Movies reveal the fine-grained organization of infant visual cortex

Ellis, C. T.[1], Yates, T. S.[2], Arcaro, M. J.[3], & Turk-Browne, N. B.[2,4]

[1]*Department of Psychology, Stanford University, Palo Alto, CA 94304*

[2]*Department of Psychology, Yale University, New Haven, CT 06511*

[3]*Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104*

[4]*Wu Tsai Institute, Yale University, New Haven, CT 06511*

** Corresponding author and lead contact: cte@stanford.edu*

Studying infant minds with movies is a promising way to increase engagement relative to traditional tasks. However, the spatial specificity and functional significance of movie-evoked activity in infants remains unclear. Here we investigated what movies can reveal about the organization of the infant visual system. We collected fMRI data from 15 awake infants aged 5–23 months who attentively watched a movie. The activity evoked by the movie reflected the functional profile of visual areas. Namely, homotopic areas from the two hemispheres responded similarly to the movie, whereas distinct areas responded dissimilarly, especially across dorsal and ventral visual cortex. Moreover, visual maps that typically require time-intensive and complicated retinotopic mapping could be predicted from movie-evoked activity in both data-driven analyses (i.e., independent components analysis) at the individual level and by using functional alignment into a common low-dimensional embedding to generalize across participants. These results suggest that the infant visual system is already structured

to process dynamic, naturalistic information and that fine-grained cortical organization can be discovered from movie data.

Keywords: naturalistic tasks, child development, fMRI, ventral and dorsal visual streams, retinotopic mapping, functional alignment

Studying the function and organization of the youngest human brains remains a challenge. Among the most important obstacles facing this research is that infants are unable to maintain focus for long periods of time and struggle to complete traditional cognitive tasks[1]. Movies can be a useful tool for studying the developing mind[2], as has been shown in older children[3–5]. The dynamic, continuous, and content-rich nature of movie stimuli[6,7] make them effective at capturing infant attention[8,9]. Here, we examine what can be revealed about the functional organization of the infant brain during movie-watching.

We focus on visual cortex because its organization at multiple spatial scales is well understood from traditional, task-based fMRI. The mammalian visual cortex is divided into multiple areas with partially distinct functional roles[10,11]. Within visual areas, there are orderly, topographic representations, or maps, of visual space[12,13]. These maps capture information about the location and spatial extent of visual stimuli with respect to fixation. Thus, maps reflect sensitivity to polar angle, measured via alternations between horizontal and vertical meridians that define area boundaries[14,15], and sensitivity to spatial frequency, reflected in gradients of sensitivity to high and low spatial frequencies from foveal to peripheral vision, respectively[16]. Previously, we reported that these maps could be revealed by a retinotopy task in infants as young as 5 months of age[17]. However, each session of infant fMRI allows for only 5–10 minutes of data collection[1], and so requiring retinotopic mapping tasks to define regions of interest would leave little time for other tasks. Moreover, retinotopy tasks typically require controlled behavior (e.g., fixation) which can

lead to attrition.

The primary goal of the current study is to investigate the organization of visual cortex with movie-watching data. Movies drive strong and naturalistic responses in sensory regions while minimizing task demands[6,7,18]. Movies have been useful in awake infant fMRI for studying event segmentation[19], functional alignment[20], and brain networks[21]. However, this past work did not address the granularity and specificity of cortical organization that movies can identify. For example, infants have similar evoked responses to movie content across large portions of visual cortex[19], but the structure of those responses remains unclear, especially how they vary across visual areas. There are several reasons for skepticism that movies could evoke detailed, retinotopic organization: Movies may not fully sample the stimulus parameters (e.g., spatial frequencies) or visual functions needed to find topographic maps and areas in visual cortex. Even if movies contain the necessary visual properties, they may unfold at a faster rate than can be detected by fMRI. Additionally, naturalistic stimuli may not drive visual responses as robustly as experimenter-defined stimuli that are designed for retinotopic mapping with discrete onsets and high contrast. Finally, the complexity of movie stimuli may result in variable attention between participants, impeding discovery of reliable visual structure across individuals.

We conducted several analyses to probe different kinds of visual granularity in infant movie-watching fMRI data. First, we asked whether distinct areas of the infant visual cortex have different functional profiles. Second, we asked whether the topographic organization of visual areas can be recovered within participants. Third, we asked whether this within-area organization is aligned

across participants. These three analyses assess key indicators of the mature visual system: functional specialization between areas, organization within areas, and consistency between individuals. Throughout, we explored the feasibility of *only* relying on movies to perform retinotopic mapping of infant visual areas.

Results

We performed fMRI in awake, behaving infants using a protocol described previously[1]. The dataset consisted of 15 sessions of infant participants (4.8–23.1 months old) who had both movie-watching data and retinotopic mapping data collected in the same session (Table S1). All available movies from each session were included (Table S2).

The retinotopic-mapping data from the same infants[17] allowed us to generate infant-specific meridian maps (horizontal versus vertical stimulation) and spatial frequency maps (high versus low stimulation). The meridian maps were used to define regions of interest (ROIs) for visual areas V1, V2, V3, V4, and V3A/B.

Evidence of area organization with homotopic similarity To determine what movies can reveal about the organization of areas in visual cortex, we compared activity across left and right hemispheres. Namely, we correlated timecourses of movie-related BOLD activity between retinotopically defined, participant-specific ROIs (7.3 regions per participant per hemisphere, range: 6–8)[22,23]. Higher correlations between the same (i.e., homotopic) areas than different areas indicates differentiation of function between areas. Moreover, other than V1, homotopic visual areas are

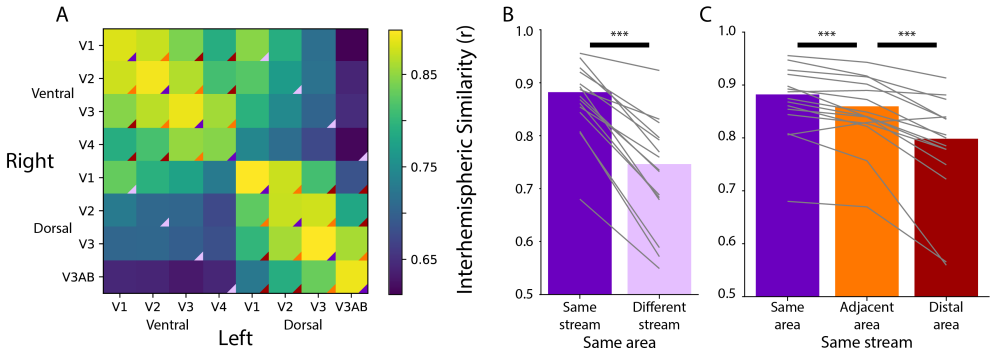

Figure 1: Homotopic correlations between retinotopic areas. (A) Average correlation of the time-course of activity evoked during movie watching for all areas. This is done for the left and right hemisphere separately, creating a matrix that is not diagonally symmetric. The color triangles overlaid on the corners of the matrix cells indicate which cells contributed to the summary data of different comparisons in subpanels B and C. (B) Across-hemisphere similarity of the same visual area from the same stream (e.g., left ventral V1 and right ventral V1) and from different streams (e.g., left ventral V1 and right dorsal V1). (C) Across-hemisphere similarity in the same stream when matching the same area (e.g., left ventral V1 and right ventral V1), matching to an adjacent area (e.g., left ventral V1 and right ventral V2), or matching to a distal area (e.g., left ventral V1 and right ventral V4). Grey lines represent individual participants. *** = $p<0.001$ from bootstrap resampling

anatomically separated across the hemispheres, so similar responses are unlikely to be attributable to spatial autocorrelation.

Homotopic areas (e.g., left ventral V1 and right ventral V1; diagonal of Figure 1A) were highly correlated (M=0.88, range of area means: 0.85–0.90), and more correlated than non-homotopic areas, such as the same visual area across streams (e.g., left ventral V1 and right dorsal V1; Figure 1B; $\Delta_{Fisher\,Z}$ M=0.42, $p<0.001$). To clarify, we use the term 'stream' to liberally distinguish visual regions that are *more* dorsal or *more* ventral, as opposed to the functional definition used in reference to the 'what' and 'where' streams[24]. Within stream (Figure 1C), homotopic areas were more correlated than adjacent areas in the visual hierarchy (e.g., left ventral V1 and right ventral

V2; $\Delta_{Fisher\,Z}$ M=0.09, p<0.001), and adjacent areas were more correlated than distal areas (e.g., left ventral V1 and right ventral V4; $\Delta_{Fisher\,Z}$ M=0.20, p<0.001). Hence, movies elicit distinct processing dynamics across areas of infant visual cortex defined independently using retinotopic mapping.

We previously found[17] that an anatomical segmentation of visual cortex[25] could identify these same areas reasonably well. Indeed, the results above were replicated when using visual areas defined anatomically (Figure S1). However, a key advantage of anatomical segmentation is that it can define visual areas not mapped by a functional retinotopy task. This could help address limitations of the analyses above, namely that there was a variable number of retinotopic areas identified across infants and these areas covered only part of visually responsive cortex. Focusing on broader areas that include portions of the ventral and dorsal stream in the adult visual cortex[24,25], we tested for functional differentiation of these streams in infants. We applied multi-dimensional scaling (MDS) — a data-driven method for assessing the clustering of data — to the average cross-correlation matrix across participants (Figure S1). The stress of fitting these data with a two-dimensional MDS was in the acceptable range (0.076). Clear organization was present (Figure 2): areas in the adult-defined ventral stream (e.g., VO, PHC) differentiated from areas in the adult-defined dorsal stream (e.g., V3A/B). Indeed, we see a slight separation between canonical dorsal areas and the recently defined lateral pathway[26] (e.g., LO1, hMT), although more evidence is needed to substantiate this distinction. Again, this organization cannot be attributed to mere spatial autocorrelation within stream because analyses were conducted across hemispheres (at significant anatomical distance). These results thus provide evidence of a dissociation in the functional profile

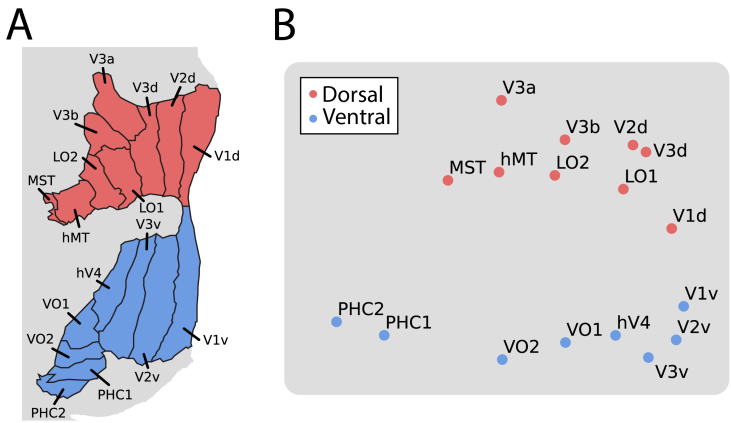

Figure 2: Multi-dimensional scaling (MDS) of movie-evoked activity in visual cortex. A) Anatomically defined areas[25] used for this analysis, separated into dorsal (red) and ventral (blue) visual cortex, overlaid on a flatmap of visual cortex. B) The timecourse of functional activity for each area was extracted and compared across hemispheres (e.g., left V1 was correlated with right V1). This matrix was averaged across participants and used to create a Euclidean dissimilarity matrix. MDS captured the structure of this matrix in two dimensions with suitably low stress. The plot shows a projection that emphasizes the similarity to the brain's organization.

of anatomically defined ventral and dorsal streams during infant movie-watching.

Evidence of within-area organization with independent components analysis We next explored whether movies can reveal fine-grained organization *within* visual areas by using independent components analysis (ICA) to discover visual maps in individual infant brains[22,27,28]. ICA is a method for decomposing a source into constituent signals by finding components that account for independent variance. When applied to fMRI data (using MELODIC in FSL), these components have spatial structure that varies in strength over time. Many of these components reflect noise (e.g., motion, breathing) or task-related signals (e.g., face responses), while other components reflect the functional architecture of the brain (e.g., topographic maps)[22,27]. We visually inspected each component and categorized it as a potential spatial frequency map, a potential meridian map,

or neither. This process was blind to the ground-truth of what the visual maps look like for that participant from the retinotopic mapping task, simulating what would be possible if retinotopy data from the participants were unavailable. Success in this process requires that 1) movies and ICA can reveal retinotopic organization and 2) experimenters can accurately identify these components.

Multiple maps could be identified because there were more than one candidate that the experimenter thought was a suitable map. Across infant participants, we identified an average of 2.4 (range: 0–5) components as potential spatial frequency maps and 1.1 (range: 0–4) components as potential meridian maps. To evaluate the quality of these maps, we compared them to the ground-truth of that participant's task-evoked maps (Figure 3). Spatial frequency and meridian maps are defined by their systematic gradients of intensity across the cortical surface[29]. Lines drawn parallel to area boundaries show monotonic gradients on spatial frequency maps, with stronger responses to high spatial frequency at the fovea, and stronger responses to low spatial frequencies in the periphery. By contrast, lines drawn perpendicular to the area boundaries show oscillations in sensitivity to horizontal and vertical meridians on meridian maps.

To assess the selected component maps, we correlated the gradients (described above) of the task-evoked and component maps. Figure 4A shows the absolute correlations between the task-evoked maps and the manually identified spatial frequency components (M=0.52, range: 0.23–0.85). To evaluate whether movies are a viable method for defining retinotopic maps, we tested whether the task-evoked retinotopic maps were more similar to manually identified components than other components. We identified the best component in 6 of 13 participants (Figure 4B).

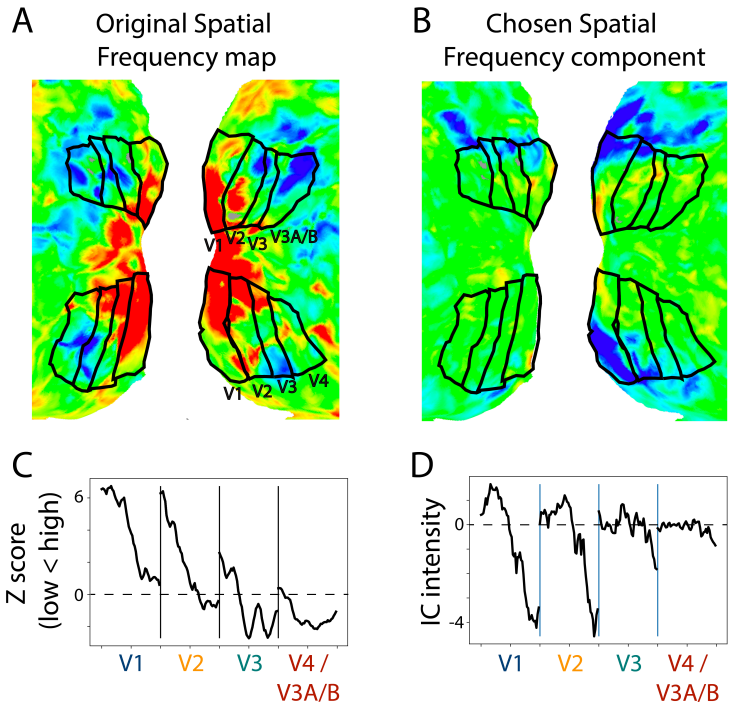

Figure 3: Example retintopic task vs. ICA-based spatial frequency maps. A) Spatial frequency map of a 17.1 month old infant. The retinotopic task data are from a prior study[17]. The view is of the flattened occipital cortex with visual areas traced in black. B) Component captured by ICA of movie data from the same participant. This component was chosen as a spatial frequency map in this participant. The sign of ICA is arbitrary so it was flipped here for visualization. C) Gradients in spatial frequency within-area from the task-evoked map in subpanel A. Lines parallel to the area boundaries (emanating from fovea to periphery) were manually traced and used to capture the changes in response to high versus low spatial frequency stimulation. D) Gradients in the component map. These gradients are similar to those found in the task-evoked spatial frequency map, confirming that this is an appropriate component.

The percentile of the average manually identified component was high (M=63.8 percentile, range: 26.7–98.1) and significantly above chance (ΔM=13.8, CI=[3.3–23.9], p=.010). This illustrates that the manually identified components derived from movie-watching data are similar to the spatial frequency maps derived from retinotopic mapping. This provides support for using data-driven techniques like ICA to reveal functional organization within visual areas in infants. The fact that

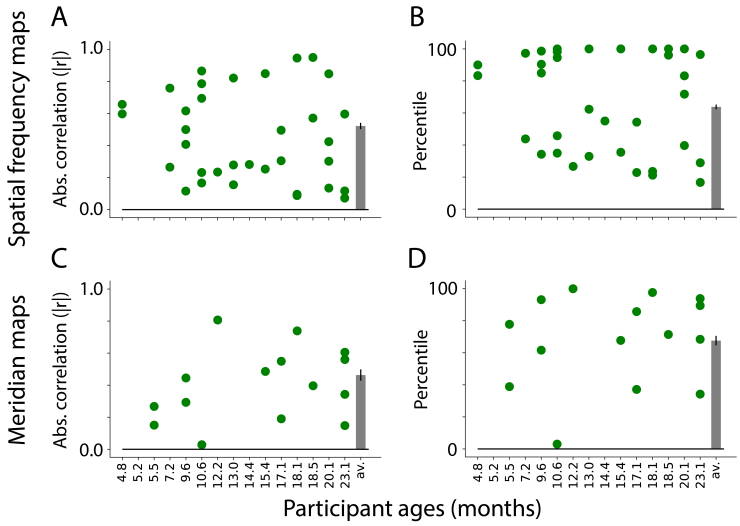

Figure 4: Similarity between visual maps from the retinotopy task and ICA applied to movies. A) Absolute correlation between the task-evoked and component spatial frequency maps (absolute values used because sign of ICA maps is arbitrary). Each dot is a manually identified component. At least one component was identified in 13 out of 15 participants. The bar plot is the average across participants. The error bar is the standard error across participants. B) Ranked correlations for the manually identified spatial frequency components relative to all components identified by ICA. Bar plot is same as A. C) Same as A but for meridian maps. At least one component was identified in 9 out of 15 participants. D) Same as B but for meridian maps.

this can work also indicates the underlying architecture of the infant visual system influences how movies are processed.

We performed the same analyses on the meridian maps. As noted above, the lines were now traced perpendicular to the boundaries. Figure 4C shows the correlation between the task-evoked meridian maps and the manually identified components (M=0.46, range: 0.03–0.81). Compared to all possible components identified by ICA, the best possible component was identified for 1 out of 9 participants (Figure 4D). Although the percentile of the average manually identified component was numerically high (M=67.6 percentile, range: 3.0–100.0), it was not significantly above

chance (ΔM=17.6, CI=[-1.8–33.0], p=.074). This difference in performance compared to spatial frequency is also evident in the fact that fewer components were identified as potential meridian maps, and that several participants had no such maps. Because it is possible, albeit less likely, to identify meridian maps from ICA, the structure may be present in the data but more susceptible to noise or gaze variability. Spatial frequency maps have a coarser structure than meridian maps, and are more invariant to fixation, which may explain why they are easier to identify.

Evidence of within area organization with shared response modeling Finally, we investigated whether the organization of visual cortex in one infant can be predicted from movie-watching data in *other* participants using functional alignment[30]. For such functional alignment to work, stimulus-driven responses to the movie must be shared across participants. These analyses also benefit from greater amounts of data, so we expanded the sample in two ways (Table S2): First, we added 71 movie-watching datasets from additional infants who saw the same movies but didn't have usable retinotopy data (and thus were not included in the analyses above that compared movie and retinotopy data within participant). Second, we recruited 8 adult participants to watch a subset of the movies we showed infants and to complete the retinotopy task that infants completed. Additionally, 41 datasets from adults who had seen the movies shown to infants but did not have retinotopy data were included in these analyses.

With this expanded dataset, we used shared response modeling (SRM)[31] to predict visual maps from other participants (Figure 5). Specifically, we held out one participant for testing purposes and used SRM to learn a low-dimensional, shared feature space from the movie-watching

data of the remaining participants in a mask of occipital cortex. This shared space represented the responses to that movie in visual cortex that were shared across participants, agnostic to the precise localization of these responses across voxels in each individual. The number of features in the shared space (K=10) was determined via a cross-validation procedure on movie-watching data in adults (Figure S2). We then mapped the held-out participant's movie data into the learned shared space without changing the shared space.

The task-evoked retinotopic maps from all but the held-out participant were transformed into this shared space and averaged, separately for each map type. Taking the inverse of the held-out participant's mapping allowed us to transform the averaged shared space representation of visual maps into the held-out participant's brain space. This predicted visual organization was compared to the participant's actual visual map from the retinotopy task (using the same methods as for ICA). Critically, predicting the retinotopic maps used no retinotopy data from the held-out participant. Moreover, it is completely unconstrained anatomically (except for a liberal occipital lobe mask). Hence, the similarity of the SRM-predicted map to the task-evoked map is due to representations of visual space in other participants being mapped into the shared space.

We trained SRMs on two populations to predict a held-out infant's maps: (1) other infants and (2) adults. There may be advantages to either approach: infants are likely more similar to each other than adults in terms of how they respond to the movie; however, their data is more contaminated by motion. When using the infants to predict a held out infant, the spatial frequency map (Figure 6A) and meridian map (Figure 6C) predictions are moderately correlated with task-

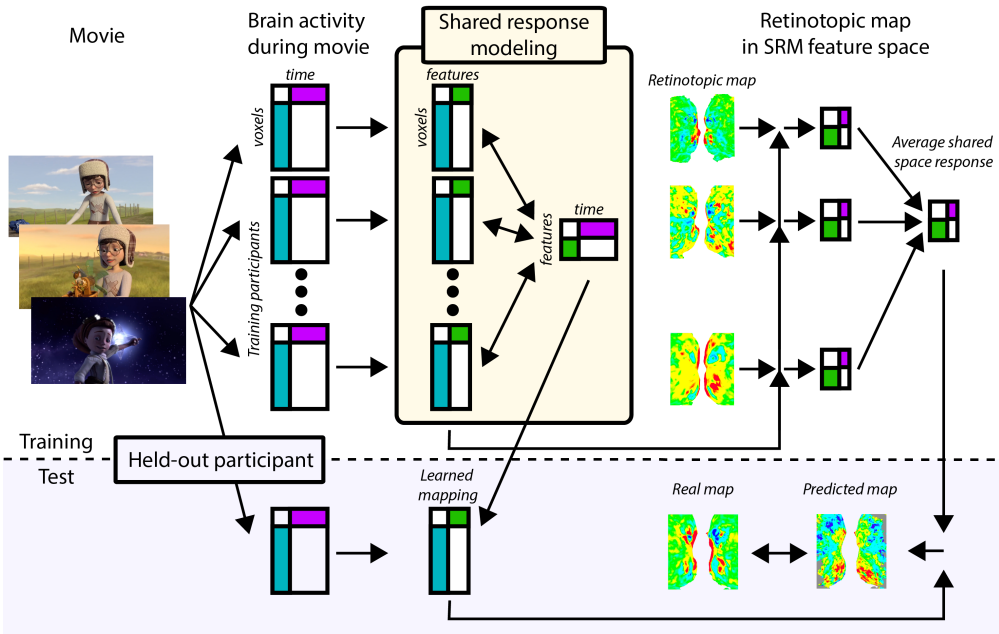

Figure 5: Pipeline for predicting visual maps from movie data. All participants watched the same movie, and one participant's data were held out. The remaining participants were mapped into a lower-dimensional feature space using shared response modeling (SRM)[31]. The visual maps from these participants were transformed into the shared space and averaged. By mapping the held-out participant into the shared space (with only their movie data), the average visual maps in shared space could be transformed into their brain space, resulting in a predicted retinotopic map that can be validated against their real map from the retinotopy task.

evoked retinotopy data (spatial frequency: M=0.46, range: -0.06-0.78; meridian: M=0.24, range: -0.12-0.78).

To evaluate whether success was due to fitting the shared response, we flipped the held-out participant's training data (i.e., the first timepoint became the last timepoint and vice versa) so that an appropriate fit should not be learnable. The vertical lines for each movie in Figure 6 indicate the change in performance for this baseline. Indeed, flipping significantly worsened prediction of the spatial frequency map ($\Delta_{Fisher\ Z}$ M=0.52, CI=[0.24–0.80], p<.001) and the meridian map ($\Delta_{Fisher\ Z}$

M=0.24, CI=[0.02–0.49], p=.034). Hence, the movie-evoked response enables the mapping of other infants' retinotopic maps into a held-out infant.

Using adult data to predict infant data also results in maps similar to task-evoked spatial frequency maps (Figure 6B; M=0.56, range: 0.17–0.79) and meridian maps (Figure 6D; M=0.34, range: -0.27-0.64). Again, flipping the SRM training data significantly worsened prediction of the held-out participant's spatial frequency map ($\Delta_{Fisher\ Z}$ M=0.40, CI=[0.17–0.65], p<.001) and meridian map ($\Delta_{Fisher\ Z}$ M=0.33, CI=[0.12–0.55], p=.002). There was no significant difference in SRM performance when using adults versus infants as the training set (spatial frequency: $\Delta_{Fisher\ Z}$ M=0.14, CI=[-0.00–0.27], p=.054; meridian: $\Delta_{Fisher\ Z}$ M=0.11, CI=[-0.05–0.28], p=.179). In sum, SRM could be used to predict visual maps with moderate accuracy. This indicates that functional alignment methods like SRM can partially capture the retinotopic organization of visual cortex from infant movie-watching data.

Discussion

We present evidence that movies can reveal the organization of infant visual cortex at different spatial scales. We found that movies evoke differential function across areas, topographic organization of function within areas, and this topographic organization is shared across participants.

We show that the movie-evoked response in a visual area is more similar to the same area in the other hemisphere than to different areas in the other hemisphere. By comparing across anatomically distant hemispheres, we reduced the impact of spatial autocorrelation and isolated

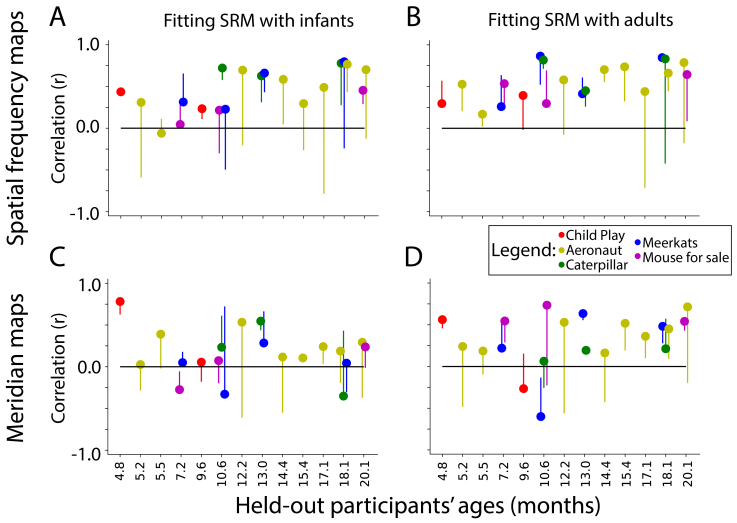

Figure 6: Similarity of SRM-predicted maps and task-evoked retinotopic maps. Correlation between the gradients of the A) spatial frequency maps and C) meridian maps predicted with SRM from other infants and task-evoked retinotopy maps. B, D) Same as A, except using adult participants to train the SRM and predict maps. Dot color indicates the movie used for fitting the SRM. The end of the line indicates the correlation of the task-evoked retinotopy map and the predicted map when using flipped training data for SRM. Hence, lines extending below the dot indicate that the true performance was higher than a baseline fit. The bar plot is the average across participants. The error bar is the standard error across participants.

the stimulus-driven signals in the brain activity[22,23,32]. The greater across-hemisphere similarity for same versus different areas provides some of the first evidence that visual areas and streams are functionally differentiated in infants as young as 5 months old. Previous work suggests that functions of the dorsal and ventral streams are detectable in young infants[33] but that the localization of these functions is immature[34]. Despite this, we find that the areas of infant visual cortex that will mature into the dorsal and ventral streams have distinct activity profiles during movie watching.

Additionally, we used a data-driven approach (ICA) to uncover the retinotopic organization of visual cortex in infants. We observed components that were highly similar to a spatial frequency

map obtained from the same infant in a retinotopy task. Importantly, the components could be identified without knowledge of these ground-truth maps. This was also true for the meridian maps, to a lesser degree. One caveat for interpreting these results is that although some of the components are *similar* to a spatial frequency map or meridian map, they could reflect a different kind of visual map. For instance, the spatial frequency map is highly correlated with the eccentricity map[16,35–37] (which itself is related to receptive field size). This means it is inappropriate to make strong claims about the underlying function of the components based on their similarity to visual maps alone. Nonetheless, these results do show that it is possible to discover visual maps in infants with movie-watching data and ICA.

We also asked whether functional alignment[20,31] could be used to detect visual maps in infants. Using a shared response model[31] trained on movie-watching data of infants or adults, we transformed the visual maps of other individuals into a held-out infant's brain to evaluate the fit to visual maps from a retinotopy task[30]. Like ICA, this was more successful for the spatial frequency maps, but it was still possible in some cases with the meridian maps. This is remarkable because the complex pattern of brain activity underlying these visual maps could be 'compressed' by SRM into only 10 dimensions in the shared space (i.e., the visual maps were summarized by a vector of 10 values). The weight matrix that 'decompressed' visual maps from this low-dimensional space into the held-out infant was learned from their movie-watching data alone. Hence, success with this approach means that visual maps are engaged during infant movie-watching. Furthermore, this result shows that functional alignment is practical for studies in awake infants that produce small amounts of data. Functional alignment can increase signal quality[30,31,38] and reveal chang-

ing function over development[39], so it may prove especially useful for infant fMRI[40]. In sum, movies evoke sufficiently reliable activity across infants and adults to find a shared response, and this shared response contains information about the organization of infant visual cortex.

The fact that movies are engaging to infants and serve multiple purposes[2,19] makes them an efficient approach for crude retinotopic mapping in infants, given the limited amount of awake data that can be obtained per session. To be clear, we are not suggesting that movies work well enough to *replace* a retinotopy task when accurate maps are needed. For instance, even though ICA found components that were highly correlated with the spatial frequency map, we also selected some components that turned out to have lower correlations. Without knowing the ground-truth from a retinotopy task, there would be no way to weed these out. On a positive note, our analyses provide additional support for our previous conclusion[17] that infant visual areas can be defined anatomically. Not only do visual areas have highly similar movie responses across hemispheres, but the function of visual areas showed interpretable structure when visualized with multi-dimensional scaling. Hence, some combination of anatomical landmarks and movie data may be adequate for defining infant visual areas for most research questions.

In conclusion, movies evoke activity in infants that recapitulate the organization of the visual cortex. This activity is differentiated across visual areas and contains information about the visual maps at the foundation of visual processing. The work presented here is another demonstration of the power of content-rich, dynamic, and naturalistic stimuli to reveal insights in cognitive neuroscience.

Methods

Participants Infant participants with retinotopy data were previously reported in another study[17]. Of those 17 original sessions, 15 had usable movie data collected in the same session and thus could be included in the current study. In this subsample, the age range was 4.8–23.1 months (M=13.0; 12 female; Table S1). The combinations of movies that infants saw were inconsistent, so the types of comparisons vary across analyses reported here. In brief, all possible infant participant sessions (15) were used in the Homotopy analyses and ICA, whereas two of these sessions (ages = 18.5, 23.1 months) could not be used in the SRM analyses. Table S1 reports demographic information for the infant participants. Table S2 reports participant information about each of the movies. It also reports the number and age of participants that were used to bolster the SRM analyses.

The adult sample used for SRM was collected to include both retinotopy and movie watching data (N=8, 3 females). The adult participants saw the five most common movies that were seen by infants in our retinotopy sample (3–8 total sessions per movie in the main infant sample). To support the SRM analyses, we also utilized any other available adult data from sessions in which we had shown the main movies in otherwise identical circumstances (Table S2).

Participants were recruited through fliers, word of mouth, or the Yale Baby School. This study was approved by the Human Subjects Committee at Yale University. Adults provided informed consent for themselves or their child.

Materials Our experiment display code can be found here: https://github.com/ntblab/experiment_menu/tree/Movies/ and https://github.com/ntblab/experiment_menu/tree/retinotopy/. The code used to perform the data analyses is available at https://github.com/ntblab/infant_neuropipe/tree/predict_retinotopy/; this code uses tools from the Brain Imaging Analysis Kit[41]; https://brainiak.org/docs/). Raw and preprocessed functional and anatomical data will be available here: https://datadryad.org/PENDING_URL but is available from this temporary link while under review: https://drive.google.com/drive/folders/1zKWLluNhUz48MZMAS8-I-xEINpLWYVo_?usp=sharing).

Data acquisition Data were collected at the Brain Imaging Center (BIC) in the Faculty of Arts and Sciences at Yale University. We used a Siemens Prisma (3T) MRI and only the bottom half of the 20-channel head coil. Functional images were acquired with a whole-brain T2* gradient-echo EPI sequence (TR=2s, TE=30ms, flip angle=71, matrix=64x64, slices=34, resolution=3mm iso, interleaved slice acquisition). Anatomical images were acquired with a T1 PETRA sequence for infants (TR1=3.32ms, TR2=2250ms, TE=0.07ms, flip angle=6, matrix=320x320, slices=320, resolution=0.94mm iso, radial slices=30000) and a T1 MPRAGE sequence for adults, with the top of the head coil attached, (TR=2300ms, TE=2.96ms, TI=900ms, flip angle=9, iPAT=2, slices=176, matrix=256x256, resolution=1.0mm iso).

Procedure Our approach for collecting fMRI data from awake infants has been described in a previous methods paper[1], with important details repeated below. Infants were first brought in for a mock scanning session to acclimate them and their parent to the scanning environment. Scans were scheduled when the infants were typically calm and happy. Participants were carefully screened for

metal. We applied hearing protection in three layers for the infants: silicon inner ear putty, over-ear adhesive covers and ear muffs. For the infants that were played sound (see below), Optoacoustics noise cancelling headphones were used instead of the ear muffs. The infant was placed on a vacuum pillow on the bed that comfortably reduced their movement. The top of the head coil was not placed over the infant in order to maintain comfort. Stimuli were projected directly on to the surface of the bore. A video camera (High Resolution camera, MRC systems) recorded the infant's face during scanning. Adult participants underwent the same procedure with the following exceptions: they did not attend a mock scanning session, hearing protection was only two layers (earplugs and Optoacoustics headphones), and they were not on a vacuum pillow. Some infants participated in additional tasks during their scanning session.

When the infant was focused, experimental stimuli were shown using Psychtoolbox[42] for MATLAB. The details for the retinotopy task are explained fully elsewhere[17]. In short, we showed two types of blocks. For the meridian mapping blocks, a bow tie cut-out of a colorful, large, flickering checkerboard was presented in either a vertical or horizontal orientation[43]. 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[22]. For all blocks, a smaller (1.5 visual degree) grayscale movie was played at center to encourage fixation. Each block type contained two phases of stimulation. The first phase consisted of one of the conditions (e.g., horizontal or high) for 20s, followed immediately by the second phase with the other condition of the same block type (e.g., vertical or low, respectively) for 20s. At the end of each block there was at least 6s rest before the start of the next block. Infant participants saw up to 12 blocks of this

stimulus, resulting in 24 epochs of stimuli. Adults all saw 12 blocks.

Participants saw a broad range of movies in this study (Table S3), some of which have been reported previously[19,21]. The movie titled 'Child Play' comprises the concatenation of four silent videos that range in duration from 64–143s and were shown in the same order (with 6s in-between). They extended 40.8° wide by 25.5° high on the screen. The other movies were stylistically similar, computer-generated animations that each lasted 180s. These movies extended 45.0° wide by 25.5° high. Some of the movies were collected as part of an unpublished experiment in which we either played the full movie or inserted drops every 10s (i.e., the screen went blank while the audio continued). We included the 'Dropped' movies in the Homotopy analyses and ICA (average number of 'Dropped' movies per participant: 0.9, range: 0–3); however, we did not include them in the SRM analyses. Moreover, we only included 4 (out of 17) of these movies in the SRM analyses because there were insufficient numbers of infant participants to enable the training of the SRM.

Gaze Coding The infant gaze coding procedure for the retinotopy data was the same as reported previously[17]. The gaze coding for the movies was also the same as reported previously[19,21]. Participants looked at the screen for an average of 93.7% of the time (range: 78–99) for the movies used in the homotopy and ICA analyses, and 94.5% of the time (range: 82–99) for the movies used in the SRM analyses (Table S1) Adult participants were not gaze coded, but they were monitored online for inattentiveness. One adult participant was drowsy so they were manually coded. This resulted in the removal of four out of the 24 epochs of retinotopy.

Preprocessing We used FSL's FEAT analyses with modifications in order to implement infant-specific preprocessing of the data[1]. If infants participated in other experiments during the same functional run (14 sessions), the data was split to create a pseudorun. Three burn-in volumes were discarded from the beginning of each run/pseudorun when available. To determine the reference volume for alignment and motion correction, the Euclidean distance between all volumes was calculated and the volume that minimized the distance between all points was chosen as reference (the 'centroid volume'). Adjacent timepoints with greater than 3mm of movement were interpolated. To create the brain mask we calculated the SFNR[44] for each voxel in the centroid volume. This produced a bimodal distribution reflecting the signal properties of brain and non-brain voxels. We thresholded the brain voxels at the trough between these two peaks. We performed Gaussian smoothing (FWHM=5mm). Motion correction with 6 degrees of freedom was performed using the centroid volume. AFNI's despiking algorithm attenuated voxels with aberrant timepoints. The data for each movie were z-scored in time.

We registered the centroid volume to a homogenized and skull-stripped anatomical volume from each participant. Initial alignment was performed using FLIRT with a normalized mutual information cost function. This automatic registration was manually inspected and then corrected if necessary using mrAlign from mrTools[45].

The final step common across analyses created a transformation into surface space. Surfaces were reconstructed from iBEAT v2.0[46]. These surfaces were then aligned into standard Buckner40 standard surface space[47] using FreeSurfer[47].

Additional preprocessing steps were taken for the SRM analyses. For each individual movie (including each movie that makes up 'Child Play'), the fMRI data was time-shifted by 4s and the break after the movie finished was cropped. This was done to account for hemodynamic lag, so that the first TR and last TR of the data approximately[48] corresponded to the brain's response to the first and last 2s of the movie, respectively.

Occipital masks were aligned to the participant's native space for the SRM analyses. To produce these, a mapping from native functional space to standard space was determined. This was enabled using non-linear alignment of the anatomical image to standard space using ANTs[49]. For infants, an initial linear alignment with 12 DOF was used to align anatomical data to the age-specific infant template[50], followed by non-linear warping using diffeomorphic symmetric normalization. Then, we used a predefined transformation (12 DOF) to linearly align between the infant template and adult standard. For adults, we used the same alignment procedure, except participants were directly aligned to adult standard. We used the occipital mask from the MNI structural atlas[51] in standard space and used the inverted transform to put it into native functional space.

Analysis

Retinotopy For our measure of task-evoked retinotopy in infants, we used the outputs of the retinotopy analyses from our previous paper[17] that are publicly released. In brief, we performed separate univariate contrasts between conditions in the study (horizontal>vertical, high spatial frequency>low spatial frequency). We then mapped these contrasts into surface space. Then, in surface space rendered by AFNI[52], we demarcated the visual areas V1, V2, V3, V4, and V3A/B

using traditional protocols based on the meridian map contrast[53]. We traced lines perpendicular and parallel to the area boundaries to quantify gradients in the visual areas. The anatomically-defined areas of interest[25] used in Figure 2 were available in this standard surface space.

Homotopy The homotopy analyses compared the time course of functional activity across visual areas in different hemispheres of each infant. For the participants that had more than one movie in a session (N=9), all the movies were concatenated along with burn out time between the movies (Mean number of movies per participant=2.7, range: 1–6, Mean duration of movies=540.7s, range: 186–1116). For the areas that were defined with the retinotopy task (average number of areas traced in each hemisphere = 7.3, range: 6.0–8.0), the functional activity was averaged within area and then Pearson correlated between all other areas. The resulting cross-correlation matrix was Fisher Z transformed before different cells were averaged or compared. If infants did not have an area traced then those areas were ignored in the analyses. We grouped visual areas according to stream, where areas that are *more* dorsal of V1 were called 'dorsal' stream and areas *more* ventral were called 'ventral' stream. To assess the functional similarity of visual areas, Fisher Z correlations between the same areas in the same stream were averaged, and compared to the correlations of approximately equivalent areas from different streams (e.g., dorsal V2 compared with ventral V2). The averages for each of the two conditions (same stream vs. different stream) were evaluated statistically using bootstrap resampling[54]. Specifically, we computed the mean difference between conditions in a pseudosample, generated by sampling participants with replacement. We created 10,000 such pseudosamples and took the proportion of differences that showed a different sign than the true mean, multiplied by two to get the two-tailed p-value. To evaluate how distance affects

similarity, we additionally compared with bootstrap resampling the Fisher Z correlations of areas across hemispheres in the same stream: same area to adjacent areas (e.g., ventral V1 with ventral V2), to distal areas (e.g., ventral V1 with ventral V3). Before reporting the results in the figures, the Fisher Z values were converted back into Pearson correlation values.

As an additional analysis to the one described above, we used an atlas of anatomically-defined visual areas from adults[25] to define both early and later visual areas. Specifically, we used the areas labeled as part of the ventral and dorsal stream (excluding the intraparietal sulcus and frontal eye fields since they often cluster separately[55]), and then averaged the functional response within each area. The functional responses were then correlated across hemispheres, as in the main analysis. Multi-dimensional scaling was then performed on the cross-correlation matrix, and the dimensionality that fell below the threshold for stress (0.2) was chosen. In this case, that was a dimensionality of 2 (stress=0.076). We then visualized the resulting output of the data in these two dimensions.

Independent Components Analysis (ICA) To conduct ICA, we provided the preprocessed movie data to FSL's MELODIC[27]. Like in the homotopy analyses, we used all of the movie data available per session. The algorithm found a range of components across participants (M=76.4 components, range: 31–167). With this large number of possible components, an individual coder (CE) sorted through them to determine whether each one looked like a meridian map, spatial frequency map, or neither (critically, without referring to the ground truth from the retinotopy task). We initially visually inspected each component in volumetric space, looking for the following features: First,

we searched for whether there was a strong weighting of the component in visual cortex. Second, we looked for components that had a symmetrical pattern in visual cortex between the two hemispheres. To identify the spatial frequency maps, we looked for a continuous gradient emanating out from the early visual cortex. For meridian maps, we looked for sharp alternations in the sign of the component, particularly near the midline of the two hemispheres. Based on these criteria, we then chose a small set of components that were further scrutinized in surface space. On the surface, we looked for features that clearly define a visual map topography. Again, this selection process was blind to the task-evoked retinotopic maps, so that a person without retinotopy data could take the same steps and potentially find maps.

These components were then tested against that participant's task-evoked retinotopic maps. If the component was labeled as a potential spatial frequency map, we tested whether there was a monotonic gradient from fovea to periphery. Specifically, we measured the component response along lines drawn parallel to the area boundaries, averaged across these lines, and then correlated this pattern with the same response in the actual map. The absolute correlation was used because the sign of ICA is arbitrary. For each participant, we then ranked the components to ask if the ones that were chosen were the best ones possible out of all those derived from MELODIC. To test whether the identified components were better than the non-identified components, we ranked all the components correlation to the task-evoked maps. This ranking was converted into a percentile, where 100% means it is the best possible component. We took the identified component's percentile (or averaged the percentiles if there were multiple components chosen) and compared it to chance (50%). This difference from chance was used for bootstrap resampling to evaluate whether

the identified components were significantly better than chance. We performed the same kind of analysis for meridian maps, except in this case the lines used for testing were those drawn perpendicular to the areas. In this case, we were testing whether the components showed oscillations in the sign of the intensity.

Shared Response Modeling (SRM) We based our SRM analyses on previous approaches[30]. SRM embeds the brain activity of multiple individuals viewing a common stimulus into a shared space with a small number of feature dimensions. Each voxel of each participant is assigned a weight for each feature. The weight reflects how much the voxel loads onto that feature. For our study, the SRM was either trained on infant movie-watching data or adult movie-watching data to learn the shared response, and the mapping of the training participants into this shared space. For the infant SRM, we used a leave-one-out approach. We took a movie that the held-out infant saw (e.g., 'Aeronaut') and considered all other infant participants that saw that movie (including additional participants without any retinotopy data). We fit an SRM model on all of the participants except the held-out one. This model has 10 features, as was determined based on cross-validation with adult data (Figure S2). We used an occipital anatomical mask to fit the SRM. Using the learned individual weight matrices, the retinotopic maps from the infants in the training set were then transformed into the shared space and averaged across participants. The held-out participant's movie data were used to learn a mapping to the learned SRM features. By applying the inverse of this mapping, we transformed the averaged visual maps of the training set in shared space into the brain space of the held-out participant to predict their visual maps. Using the same methods as described for ICA above, we compared the task-evoked and predicted gradient responses. These

analysis steps were also followed for the adult SRM, with the difference being that the group of participants used to create the SRM model and to create the averaged visual maps were adults. As with the infant SRM, additional adult participants without retinotopy data were used for training. Across both types of analysis, the held-out participant was completely ignored when fitting the SRM, and no retinotopy data went into training the SRM.

To test the benefit of SRM, we performed a control analysis in which we scrambled the movie data from the held-out participant before learning their mapping into the shared space. Specifically, we flipped the timecourse of the data so that the first timepoint became the last, and vice versa. By creating a mismatch in the movie sequence across participants, this procedure should result in meaningless weights for the held-out participant and, in turn, the prediction of visual maps using SRM will fail. We compared 'real' and 'flipped' SRM procedures by computing the difference in fit (transformed into Fisher Z) for each movie, and then averaging that difference within participant. Those differences were then bootstrap resampled to evaluate significance. We also performed bootstrap resampling to compare the 'real' SRM accuracy when using infants versus adults for training.

Contributions C.T.E. and M.J.A. conceived of the analyses. C.T.E., T.S.Y., & N.B.T-B. collected the data. C.T.E. & T.S.Y. preprocessed the data. C.T.E. performed the analyses. All authors contributed to the drafting of the manuscript.

Acknowledgements We are thankful to the families of infants who participated. We also acknowledge the hard work of the Yale Baby School team, including L. Rait, J. Daniels, A. Letrou, and K. Armstrong for recruitment, scheduling, and administration, and L. Skalaban, A. Bracher, D. Choi, and J. Trach for help in infant fMRI data collection. Thank you to J. Wu, J. Fel, and A. Klein for help with gaze coding, and R. Watts for technical support. We are grateful for internal funding 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 and the James S. McDonnell Foundation (https://doi.org/10.37717/2020-1208).

References

1. Ellis, C. T. *et al.* Re-imagining fmri for awake behaving infants. *Nature Communications* **11** (2020).

2. Vanderwal, T., Eilbott, J. & Castellanos, F. X. Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience* **36**, 100600 (2019).

3. Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C. & Castellanos, F. X. Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage* **122**, 222–232 (2015).

4. Richardson, H., Lisandrelli, G., Riobueno-Naylor, A. & Saxe, R. Development of the social brain from age three to twelve years. *Nature Communications* **9**, 1027 (2018).

5. Alexander, L. M. *et al.* An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data* **4**, 1–26 (2017).

6. Nastase, S. A., Goldstein, A. & Hasson, U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage* **222**, 117254 (2020).

7. Finn, E. S., Glerean, E., Hasson, U. & Vanderwal, T. Naturalistic imaging: The use of ecologically valid conditions to study brain function. *NeuroImage* **247** (2022).

8. Franchak, J. M., Heeger, D. J., Hasson, U. & Adolph, K. E. Free viewing gaze behavior in infants and adults. *Infancy* **21**, 262–287 (2016).

9. Tran, M., Cabral, L., Patel, R. & Cusack, R. Online recruitment and testing of infants with mechanical turk. *Journal of Experimental Child Psychology* **156**, 168–178 (2017).

10. Brodmann, K. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* (Barth, 1909).

11. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, NY: 1991)* **1**, 1–47 (1991).

12. Kaas, J. H. Topographic maps are fundamental to sensory processing. *Brain research bulletin* **44**, 107–112 (1997).

13. White, L. E. & Fitzpatrick, D. Vision and cortical map development. *Neuron* **56**, 327–338 (2007).

14. Fox, P. T., Miezin, F. M., Allman, J. M., Van Essen, D. C. & Raichle, M. E. Retinotopic organization of human visual cortex mapped with positron-emission tomography. *Journal of Neuroscience* **7**, 913–922 (1987).

15. Schneider, W., Noll, D. C. & Cohen, J. D. Functional topographic mapping of the cortical ribbon in human vision with conventional mri scanners. *Nature* **365**, 150–153 (1993).

16. Henriksson, L., Nurminen, L., Hyvärinen, A. & Vanni, S. Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision* **8**, 1–13 (2008).

17. Ellis, C. T. *et al.* Retinotopic organization of visual cortex in human infants. *Neuron* **109**, 2616–2626 (2021).

18. Loiotile, R. E., Cusack, R. & Bedny, M. Naturalistic audio-movies and narrative synchronize "visual" cortices across congenitally blind but not sighted individuals. *Journal of Neuroscience* **39**, 8940–8948 (2019).

19. Yates, T. S. *et al.* Neural event segmentation of continuous experience in human infants. *Proceedings of the National Academy of Sciences* **119**, e2200257119 (2022).

20. Turek, J. S., Ellis, C. T., Skalaban, L. J., Turk-Browne, N. B. & Willke, T. L. Capturing shared and individual information in fmri data. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 826–830 (IEEE, 2018).

21. Yates, T. S., Ellis, C. T. & Turk-Browne, N. B. Functional networks in the infant brain during sleep and wake states. *bioRxiv* 2023–02 (2023).

22. Arcaro, M. J. & Livingstone, M. S. A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife* **6**, e26196 (2017).

23. Li, M. *et al.* Development of visual cortex in human neonates is selectively modified by postnatal experience. *eLife* **11**, e78733 (2022).

24. Ungerleider, L. G. & Mishkin, M. Two cortical visual systems. In Ingle D, M. R., Goodale MA (ed.) *Analysis of visual behavior*, 549–586 (MIT Press, Cambridge, 1982).

25. Wang, L., Mruczek, R. E., Arcaro, M. J. & Kastner, S. Probabilistic maps of visual topography in human cortex. *Cerebral Cortex* **25**, 3911–3931 (2014).

26. Weiner, K. S. & Gomez, J. Third visual pathway, anatomy, and cognition across species. *Trends in Cognitive Sciences* **25**, 548–549 (2021).

27. Beckmann, C. F., DeLuca, M., Devlin, J. T. & Smith, S. M. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1001–1013 (2005).

28. Lu, K.-H., Jeong, J. Y., Wen, H. & Liu, Z. Spontaneous activity in the visual cortex is organized by visual streams. *Human Brain Mapping* **38**, 4613–4630 (2017).

29. Arcaro, M. J., McMains, S. A., Singer, B. D. & Kastner, S. Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience* **29**, 10638–10652 (2009).

30. Guntupalli, J. S. *et al.* A model of representational spaces in human cortex. *Cerebral cortex* **26**, 2919–2934 (2016).

31. Chen, P.-H. *et al.* A reduced-dimension fmri shared response model. In *NIPS*, vol. 28, 460–468 (2015).

32. Smyser, C. D. *et al.* Longitudinal analysis of neural network development in preterm infants. *Cerebral Cortex* **20**, 2852–2862 (2010).

33. Wattam-Bell, J. *et al.* Reorganization of global form and motion processing during human visual development. *Current Biology* **20**, 411–415 (2010).

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

35. Smith, A. T., Singh, K. D., Williams, A. & Greenlee, M. W. Estimating receptive field size from fmri data in human striate and extrastriate visual cortex. *Cerebral Cortex* **11**, 1182–1190 (2001).

36. Srihasam, K., Vincent, J. L. & Livingstone, M. S. Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature Neuroscience* **17**, 1776–1783 (2014).

37. Tolhurst, D. & Thompson, I. On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **213**, 183–199 (1981).

38. Busch, E. L. *et al.* Hybrid hyperalignment: A single high-dimensional model of shared information embedded in cortical patterns of response and functional connectivity. *NeuroImage* **233**, 117975 (2021).

39. Yates, T. S., Ellis, C. T. & Turk-Browne, N. B. Emergence and organization of adult brain function throughout child development. *NeuroImage* **226**, 117606 (2021).

40. Ellis, C. T. & Turk-Browne, N. B. Infant fmri: A model system for cognitive neuroscience. *Trends in Cognitive Sciences* **22**, 375–387 (2018).

41. Kumar, M. *et al.* BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis. *PLOS Computational Biology* **16**, e1007549 (2020).

42. Kleiner, M. *et al.* What's new in psychtoolbox-3. *Perception* **36**, 1 (2007).

43. Tootell, R. *et al.* Functional analysis of human mt and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience* **15**, 3215–3230 (1995).

44. Friedman, L. & Glover, G. H. Report on a multicenter fmri quality assurance protocol. *Journal of Magnetic Resonance Imaging* **23**, 827–839 (2006).

45. Gardner, J., Merriam, E., Schluppeck, D., Besle, J. & Heeger, D. mrtools: Analysis and visualization package for functional magnetic resonance imaging data. *Zenodo* **10** (2018).

46. Wang, L. *et al.* ibeat v2. 0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction. *Nature Protocols* 1–32 (2023).

47. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage* **9**, 179–194 (1999).

48. Poppe, T., Moore, J. W. & Arichi, T. Individual focused studies of functional brain development in early human infancy. *Current Opinion in Behavioral Sciences* **40**, 137–143 (2021).

49. Avants, B. B. *et al.* A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage* **54**, 2033–2044 (2011).

50. Fonov, V. *et al.* Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* **54**, 313–327 (2011). URL http://www.sciencedirect.com/science/article/pii/S1053811910010062.

51. Mazziotta, J. *et al.* A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **356**, 1293–1322 (2001).

52. Cox, R. W. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research* **29**, 162–173 (1996).

53. Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).

54. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 54–75 (1986).

55. Haak, K. V. & Beckmann, C. F. Objective analysis of the topological organization of the human cortical visual connectome suggests three visual pathways. *Cortex* **98**, 73–83 (2018).

Supplementary materials

Movies reveal the fine-grained organization of infant visual cortex

Ellis, C. T., Yates, T. S., Arcaro, M. J., & Turk-Browne, N. B.

Table S1: Demographic and dataset information for infant participants in the study. 'Age' is recorded in months. 'Sex' is the assigned sex at birth. 'Retinotopy areas' is the number of areas segmented from task-evoked retinotopy, averaged across hemispheres. Information about the movie data is separated based on analysis type: whereas all movie data is used for homotopy and ICA analyses, a subset of data is used for SRM. 'Num.' is the number of movies used. 'Length' is the duration in seconds of the run used for these analyses (includes both movie and rest periods). 'Drops' is the number of movies that include dropped periods. Runs' says how many runs or pseudoruns of movie data there were. 'Gaze' is the percentage of the data where the participants were looking at the movie.

ID	Age	Sex	Retinotopy	Homotopy and ICA					SRM	
			Areas	Num.	Length	Drops	Runs	Gaze	Num.	Gaze
s2077_1_1	4.8	M	6.0	1	430	0	1	97	1	97
s2097_1_1	5.2	M	8.0	1	186	0	1	96	1	96
s4047_1_1	5.5	F	7.5	1	186	0	1	99	1	99
s7017_1_3	7.2	F	7.0	4	744	2	2	97	2	98
s7047_1_1	9.6	F	7.0	1	432	0	1	91	1	91
s7067_1_4	10.6	F	7.5	6	1110	3	2	98	3	99
s8037_1_2	12.2	F	7.5	1	186	0	1	95	1	95
s4607_1_4	13.0	F	7.0	3	558	1	1	93	2	90
s1607_1_4	14.4	M	6.0	1	372	0	2	93	1	93
s6687_1_4	15.4	F	8.0	1	186	0	1	82	1	82
s8687_1_5	17.1	F	8.0	1	186	0	1	98	1	98
s6687_1_5	18.1	F	8.0	5	930	2	2	94	3	92
s4607_1_7	18.5	F	7.5	4	744	2	2	78	0	NaN
s6687_1_6	20.1	F	6.5	6	1116	2	3	97	2	98
s8687_1_8	23.1	F	7.5	4	744	2	1	97	0	NaN
Mean	13.0	.	7.3	2.7	540.7	0.9	1.5	93.7	1.3	94.5

Table S2: Number of participants per movie. The first column is the movie name, where 'Drop-' indicates that it was a movie containing alternating epochs of blank screens. 'SRM' indicates whether the movie is used in SRM analyses. The movies that are not included in SRM are used for homotopy and ICA. 'Ret. infants' and 'Ret. adults' refers to the number of participants with retinotopy data that saw this movie. 'Infant SRM' and 'Adult SRM' refer to the number of additional participants available to use for training the SRM but who did not have retinotopy data. 'Infant Ages' is the average age in months of the infant participants included in the SRM, with the range of ages included in parentheses.

Movie name	SRM	Ret. infants	Ret. adults	Infant SRM	Infant Ages	Adult SRM
Child_Play	1	2	8	20	13.7 (3.3–32.0)	9
Aeronaut	1	8	8	35	10.1 (3.6–20.1)	32
Caterpillar	1	3	8	6	13.0 (6.6–18.2)	0
Meerkats	1	4	8	6	13.4 (7.2–18.2)	0
Mouseforsale	1	3	8	4	14.7 (7.2–20.1)	0
Elephant	0	1	0	0		0
MadeinFrance	0	1	0	0		0
Clocky	0	1	0	0		0
Gopher	0	1	0	0		0
Foxmouse	0	2	0	0		0
Drop-Caterpillar	0	4	0	0		0
Drop-Meerkats	0	3	0	0		0
Drop-Mouseforsale	0	1	0	0		0
Drop-Elephant	0	1	0	0		0
Drop-MadeinFrance	0	1	0	0		0
Drop-Clocky	0	2	0	0		0
Drop-Ballet	0	1	0	0		0
Drop-Foxmouse	0	1	0	0		0

Table S3: Details for each movie used in this study.'Name' specifies the movie name. 'Duration' specifies the duration of the movie in seconds. Movies were edited to standardize length and remove inappropriate content. 'Sound' is whether sound was played during the movie. These sounds include background music, animal noises, and sound effects, but no language. 'Description' gives a brief description of the movie, as well as a current link to it when appropriate. All movies are provided in the data release.

Name	Duration	Sound	Description
Child_Play	406	0	Four photo-realistic clips from 'Daniel Tiger' showing children playing. The clips showed the following: 1. children playing in an indoor playground (84s); 2. a family making frozen banana desserts (64s); 3. a child visiting the doctor (115s); 4. children helping with indoor and outdoor chores (143s).
Aeronaut	180	0	A computer-generated segment from a short film titled "Soar" (https://vimeo.com/148198462) and described here[19].
Caterpillar	180	1	A computer-generated segment from a short film titled "Sweet Cocoon" (https://www.youtube.com/watch?v=yQ1ZcNpbwOA). This video depicts a caterpillar trying to fit into its cocoon so it can become a butterfly.
Meerkats	180	1	A computer-generated segment from the short film titled "Catch It" (https://www.youtube.com/watch?v=c88QE6yGhfM). It depicts a gang of meerkats who take back a treasured fruit from a vulture.
Mouse for Sale	180	1	A computer-generated segment from a short film of the same name (https://www.youtube.com/watch?v=UB3nKCNUBB4). It shows a mouse in a pet store who is teased for having big ears.
Elephant	180	1	A computer-generated segment from a short film of the same name (https://www.youtube.com/watch?v=h_aC8pGY1aY). It shows an elephant in a china shop.
Made in France	180	1	A computer-generated segment from a short film of the same name (https://www.youtube.com/watch?v=Her3d1DH7yU). It shows a mouse making cheese.
Clocky	180	1	A computer-generated segment from a short film of the same name (https://www.youtube.com/watch?v=8VRD5KOFK94). It shows a clock preparing to wake up its owner.
Gopher	180	1	A computer-generated segment named 'Gopher broke' (https://www.youtube.com/watch?v=tWufIUbXubY). It shows a gopher collecting food.
Foxmouse	180	1	A computer-generated segment named 'The short story of a fox and a mouse' (https://www.youtube.com/watch?v=k6kCwj0Sk4s). It shows a fox playing with a mouse in the snow.
Ballet	180	1	A computer-generated segment named 'The Duet' (https://www.youtube.com/watch?v=GuX52wkCIJA). This is an artistic rendition of growing up and falling in love.

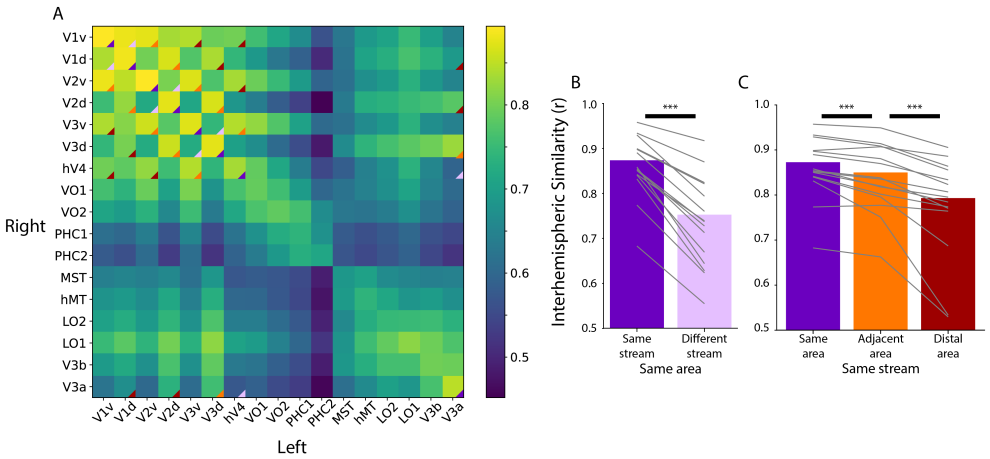

Figure S1: Homotopic correlations between anatomically defined areas. A) Average correlation of the time course of activity evoked during movie watching for ventral and dorsal areas in an anatomical segmentation[25]. This is done for the left and right hemispheres separately, which is why the matrix is not diagonally symmetric. The triangles overlaid on the matrix corner highlights the area-wise comparisons used in B and C. Only areas that we were able to retinotopically mapped (i.e., those that overlap with Figure 1) were used for this analysis. B) Correlation of the same area and same stream (e.g., left ventral V1 and right ventral V1) versus the same area and different stream (e.g., left ventral V1 and right dorsal V1). Difference with bootstrap resampling: $\Delta_{Fisher\ Z}$ M=0.37, p<0.001. C) Correlation within the same stream between the same areas, adjacent areas (e.g., left ventral V1 and right ventral V2), or distal areas (e.g., left ventral V1 and right ventral V4). Difference with bootstrap resampling: Same > Adjacent $\Delta_{Fisher\ Z}$ M=0.09, p<0.001; Adjacent > Distal $\Delta_{Fisher\ Z}$ M=0.18, p<0.001. Grey lines represent individual participants. *** = p<0.001 from bootstrap resampling

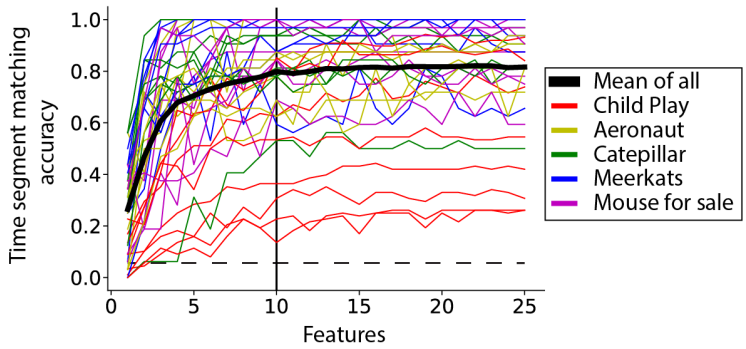

Figure S2: Cross-validation of the number of features in SRM. The movie data from all adult participants (Table S2) was split in half, with a 10 TR buffer between sets. The data were masked only to include occipital lobe voxels. The first half of the movie was used for training the SRM in all but one participant. The number of features learned by the SRM was varied across analyses from 1–25. The second half of the movie was then used to generate a shared response (i.e., the activity time course in each feature). To test the SRM, the held-out participant's first half of data is used to learn a mapping of that participant into the SRM space (this mapping does not change the features learned and is not based on the second half of data). The second half of the held-out participant's data is then mapped into the shared response space, like the other participants. Time-segment matching was performed on the shared response[20,31]. In brief, time-segment matching tests whether a segment of the data (10 TRs) in the held-out participant can be matched to its correct time point based on the other participants. This tests whether the SRM succeeds in making the held-out participant similar to the others. This analysis was performed on each participant and movie separately (each has a line). The dashed line is chance for time-segment matching, averaged across all movies and participants. The black solid line at features=10 reflects the number of features chosen.