, 2017; Henriksson et al., 2008; Singh et al., 2000; Smith et al., 2001; Srihasam et al., 2014) and electrophysiology (Tolhurst and Thompson, 1981). However, it is unclear whether the overlap of these maps exists in human infants like it does in adults and non-human primates. Moreover, the spatial frequencies chosen to represent high and low could both be considered low by traditional standards. Although the contrast of high versus low spatial frequency is consistent with what would be expected with a traditional range, it remains possible that different or more robust results would be obtained if even higher spatial frequencies had been used. Indeed, supplemental analyses found that the high spatial frequency condition may have behaved more like an intermediate spatial frequency condition. Together, this suggests that future research should examine the relationship between spatial frequency, eccentricity, and receptive field size in human infants using a more extreme range of spatial frequency values.

In sum, we found robust evidence for retinotopic organization in awake behaving human infants ranging in age from 5 to 23 months old. We identified four areas in both ventral and dorsal streams, spanning striate and extrastriate cortex, and these areas were reliable across sessions. The spatial frequency sensitivity within and across the areas mirrored what has been observed in adults. We found limited evidence of developmental change in retinotopic organization, other than the size of striate

cortex. Together, these results suggest that the early infant visual system has the basic cortical architecture needed for visual processing. By clarifying the timeline of retinotopic organization, our results could constrain theories of infant visual development, as well as the neural basis of infant visual dysfunction in disorders, such as amblyopia (Braddick and Atkinson, 2011). More broadly, retinotopy gave credibility to fMRI research during its earliest days by validating that architectural features of visual cortex first identified in animals could be measured in human BOLD (Engel et al., 1994; Schneider et al., 1993; Sereno et al., 1995). Likewise, the current study shows the promise of fMRI with awake behaving infants to reveal the function of the infant brain and to track changes across early postnatal development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Participants
- METHOD DETAILS
 - Data acquisition
 - Procedure
 - Gaze coding
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Preprocessing
 - GLM analysis
 - Visual area and gradient tracing
 - Comparing visual areas
 - Visual area size
 - Inter-rater reliability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.neuron.2021.06.004>.

ACKNOWLEDGMENTS

Thank you to N. Wilson and N. Córdova for help with initial experiment design. Thank you to K. Armstrong, L. Rait, J. Daniels, and the entire Yale Baby School team for recruitment, scheduling, and administration. Thank you to A. Bracher for help with scanning. Thank you to J. Fel and J. Wu for help with gaze coding. Thank you to K. Anderson for help with the iBEAT to Freesurfer conversion. Thank you to R. Watts for technical support. We are grateful for internal funding from the Department of Psychology and Princeton Neuroscience Institute at Princeton University and from the Department of Psychology and Faculty of Arts and Sciences at Yale University. N.B.T.-B. was further supported by the Canadian Institute for Advanced Research.

AUTHOR CONTRIBUTIONS

C.T.E., M.J.A., and N.B.T.-B. initially created the experiment design. C.T.E., L.J.S., V.R.B., and N.B.T.-B. did early piloting. C.T.E., T.S.Y., and N.B.T.-B. collected the data. C.T.E. and T.S.Y. preprocessed the data. C.T.E. and

M.J.A. performed the analyses. All authors contributed to the drafting of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: December 1, 2020

Revised: May 7, 2021

Accepted: June 4, 2021

Published: July 5, 2021

REFERENCES

- Angelucci, A., and Rosa, M.G. (2015). Resolving the organization of the third tier visual cortex in primates: a hypothesis-based approach. *Vis. Neurosci.* 32, E010.
- Arcaro, M.J., and Livingstone, M.S. (2017). A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife* 6, e26196.
- Arcaro, M.J., McMains, S.A., Singer, B.D., and Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *J. Neurosci.* 29, 10638–10652.
- Arcaro, M.J., Schade, P.F., Vincent, J.L., Ponce, C.R., and Livingstone, M.S. (2017). Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* 20, 1404–1412.
- Argall, B.D., Saad, Z.S., and Beauchamp, M.S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. *Hum. Brain Mapp.* 27, 14–27.
- Aslin, R.N., and Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Percept. Psychophys.* 17, 293–302.
- Baker, T.J., Tse, J., Gerhardstein, P., and Adler, S.A. (2008). Contour integration by 6-month-old infants: discrimination of distinct contour shapes. *Vision Res.* 48, 136–148.
- Baker, T.J., Norcia, A.M., and Candy, T.R. (2011). Orientation tuning in the visual cortex of 3-month-old human infants. *Vision Res.* 51, 470–478.
- Banks, M.S., Stephens, B.R., and Hartmann, E.E. (1985). The development of basic mechanisms of pattern vision: spatial frequency channels. *J. Exp. Child Psychol.* 40, 501–527.
- Biagi, L., Crespi, S.A., Tosetti, M., and Morrone, M.C. (2015). BOLD response selective to flow-motion in very young infants. *PLoS Biol.* 13, e1002260.
- Boothe, R.G., Dobson, V., and Teller, D.Y. (1985). Postnatal development of vision in human and nonhuman primates. *Annu. Rev. Neurosci.* 8, 495–545.
- Bourne, J.A., and Rosa, M.G. (2006). Hierarchical development of the primate visual cortex, as revealed by neurofilament immunoreactivity: early maturation of the middle temporal area (MT). *Cereb. Cortex* 16, 405–414.
- Braddick, O., and Atkinson, J. (2011). Development of human visual function. *Vision Res.* 51, 1588–1609.
- Braddick, O.J., Wattam-Bell, J., and Atkinson, J. (1986). Orientation-specific cortical responses develop in early infancy. *Nature* 320, 617–619.
- Burkhalter, A., Bernardo, K.L., and Charles, V. (1993). Development of local circuits in human visual cortex. *J. Neurosci.* 13, 1916–1931.
- Chapman, B., Gödecke, I., and Bonhoeffer, T. (1999). Development of orientation preference in the mammalian visual cortex. *J. Neurobiol.* 41, 18–24.
- Chino, Y.M., Smith, E.L., 3rd, Hata, S., and Cheng, H. (1997). Postnatal development of binocular disparity sensitivity in neurons of the primate visual cortex. *J. Neurosci.* 17, 296–307.

- Condé, F., Lund, J.S., and Lewis, D.A. (1996). The hierarchical development of monkey visual cortical regions as revealed by the maturation of parvalbumin-immunoreactive neurons. *Brain Res. Dev. Brain Res.* **96**, 261–276.
- Conner, I.P., Sharma, S., Lemieux, S.K., and Mendola, J.D. (2004). Retinotopic organization in children measured with fMRI. *J. Vis.* **4**, 509–523.
- Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173.
- Crowley, J.C., and Katz, L.C. (1999). Development of ocular dominance columns in the absence of retinal input. *Nat. Neurosci.* **2**, 1125–1130.
- Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194.
- Deen, B., Richardson, H., Dilks, D.D., Takahashi, A., Keil, B., Wald, L.L., Kanwisher, N., and Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nat. Commun.* **8**, 13995.
- Dice, L.R. (1945). Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302.
- Distler, C., Bachevalier, J., Kennedy, C., Mishkin, M., and Ungerleider, L.G. (1996). Functional development of the corticocortical pathway for motion analysis in the macaque monkey: a 14C-2-deoxyglucose study. *Cereb. Cortex* **6**, 184–195.
- Dougherty, R.F., Koch, V.M., Brewer, A.A., Fischer, B., Modersitzki, J., and Wandell, B.A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J. Vis.* **3**, 586–598.
- Dumoulin, S.O., and Wandell, B.A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660.
- Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–75.
- Ellis, C.T., and Turk-Browne, N.B. (2018). Infant fMRI: A model system for cognitive neuroscience. *Trends Cogn. Sci.* **22**, 375–387.
- Ellis, C.T., Skalaban, L.J., Yates, T.S., Bejanki, V.R., Córdova, N.I., and Turk-Browne, N.B. (2020). Re-imagining fMRI for awake behaving infants. *Nat. Commun.* **11**, 4523.
- Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.-J., and Shadlen, M.N. (1994). fMRI of human visual cortex. *Nature* **369**, 525.
- Farley, B.J., Yu, H., Jin, D.Z., and Sur, M. (2007). Alteration of visual input results in a coordinated reorganization of multiple visual cortex maps. *J. Neurosci.* **27**, 10299–10310.
- Fize, D., Vanduffel, W., Nelissen, K., Denys, K., Chef d'Hotel, C., Faugeras, O., and Orban, G.A. (2003). The retinotopic organization of primate dorsal V4 and surrounding areas: A functional magnetic resonance imaging study in awake monkeys. *J. Neurosci.* **23**, 7395–7406.
- Fonov, V., Evans, A.C., Botteron, K., Almlí, C.R., McKinstry, R.C., and Collins, D.L.; Brain Development Cooperative Group (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* **54**, 313–327.
- Fox, P.T., Miezin, F.M., Allman, J.M., Van Essen, D.C., and Raichle, M.E. (1987). Retinotopic organization of human visual cortex mapped with positron-emission tomography. *J. Neurosci.* **7**, 913–922.
- Gao, W., Lin, W., Grewen, K., and Gilmore, J.H. (2017). Functional connectivity of the infant human brain: plastic and modifiable. *Neuroscientist* **23**, 169–184.
- Gattass, R., Lima, B., Soares, J.G., and Ungerleider, L.G. (2015). Controversies about the visual areas located at the anterior border of area V2 in primates. *Vis. Neurosci.* **32**, E019.
- Gomez, J., Natu, V., Jeska, B., Barnett, M., and Grill-Spector, K. (2018). Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* **9**, 788.
- Gomez, J., Barnett, M., and Grill-Spector, K. (2019). Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nat. Hum. Behav.* **3**, 611–624.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., and Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* **34**, 479–490.
- Henriksson, L., Nurminen, L., Hyvärinen, A., and Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *J. Vis.* **8**, 5.1–13.
- Kaas, J.H. (1997). Topographic maps are fundamental to sensory processing. *Brain Res. Bull.* **44**, 107–112.
- Kastner, S., De Weerd, P., and Ungerleider, L.G. (2000). Texture segregation in the human visual cortex: A functional MRI study. *J. Neurophysiol.* **83**, 2453–2457.
- Kiorpes, L. (2016). The puzzle of visual development: behavior and neural limits. *J. Neurosci.* **36**, 11384–11393.
- Kiorpes, L., and Bassin, S.A. (2003). Development of contour integration in macaque monkeys. *Vis. Neurosci.* **20**, 567–575.
- Kiorpes, L., and Movshon, J. (2014). Neural limitations on visual development in primates: beyond striate cortex. In *The Visual Neurosciences*, Second Edition, L. Chalupa and J.S. Werner, eds. (MIT).
- Kosakowski, H., Cohen, M., Takahashi, A., Keil, B., Kanwisher, N., and Saxe, R. (2021). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *PsyArXiv*. <https://doi.org/10.31234/osf.io/7hqcu>.
- Kovács, I. (2000). Human development of perceptual organization. *Vision Res.* **40**, 1301–1310.
- Kovács, I., Kozma, P., Fehér, A., and Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proc. Natl. Acad. Sci. USA* **96**, 12204–12209.
- Lewis, T.L., and Maurer, D. (2005). Multiple sensitive periods in human visual development: evidence from visually deprived children. *Dev. Psychobiol.* **46**, 163–183.
- Li, G., Nie, J., Wang, L., Shi, F., Gilmore, J.H., Lin, W., and Shen, D. (2014). Measuring the dynamic longitudinal cortex development in infants by reconstruction of temporally consistent cortical surfaces. *Neuroimage* **90**, 266–279.
- Li, G., Wang, L., Shi, F., Gilmore, J.H., Lin, W., and Shen, D. (2015). Construction of 4D high-definition cortical surface atlases of infants: Methods and applications. *Med. Image Anal.* **25**, 22–36.
- Li, G., Wang, L., Yap, P.-T., Wang, F., Wu, Z., Meng, Y., Dong, P., Kim, J., Shi, F., Reik, I., et al. (2019). Computational neuroanatomy of baby brains: a review. *Neuroimage* **185**, 906–925.
- Lyon, D.C., and Connolly, J.D. (2012). The case for primate V3. *Proc. R. Soc. B* **279**, 625–633.
- McLaughlin, T., and O'Leary, D.D. (2005). Molecular gradients and development of retinotopic maps. *Annu. Rev. Neurosci.* **28**, 327–355.
- Norcia, A.M., Tyler, C.W., and Hamer, R.D. (1990). Development of contrast sensitivity in the human infant. *Vision Res.* **30**, 1475–1486.
- Orban, G.A., Van Essen, D., and Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn. Sci.* **8**, 315–324.
- Patel, A., Maurer, D., and Lewis, T.L. (2010). The development of spatial frequency discrimination. *J. Vis.* **10**, 1–10.
- Purves, D., and LaMantia, A. (1993). Development of blobs in the visual cortex of macaques. *J. Comp. Neurol.* **334**, 169–175.
- Schneider, W., Noll, D.C., and Cohen, J.D. (1993). Functional topographic mapping of the cortical ribbon in human vision with conventional MRI scanners. *Nature* **365**, 150–153.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893.
- Shipp, S., Watson, J.D., Frackowiak, R.S., and Zeki, S. (1995). Retinotopic maps in human prestriate visual cortex: the demarcation of areas V2 and V3. *Neuroimage* **2**, 125–132.
- Siegel, J.S., Power, J.D., Dubis, J.W., Vogel, A.C., Church, J.A., Schlaggar, B.L., and Petersen, S.E. (2014). Statistical improvements in functional

- magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* 35, 1981–1996.
- Singh, K.D., Smith, A.T., and Greenlee, M.W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* 12, 550–564.
- Siu, C.R., and Murphy, K.M. (2018). The development of human visual cortex and clinical implications. *Eye Brain* 10, 25–36.
- Smith, A.T., Singh, K.D., Williams, A.L., and Greenlee, M.W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* 11, 1182–1190.
- Srihasam, K., Vincent, J.L., and Livingstone, M.S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17, 1776–1783.
- Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al.; Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182.
- Tolhurst, D.J., and Thompson, I.D. (1981). On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proc. R. Soc. Lond. B Biol. Sci.* 213, 183–199.
- Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., and Belliveau, J.W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15, 3215–3230.
- van den Boomen, C., and Peters, J.C. (2017). Spatial frequency discrimination: Effects of age, reward, and practice. *PLoS ONE* 12, e0169800.
- van den Heuvel, M.I., and Thomason, M.E. (2016). Functional connectivity of the human brain in utero. *Trends Cogn. Sci.* 20, 931–939.
- Van Grootel, T.J., Meeson, A., Munk, M.H.J., Kourtzi, Z., Movshon, J.A., Logothetis, N.K., and Kiorpes, L. (2017). Development of visual cortical function in infant macaques: a BOLD fMRI study. *PLoS ONE* 12, e0187942.
- Wandell, B.A., Dumoulin, S.O., and Brewer, A.A. (2007). Visual field maps in human cortex. *Neuron* 56, 366–383.
- Wang, L., Mruczek, R.E., Arcaro, M.J., and Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cereb. Cortex* 25, 3911–3931.
- Wang, L., Li, G., Shi, F., Cao, X., Lian, C., Nie, D., Liu, M., Zhang, H., Li, G., Wu, Z., et al. (2018). Volume-based analysis of 6-month-old infant brain MRI for autism biomarker identification and early diagnosis. *Med. Image Comput. Comput. Assist. Interv.* 11072, 411–419, Springer.
- White, L.E., and Fitzpatrick, D. (2007). Vision and cortical map development. *Neuron* 56, 327–338.
- Winawer, J., Horiguchi, H., Sayres, R.A., Amano, K., and Wandell, B.A. (2010). Mapping hV4 and ventral occipital cortex: the venous eclipse. *J. Vis.* 10, 1–22.
- Witthoft, N., Nguyen, M.L., Golarai, G., LaRocque, K.F., Liberman, A., Smith, M.E., and Grill-Spector, K. (2014). Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cereb. Cortex* 24, 2401–2408.
- Zhang, B., Zheng, J., Watanabe, I., Maruko, I., Bi, H., Smith, E.L., 3rd, and Chino, Y. (2005). Delayed maturation of receptive field center/surround mechanisms in V2. *Proc. Natl. Acad. Sci. USA* 102, 5862–5867.
- Zheng, J., Zhang, B., Bi, H., Maruko, I., Watanabe, I., Nakatsuka, C., Smith, E.L., 3rd, and Chino, Y.M. (2007). Development of temporal response properties and contrast sensitivity of V1 and V2 neurons in macaque monkeys. *J. Neurophysiol.* 97, 3905–3916.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------------|----------------------------|---|
| Deposited data | | |
| Raw and preprocessed data | Dryad Digital Repository | https://doi.org/10.5061/dryad.7h44j0ztm |
| Software and algorithms | | |
| MATLAB v. 2017a | Mathworks | https://www.mathworks.com |
| Python v. 3.6 | Python Software Foundation | https://www.python.org |
| iBEAT v. 2.0 | UNC-IDEA Lab | https://ibeat.wildapricot.org |
| FSL v. 5.0.9 | FMRIB | https://fsl.fmrib.ox.ac.uk/fsl/fslwiki |
| Experiment menu v. 1.1 | Yale Turk-Browne Lab | https://github.com/ntblab/experiment_menu |
| Infant neuropipe v. 1.3 | Yale Turk-Browne Lab | https://github.com/ntblab/infant_neuropipe |

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the Lead Contact, Cameron Ellis (cameron.ellis@yale.edu).

Materials availability

This study did not generate new materials.

Data and code availability

The code for running the retinotopy task can be found at: https://github.com/ntblab/experiment_menu/tree/retinotopy/. The code for performing the analyses can be found at: https://github.com/ntblab/infant_neuropipe/tree/Retinotopy/. The data, including anonymized anatomical images, surface reconstructions, manually traced areas, and both raw and preprocessed functional images can be found at: <https://doi.org/10.5061/dryad.7h44j0ztm>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Data from 17 sessions with infants aged 4.8 to 23.1 months ($M = 12.2$, $SD = 5.7$; 13 females) met our inclusion criteria of at least one usable phase per condition. The planned sample size was to have two participants in each three-month window between 3 and 24 months. We met this criteria, except we only had one participant between 21 and 24 months. This sample size is larger than previous developmental studies of retinotopy (Conner et al., 2004). Not included in the sample are data from 15 sessions without enough blocks prior to exclusions for head motion and eye gaze, or from four sessions without enough blocks even prior to exclusions. The final sample included 12 unique participants, three participants who provided two sessions of usable data, and one participant who provided three sessions. These sessions occurred at least one month apart (range = 1.8–6.0) and so the data from these sessions were treated separately, similar to prior work (Deen et al., 2017). Refer to Table S1 for information on each participant. Data was collected at the Brain Imaging Center (BIC) at Yale University. Parents provided informed consent on behalf of their child. The study was approved by the Human Subjects Committee at Yale University.

METHOD DETAILS

Data acquisition

Data were acquired with a Siemens Prisma (3T) MRI with the bottom of the 20-channel Siemens head coil. Anatomical images were acquired with a T1-weighted PETRA sequence (TR1 = 3.32ms, TR2 = 2250ms, TE = 0.07ms, flip angle = 6°, matrix = 320x320, slices = 320, resolution = 0.94mm iso, radial slices = 30000). For three of our younger, compliant participants, we also collected a T2-weighted SPACE sequence (TR = 3200ms, TE = 563ms, flip angle = 120°, matrix = 192x192, slices = 176, resolution = 1mm isotropic), and these were supplied to iBEAT to support surface reconstruction. Functional images were acquired with a whole-brain T2* gradient-echo EPI sequence (TR = 2 s, TE = 30ms, flip angle = 71°, matrix = 64x64, slices = 34, resolution = 3mm iso, interleaved slice acquisition).

Procedure

There are many challenges when conducting fMRI research in early developmental populations (Ellis and Turk-Browne, 2018; van den Heuvel and Thomason, 2016). We have described and validated our approach to awake infant fMRI in a separate methods paper (Ellis et al., 2020). In brief, families visited the lab before their first scanning session for an orientation session. The aim of this was to acclimate the infant and parent to the scanning environment. Scanning sessions were scheduled for a time when the parents thought that the infant would be calm. The infant and parent were extensively screened for any metal on or in their body. Hearing protection for the infant consisted of three layers: silicon inner ear putty, over-ear adhesive covers, and ear muffs. The infant was positioned on the scanner bed, on top of a vacuum pillow that reduced movement. The top half of the head coil was not used because the bottom elements provided sufficient coverage of the smaller infant head. This allowed for sufficient visibility to monitor the infant's comfort and allowed us to project stimuli onto the ceiling of the bore directly above the infant's face using a custom mirror system. A video camera (MRC high-resolution camera) allowed us to record the infant's face during scanning for monitoring and offline eye gaze coding.

When the infant was focused, stimuli were shown in MATLAB using Psychtoolbox (<http://psychtoolbox.org>), against a gray background. For the meridian mapping blocks, a bow tie cut-out of a pastel-colored checkerboard was presented in either a vertical or horizontal orientation (Tootell et al., 1995). The arcs of the bow ties were 45° and their diameter spanned 40 visual degrees. The checkerboard spacing increased logarithmically out from the fovea, approximating the cortical magnification factor (Tootell et al., 1995). The color of the checkerboard alternated every 125ms between the original color pattern and its negative. For the spatial frequency mapping blocks, the stimuli were grayscale Gaussian random fields of high (1.5 cycles per visual degree) or low (0.05 cycles per visual degree) spatial frequency (Arcaro and Livingstone, 2017). The difference between these spatial frequencies should be detectable to 3-month old infants (Banks et al., 1985; Norcia et al., 1990), although spatial frequency discrimination develops into adolescence (Patel et al., 2010; van den Boomen and Peters, 2017). These patterns were shown as squares spanning 40° on each edge. There were five images for each spatial frequency, one of which was shown every 500 ms. Meridian and spatial frequency blocks contained two phases of stimulation. The first phase consisted of one of the conditions (e.g., horizontal or high) for 20 s, followed immediately by the second phase with the other condition of the same block type (e.g., vertical or low, respectively) for 20 s. The order of conditions was counterbalanced across blocks. At the end of each block there was at least 6 s rest before the start of the next block. Participants alternated between blocks of the spatial frequency and meridian mapping tasks, with the goal of acquiring 8 blocks total.

To facilitate attention to the center of the stimulus, a movie was played in a small window (1.5° in diameter) at that location. For spatial frequency blocks this was overlaid on top of the stimulus, for meridian mapping blocks the bow tie was overlaid on the movie. The animated movie showed grayscale shapes moving in unpredictable patterns, including jittering, looming, and smooth motion. The movie was saturated in order to minimize the amount of high contrast changes.

Gaze coding

Infant gaze was coded offline by two or three coders ($M = 2.12$) blind to condition. The coders determined whether the eyes were oriented left, right, up, down, center, off-screen (i.e., blinking or looking away), or were undetected (i.e., out of the camera's field of view). Codes for the directional gazes (i.e., left, right, up, and down) were only applied if the coder believed the infants were still looking at the screen. For instance, if the participant was looking left but off the screen then that frame was coded as off-screen. For the spatial frequency blocks, the gaze was considered acceptable if it was coded as center or any of the four directions. For the meridian mapping blocks, gaze directions perpendicular to the bow tie orientation were treated as equivalent to off-screen in preprocessing. For example, if the infant was viewing a horizontal bow tie, looking to the left, center, or right was considered acceptable.

The frame rate and resolution varied across participants, but the minimum rate was 16Hz and we always had sufficient resolution to identify the eye. The label for each frame was determined as the mode of a moving window of five frames centered on that frame across all coder reports. In case of a tie, we used modal response from the previous frame. The coders were highly reliable: when coding the same frame, coders reported the same response on 79% (range across participants = 71%–86%) of frames. Phases were excluded if the participant was coded as looking away from the stimulus for more than 25% of the time, computed separately for the two phases within each block. This was a stricter criterion than our other infant fMRI studies (Ellis et al., 2020) because of the importance of eye position for retinotopy. Across all included phases, participants looked at the stimulus 91% of the time on average (range across participants = 87%–95%).

To determine whether there were differences in the looking behavior across the conditions, we computed each participant's average proportion of looking time for every manual code that was allowable in that phase (e.g., left, center, or right for horizontal meridian). We quantified the difference in mean looking for horizontal and vertical conditions using bootstrap resampling. We also compared the mean looking time away from center per condition (e.g., left or right for horizontal meridian) with bootstrap resampling.

QUANTIFICATION AND STATISTICAL ANALYSIS

Preprocessing

Individual runs were preprocessed using FEAT in FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>), with optimizations for infant fMRI data. We discarded three volumes from the beginning of each run, along with the volumes automatically discarded by the EPI sequence. Blocks were stripped of any excess burn-in or burn-out volumes beyond the 3 TRs (6 s) that were planned. Some runs contained other experiments not discussed here ($N = 15$ sessions). In such cases, pseudo runs were created containing only the data of interest.

Blocks were sometimes separated by long pauses (> 30 s) within a session because of a break, an anatomical scan, or an intervening experiment ($N = 5$; $M = 556.5$ s break; range = 107.9–1183.4 s). We used the ‘centroid’ volume (i.e., with the minimal Euclidean distance from all other volumes) for alignment and motion correction. Slice-time correction was applied to realign the slices in each volume. Time points were excluded if the head motion between time points exceeded 3mm (average in blocks with at least one usable phase: $M = 1.9\%$, range = 0.0%–7.0%), and phases were excluded if more than 50% of TRs exceeded this threshold. We interpolated rather than removed these time points in order to avoid biasing the linear detrending (in later analyses these time points were removed). To make the mask of brain voxels, we thresholded the signal-to-fluctuating-noise ratio (SFNR) of each voxel in the centroid volume at the trough in the histogram of values. The data were smoothed with a Gaussian kernel (5mm FWHM) and linearly detrended. AFNI’s despiking algorithm attenuated aberrant time points within voxels. To account for differences across runs in intensity and variance, the blocks that were considered usable were normalized over time using Z-scoring, prior to the runs being concatenated for further analyses. For further explanation and justification of this preprocessing procedure, please refer to [Ellis et al. \(2020\)](#).

Participants were excluded if they did not have at least 1 usable phase from each of the four conditions. After these criteria were applied, participants (coincidentally) had a similar number of phases of each condition on average: 3.53 (SD = 1.33, range: 1–6) high spatial frequency, 3.47 (SD = 1.72, range: 1–6) low spatial frequency, 3.47 (SD = 1.46, range: 1–6) horizontal meridian, and 3.47 (SD = 1.46, range: 1–6) vertical meridian.

Each run’s centroid volume was registered to the infant’s anatomical scan from the same session. FLIRT with a normalized mutual information cost function was used for initial alignment. Additional manual registration was performed using mrAlign from mrTools (Gardner lab) to fix deficiencies of automatic registration. The preprocessed functional data were aligned into anatomical space with their original spatial resolution (3mm iso). This aligned data were mapped on to surface space, as described below. Whole-brain voxelwise analyses required further alignment of functional data into a standard space. For alignment to standard, the anatomical scan from each participant was automatically (FLIRT) and manually (Freeview) aligned to an age-specific MNI infant template ([Fonov et al., 2011](#)) and then aligned to the adult MNI template (MNI152). The functional data were transformed into standard space for the task versus rest contrast. To determine which voxels to consider at the group level, the intersection of brain voxels from all infant participants in standard space was used as a whole-brain mask.

For surface reconstruction, we used iBEAT v2.0 to acquire the surfaces ([Li et al., 2014, 2015, 2019](#); [Wang et al., 2018](#)). The output of the iBEAT pipeline is the inner and outer surfaces, as well as the volumetric segmentation of gray-matter and white-matter. [Figure S2](#) shows the surface reconstructions overlaid on a slice of the anatomical data for each participant. These were then inserted into a FreeSurfer-style pipeline (a walkthrough is provided in the codebase). As part of the pipeline, these surfaces were inflated into spheres and aligned to the Buckner40 template ([Dale et al., 1999](#)). To investigate the quality of the surfaces and the alignment to standard space, we used the ENIGMA consortium quality control procedure ([Thompson et al., 2014](#)), in which we evaluated defects in the projection of the Desikan-Killiany atlas onto the individual data. In particular, we quantified how many errors there were in each hemisphere for the following regions (atlas labels) that are known to be prone to poor segmentation: bankssts, precentral, postcentral, pericalcarine, parahippocampal, entorhinal, rostralanteriorcingulate, insula. The infant data showed typical amounts of errors compared to what is reported for adult data. The data were then resampled in SUMA, using an icosahedral shape, into standard space with a constant number of nodes ([Argall et al., 2006](#)). To generate flatmaps, a convexity map (based on the position of gyri and sulci) was computed using AFNI ([Cox, 1996](#)) and inflated brains were cut and flattened using the FreeSurfer procedure. Once flattened, statistics maps were projected on to the surface for evaluation.

GLM analysis

For the main analyses, a GLM was fit to the BOLD activity in each voxel using FEAT in FSL. Two separate GLMs were performed, one containing the horizontal and vertical meridian regressors, and the other containing the high and low spatial frequency regressors. Each regressor modeled phases with a boxcar lasting 20 s, convolved with a double-gamma hemodynamic response function. The six translation and rotation parameters from motion correction were included in the GLM as nuisance regressors. TRs that were excluded (i.e., had translational motion greater than 3mm) were scrubbed by including an additional regressor for each to-be-excluded time point ([Siegel et al., 2014](#)). In a follow-up analysis ([Table S2](#)), we scrubbed any time point where the participant was looking away from the stimulus for at least 25% of the TR. The condition regressors were then contrasted to find the differential evoked response. The voxelwise Z-statistic volumes for these contrasts were extracted for each participant. Similar patterns of activation were found if we used the contrast of parameter estimates, rather than the Z-statistic volumes ([Table S5](#)). For visualization purposes, we set the maximum value of these maps to be the 95th percentile of the Z-statistic value for each participant in a large, anatomically defined occipital mask.

To test for visual-evoked activity, an F -test was performed in FSL using each phase type (i.e., horizontal and vertical meridians, high and low spatial frequencies) as regressors in a GLM to identify which voxels respond to any visual stimulation ([Figure S1A](#)). As a conservative test to evaluate where the brain was most activated across participants, we Z-scored the resulting F -values within each participant. Hence, any mean differences in F -values across participants are mitigated. Instead, what matters is whether the F -values are high, relative to other voxels in that participant, in the same voxels across participants. The resulting statistic map was volumetrically aligned to standard space and then merged across all participants. We used threshold free cluster enhancement through the randomize function in FSL, resulting in voxel clusters $p < 0.05$ corrected.

Visual area and gradient tracing

The contrast maps for horizontal greater than vertical were transformed onto the flat surface map and used for tracing (Argall et al., 2006). Traditional tracing guidelines for adult humans were followed (Wandell et al., 2007). The areas traced were ventral V1, V2, V3 (also known as VP), and V4 (also known as hV4), and dorsal V1, V2, V3 and V3A/B. The border between V1 and V2 was defined by the peak in the vertical meridian, the border between V2 and V3 was defined by the peak in the horizontal meridian, the border between V3 and ventral V4 or dorsal V3A/B was defined by the peak in the vertical meridian, and the terminal border of ventral V4 and dorsal V3A/B was defined by the peak of the vertical meridian after a half cycle. It is typical to trace only these areas using a meridian mapping paradigm (Kastner et al., 2000; Shipp et al., 1995; Tootell et al., 1995), as well as early traveling wave studies (Serenio et al., 1995). This procedure is likely to lead to imprecision between the V4 boundary and the VO1 boundary and did not provide sufficient resolution to demarcate V3A and V3B. When the distinctions between areas were not clear, the maximum range of the colormap was varied. If this did not resolve the ambiguity, then the evoked response to just the vertical meridian (rather than the contrast between horizontal and vertical) was checked. If these additional steps did not clarify the boundary between areas, the area was not drawn. The peripheral extent of the area was estimated by the horizontal greater than vertical contrast and the *F*-test for each participant. Because of the movie shown at fixation, the foveal response was expected to be contaminated, hence the areas were not traced to the foveal confluence. To test how the foveal confluence varied with age, we also traced the areas that bridged between the ventral and dorsal areas of V1, V2, and V3 for each hemisphere.

To quantify alternations in sensitivity to horizontal versus vertical meridians across visual areas (Figure 2C), we traced additional lines of interest in the visual cortex using SUMA (Figure 2A; Arcaro et al., 2009). These lines were drawn perpendicular to the area boundaries, posterior to anterior, for areas that were traced. Five lines, spaced equally along the width of the areas, were drawn for each hemisphere and the dorsal and ventral areas. These lines were drawn using only the areal boundaries for guidance, making the mapper unaware of the local intensity changes within areas. Nodes in surface space on those lines were indexed for their values of the horizontal greater than vertical meridian contrast. The number of nodes along each line varied between areas and participants. In order to standardize the size of these lines for the sake of comparison, we interpolated the values along each line to contain 50 values within each area (up to 200 total). To test the selectivity at the boundary between regions, we averaged the six nodes on these interpolated lines on either side of this boundary. The values were averaged across the ventral and dorsal cortex for the V1/V2 and V2/V3 border. For the border anterior to V3, they were treated separately. The average values for each session were then bootstrap resampled to compute statistical reliability (Efron and Tibshirani, 1986). Namely, we sampled, with replacement, the difference in contrast values from all participants 10,000 times, averaging across participants on each iteration to generate a sampling distribution. Confidence intervals reflect the 2.5 and 97.5 percentiles of this distribution. For null hypothesis testing, we calculated the *p*-value as the proportion of samples whose mean was in the opposite direction from the true effect, doubled to make the test two-tailed.

The gradients in sensitivity from high to low spatial frequency (Figure 3C) were quantified by tracing lines within each area, parallel to the areal boundaries (Figure 3A). Two lines were traced from the foveal to the peripheral boundary of each area and were used to index the values of the high greater than low spatial frequency contrast. The indexed values along these lines were also interpolated to include 50 values.

To statistically test the gradients in sensitivity from high to low spatial frequency within area, we divided the lines running parallel to the area boundary into quartiles, with the first quartile containing the section of the line closest to the foveal confluence. The contrasts for ventral and dorsal areas were averaged for lines in V1, V2, and V3. We averaged the contrast of high greater than low spatial frequency for the first quartile and compared it to the fourth quartile using bootstrap resampling. To test whether the difference in contrast between quartiles varied with age, we used bootstrap resampling of the correlation, randomly sampling bivariate data from 17 participants with replacement and calculating the Pearson correlation on each of 10,000 iterations. We calculated the *p*-value as the proportion of samples resulting in a correlation with the opposite sign from the true correlation, doubled to make the test two-tailed.

To statistically test the gradient of spatial frequency tuning across (rather than within) areas, we averaged the contrast values along the whole lines and compared adjacent areas (i.e., V1 versus V2, V2 versus V3, etc.) using bootstrap resampling. Areas were averaged across dorsal and ventral, except when comparing ventral V3 with V4 and dorsal V3 with V3A/B. We used bootstrap resampling of the correlation to test the relationship between age and the averaged contrast within area.

Comparing visual areas

The consistency of the traced visual areas across participants was quantified using the Dice coefficient (Dice, 1945). This was computed by finding every node in the surface that was labeled as belonging to a visual area and comparing whether those nodes had the same label across participants. The Dice coefficient is a fraction where the number of matching nodes, multiplied by 2, is the numerator, and the total number of nodes with a label in either participant is the denominator. For a given pair of sessions, the Dice coefficient was quantified for each hemisphere separately and averaged. Only areas that were traced in both participants were considered. This means the denominator of the Dice coefficient is not inflated by labels that could not possibly match between participants. The Dice coefficient was calculated for all pairwise comparisons between sessions. Comparisons of the same participant across multiple sessions are used to determine whether there is greater reliability over time within versus between participants. To approximately match the ages of the participants being compared, we started with one of the participants with repeat sessions and calculated the age differential; we then found another participant with the most similar age difference to one of these sessions (mean

difference in age between the repeat session and the matched participant = 0.97 months). For example, consider the participant who was 5.2 months at their first session (S02) and 7.0 months at their next session (S05). To provide a between-participant control, we found the participant (S06) who was closest in age to the second session (7.2 months). We then compared the Dice coefficient of S02 and S05 with the Dice coefficient of S02 and S06. To evaluate this statistically, we used bootstrap resampling of the difference in Dice coefficients between the within- and between-participant pairs. The Dice coefficient comparing a participant to themselves across sessions was ranked against the Dice coefficients from relating that participant to all other participants. To evaluate significance of this rank, for each participant with a repeat session we generated a random rank in the possible range of values for that participant and averaged across the group. We did this 10,000 times to get a distribution of permuted ranks. We then compared the observed average rank to this permuted distribution to find the likelihood of finding a ranking as extreme as the one we observed.

To test how similarity between participants varied with age, we compared the ages of the participants to their Dice coefficient. We first tested whether similarity in age predicted the Dice coefficient. To do this, we subtracted the ages of participants and related the absolute value of this difference to the Dice coefficient. We next tested if the overall age of the participants predicted the Dice coefficient. To do this, we compared the average age of each pair of participants with the Dice coefficient for each pair. Bootstrap resampling of the session was used to evaluate the significance of these correlations. Namely, we resampled which sessions would be used to make pairs and only considered pairs from the resampled group. As above, the correlation was computed for each resample and the distribution of resampled correlations was used to determine significance. In both of these analyses, comparisons using the same participant across multiple sessions were ignored because the sampling of these participants was skewed to older ages. Because within-participant Dice coefficients were higher, this skew would have biased the correlations to be positive. Correlation was used in this analysis to compare age and Dice, despite the pairs lacking independence, because correlation normalizes the range between both metrics, unlike Euclidean distance or similar metrics.

To measure the quality of alignment to the standard template, we used two complementary metrics. First, we quantified the number of defects in the projection of the Desikan-Killiany atlas, as described above. Second, we correlated the convexity map of individual participants with the convexity map from standard space (from fsaverage; Dale et al., 1999). The convexity maps of the individual and standard surface were masked according to the relevant regions (V1–V4) that were labeled in the atlas of visual cortex (Wang et al., 2015). The correlation between the convexity map and standard space was computed for each hemisphere separately. This correlation was Fisher transformed and then averaged between hemispheres. Bootstrap resampling of the correlation was used to quantify the significance of the relationship between age and the measure of alignment.

To determine the degree to which the visual areas have adult-like localization, the Dice coefficient was calculated between the traced areas in each infant and those same areas as defined in an atlas of the visual hierarchy (Wang et al., 2015). All of the areas we traced were available for comparison in the atlas except for V3A/B, which is separated in the atlas but was combined here. The maximum probability surface was used so that each node was uniquely assigned to a specific area, if at all.

The Dice coefficient to the standard atlas was computed for each individual and related to their age to determine whether the similarity to adults changes across early development. The significance of this correlation was computed using bootstrap resampling. We also tested whether the Dice coefficient was higher between two participants than between one participant and the atlas. The Dice coefficients comparing each participant with the other participants (excluding comparisons across sessions from the same participant) were averaged and subtracted from the Dice coefficient from that participant and the atlas. This difference for each participant was then tested for significance with bootstrap resampling.

Visual area size

Our main measure of the visual area size was the length of each traced area. This length was computed on surfaces in native space by taking all pairwise distances of nodes along the traced lines running perpendicular to the area boundaries and averaging them (Figure 2A). That is, length corresponds to the distance between boundaries. Length was chosen as the primary metric over other measurements like surface area because it is less sensitive to challenges in drawing the foveal border. Nevertheless, a similar pattern of results was obtained with surface area, computed by averaging the extent of the white matter and pial surface for all nodes labeled as belonging to an area. For V1, V2, and V3, we first averaged the lengths across hemispheres and then added these averages for the ventral and dorsal areas. This was still possible if an area was traced in only the left or right hemisphere, but if either the ventral or dorsal areas had not been traced in both the left and right hemisphere then the length of this area was not estimated. The average length of the remaining participants was related to their age using bootstrap resampling. To observe size changes independent of overall brain size, a partial correlation was computed where overall brain size (as measured based on the skullstripped volume from iBEAT v2.0) was used as a covariate. Analyses using gray matter volume as a covariate produced similar results. This partial correlation was also evaluated statistically using bootstrap resampling.

Furthermore, we tested the relationship between age and area length by weighting data points according to the inter-rater reliability of area tracings. Namely, we used the overall Dice coefficient between the mappers for each individual participant as a weight in a weighted least-squares regression. The parametric statistical test of the beta coefficient was used to evaluate the strength of the relationship after weighting the Dice coefficient.

To test whether the foveal confluence varied in size with age, the area of each foveal area was divided by the area of the sum of dorsal and ventral area for that area, and then correlated with age using bootstrap resampling.

Inter-rater reliability

We had a second expert mapper independently trace each session. This mapper was naive to the participant's age and did not view the maps of the original mapper during tracing. Indeed, we encouraged the second mapper to rely on their expertise to identify the areas using whatever approach they felt was most appropriate. This resulted in an approach for identifying areal boundaries that differed from the original mapper in two ways. One difference was that this new mapper used the inflated surface view, rather than the flat map. The inflated and flattened surface views emphasize different aspects of the folding and could result in different criteria for drawing areal borders. One clear difference occurs at the horizontal meridian representation of V1. To flatten the cortical surface, a cut is made along the horizontal meridian of V1 and surface nodes along the cut are not visualized on the flattened surface view. This meant that the V1 for the second mapper extended to the calcarine midline and included the strongest horizontal meridian response, whereas the original mapper only went to the edge of the calcarine cut. The areal boundaries of both the most foveal and peripheral representations also varied between mappers since the different viewing methods emphasized different aspects of the borders to prioritize for consistency. This resulted in the mappers employing different criteria for the shape and extent of areal boundaries. Another difference was that the second mapper used sulcal landmarks to guide boundary drawing, whereas the original mapper intentionally hid the convexity map when tracing boundaries. This helped the second mapper constrain boundaries, especially in V4, to be more consistent with boundaries typically found in adults.

To compare the consistency of areas labeled by the two mappers, we considered the proportion of areas that both mappers labeled (i.e., hit rate) against the proportion that the original mapper did not label but were labeled by the second mapper (i.e., false alarm rate). This allowed us to compute the d' to quantify the detection precision across mappers. As a less precise measure, we also compared the total number of areas (out of 16) traced by each mapper for each participant. We considered the overlap between the areas drawn by the two mappers. We used the Dice coefficient and evaluated the results when using all areas. Additionally, we computed the Dice for each area separately (averaging across hemispheres when available). To compute the alignment of areal boundaries we used the lines drawn perpendicular to the area boundaries and quantified the length segment in each area. Note, we used the same lines for both mappers. If we had used different lines for the mappers then similarity in length might not mean the borders are actually in the same location. Instead, if the boundaries are aligned then the lengths should be the same so the difference will be near zero.

We re-ran the analyses that could be run with just the drawn areas identified by the second mapper. This meant we could not do some analyses, such as quantifying the gradients in spatial frequency sensitivity, because we did not have lines traced in these areas. Nonetheless, with just the areal boundaries identified from this second mapper, we were able to compute the Dice between the infants and adult atlas (akin to [Figure 5](#)), compare participants across sessions (akin to [Figure 4](#)), and quantify surface area (akin to [Figure S6](#)).