

Movies reveal the fine-grained organization of infant visual cortex

Cameron T Ellis , Tristan S Yates, Michael J Arcaro, Nicholas B Turk-Browne

Department of Psychology, Stanford University, Palo Alto, United States • Department of Psychology, Yale University, New Haven, United States • Department of Psychology, University of Pennsylvania, Philadelphia, United States • Wu Tsai Institute, Yale University, New Haven, United States

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Reviewed Preprint

v3 • January 7, 2025

Revised by authors

Reviewed Preprint

v2 • September 25, 2024

Reviewed Preprint

v1 • December 19, 2023

eLife Assessment

This study presents **valuable** evidence concerning the potential for naturalistic movie-viewing fMRI experiments to reveal some features that are correlated with the functional and topographical organization of the developing visual system in awake infants and toddlers. The data are **compelling** given the difficulty of studying this population, the methodology is original and validated, and the evidence supporting the conclusions is **convincing** and in line with prior research using resting-state and awake task-based fMRI. This study will be of interest to cognitive neuroscientists and developmental psychologists, and in particular those interested in using fMRI to investigate brain organisation in pediatric and clinical populations with limited tolerance to fMRI.

<https://doi.org/10.7554/eLife.92119.3.sa2>

Abstract

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 and toddlers 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, albeit imprecisely, 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.

Studying the function and organization of the youngest human brains remains a challenge. Despite the recent growth in infant fMRI^{1–6}, one of 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 tasks⁷. Movies can be a useful tool for studying the developing mind⁸, as has been shown in older children^{9–11}. The dynamic, continuous, and content-rich nature of movie stimuli^{12, 13} make them effective at capturing infant attention^{14, 15}. 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 between areas^{16–18}. Within visual areas, there are orderly, topographic representations, or maps, of visual space^{19, 20}. 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^{21, 22}, and sensitivity to spatial frequency, reflected in gradients of sensitivity to high and low spatial frequencies from foveal to peripheral vision, respectively²³. Previously, we reported that these maps could be revealed by a retinotopy task in infants as young as 5 months of age²⁴. However, it remains unclear whether these maps are evoked by more naturalistic task-designs.

The primary goal of the current study is to investigate whether movie-watching data re-capitulates the organization of visual cortex. Movies drive strong and naturalistic responses in sensory regions while minimizing task demands^{12, 13, 25} and thus are a proxy for typical experience. In adults, movies and resting-state data have been used to characterize the visual cortex in a data-driven fashion^{26–28}. Movies have been useful in awake infant fMRI for studying event segmentation²⁹, functional alignment³⁰, and brain networks³¹. However, this past work did not address the granularity and specificity of cortical organization that movies evoke. For example, movies evoke similar activity across infants in anatomically aligned visual areas²⁹, but it remains unclear whether responses to movie content differ *between* visual areas (e.g., is there more similarity of function within visual areas than between³²). Moreover, it is unknown whether structure *within* visual areas, namely visual maps, contributes substantially to visual evoked activity. Additionally, we wish to test whether methods for functional alignment can be used with infants. Functional alignment finds a mapping between participants using functional activity – rather than anatomy – and in adults can improve signal-to-noise, enhance across participant prediction, and enable unique analyses^{28, 33–35}.

Nonetheless, 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. If movies do show the fine-grained organization of the infant visual cortex, this suggests that this structure (e.g., visual maps) scaffolds the processing of ongoing visual information.

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.

Results

We performed fMRI in awake, behaving infants and toddlers using a protocol described previously⁷. 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**), with an average duration of 540.7s (range: 186–1116s).

The retinotopic-mapping data from the same infants²⁴ 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.

As a proof of concept that the analyses we use with infants can identify fine-grained visual organization, we ran the main analyses on an adult sample. These adults (8 participants) had both retinotopic mapping data and movie-watching data. **Figures S1**, **S2**, **S3**, and **S4** demonstrate that applying these analyses to adult movie data reveals similar structure to what we find in infants.

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. Although these analyses cannot define visual maps, they test whether visual areas have different functional signatures. 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)^{32, 36, 37}. 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 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 (Mean [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’ streams¹⁸. We found no evidence that the variability in movie duration per participant correlated with this difference ($r=0.08$, $p=.700$). 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; $\Delta_{Fisher\ Z} M=0.20$, $p<0.001$). There was no correlation between movie duration and effect (Same > Adjacent: $r=-0.01$, $p=.965$, Adjacent > Distal: $r=-0.09$, $p=.740$). Additionally, if we control for motion in the correlation between areas — in case motion transients drive consistent activity across areas — then the effects described here are negligibly different (**Figure S5**). Hence, movies elicit distinct processing dynamics across areas of infant visual cortex defined independently using retinotopic mapping.

We previously found²⁴ that an anatomical segmentation of visual cortex³⁸ could identify these same areas reasonably well. Indeed, the results above were replicated when using visual areas defined anatomically (**Figure S6**). However, a key advantage of anatomical segmentation is that it can define visual areas not mapped by a functional retinotopy task. This could help

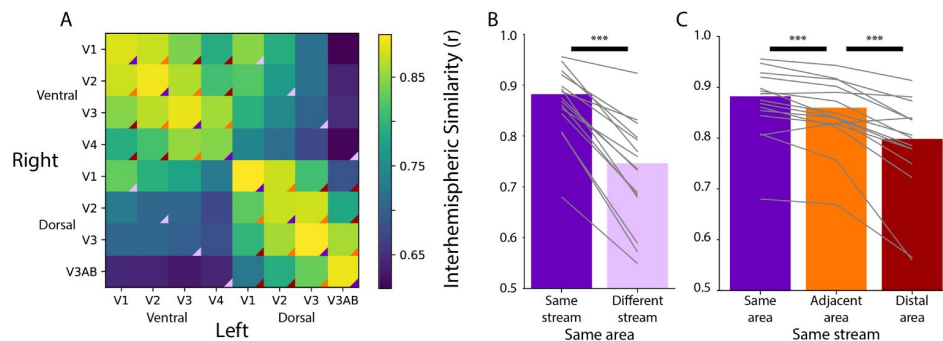


Figure 1

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

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^{18, 38}, 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 S6**)^{36, 39}. 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 pathway⁴⁰ (e.g., LO1, hMT), although more evidence is needed to substantiate this distinction. This separation between streams is striking when considering that it happens *despite* differences in visual field representations across areas: while dorsal V1 and ventral V1 represent the lower and upper visual field, respectively, V3A/B and hV4 both have full visual field maps. These visual field representations can be detected in adults⁴¹; however, they are often not the primary driver of function³⁹. We see that in infants too: hV4 and V3A/B represent the same visual space yet have distinct functional profiles. Again, this organization cannot be attributed to mere spatial autocorrelation within stream because analyses were conducted across hemispheres (at significant anatomical distance) and this pattern is preserved when accounting for motion (**Figure S5**). These results thus provide evidence of a dissociation in the functional profile 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 propose visual maps in individual infant brains^{26, 27, 36, 42, 43}. 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)^{26, 27, 36, 42, 43}. 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) retinotopic organization accounts for sufficient variance in visual activity to be identified by ICA and 2) experimenters can accurately identify these components.

Multiple maps could be identified per participant 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 surface⁴⁴. 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 (**Figure S7**). By contrast, lines drawn perpendicular to the area boundaries show oscillations in sensitivity to horizontal and vertical meridians on meridian maps (**Figure S8**). Using the same manually traced lines from the

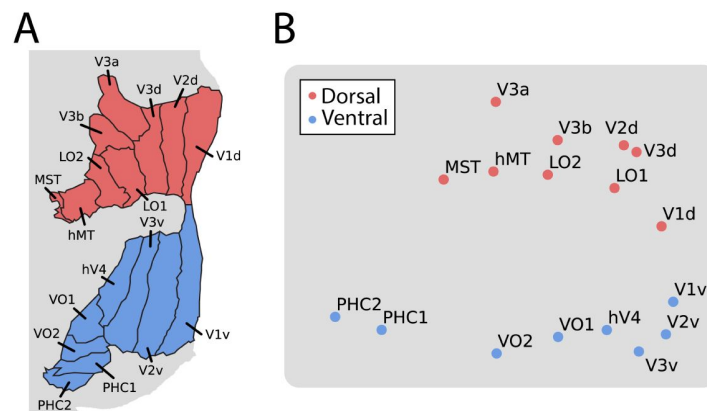


Figure 2

Multi-dimensional scaling (MDS) of movie-evoked activity in visual cortex. A) Anatomically defined areas³⁸ 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.

retinotopy task, we measured the intensity gradients in each component from the movie-watching data. We can then use the gradients of intensity in the retinotopy task-defined maps as a benchmark for comparison with the ICA-derived maps.

To assess the selected component maps, we correlated the gradients (described above) of the task-evoked and component maps. This test uses independent data: the components were defined based on movie data and validated against task-evoked retinotopic 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**). The percentile of the average manually identified component was high ($M=63.8$ percentile, range: 26.7–98.1) and significantly above chance ($\Delta M=13.8$, $CI=[3.3-23.9]$, $p=.010$). This illustrates that the manually identified components derived from movie-watching data are similar to the spatial frequency maps derived from retinotopic mapping. The fact that 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 ($\Delta M=17.6$, $CI=[-1.8-33.0]$, $p=.074$). 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. Even so, some participants have components that are highly similar to the meridian maps (e.g., s8037 1 2 or s6687 1 5 in **Figure S8**). 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. Equivalent analyses of adult data (**Figure S3**) support this conclusion: meridian maps are found in fewer adult participants.

Despite the similarity of the identified components to the retinotopic maps, it is possible that the components are noise and this similarity arose by chance. Indeed, given enough patterns of spatially smooth noise, some will resemble retinotopic organization. To test how often components derived from noise are misidentified, we made a version of each component in which the functional data were misaligned with respect to the anatomical data while preserving spatial smoothness. We then intermixed an equal number of these “rolled” components amongst the original components and randomized the order such that a coder would be blind as to whether any given component was rolled or original. The blind coder manually categorized each component as a spatial frequency component, a meridian component, or neither (identical to the steps above). It was not possible to make the coder blind for some participants whose rolled data contained visible clues because of partial voluming. In the 6 participants without such clues, only 1 of the 14 components labeled as spatial frequency or meridian, from 920 total components, was a rolled component. The fact that 13 of 14 selected components (93%) were original was extremely unlikely to have occurred by chance (binomial test: $p=0.002$). Thus, our selection procedure rarely identified components as retinotopic in realistic noise.

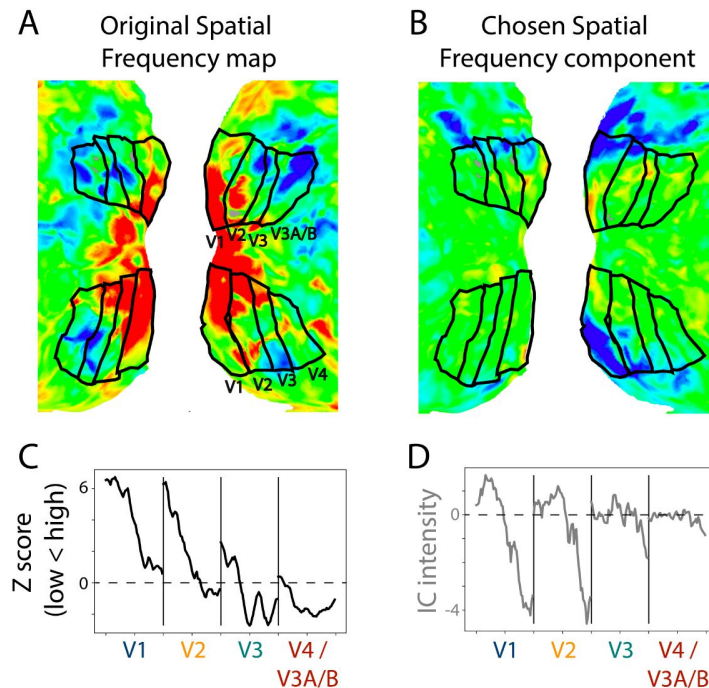


Figure 3

Example retinotopic task vs. ICA-based spatial frequency maps. A) Spatial frequency map of a 17.1 month old toddler. The retinotopic task data are from a prior study²⁴. 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. We used the same lines that were manually traced on the task-evoked map to assess the change in the component's response. We found a monotonic trend within area from medial to lateral, just like we see in the ground truth. This is one example result, find all participants in **Figure S7**.

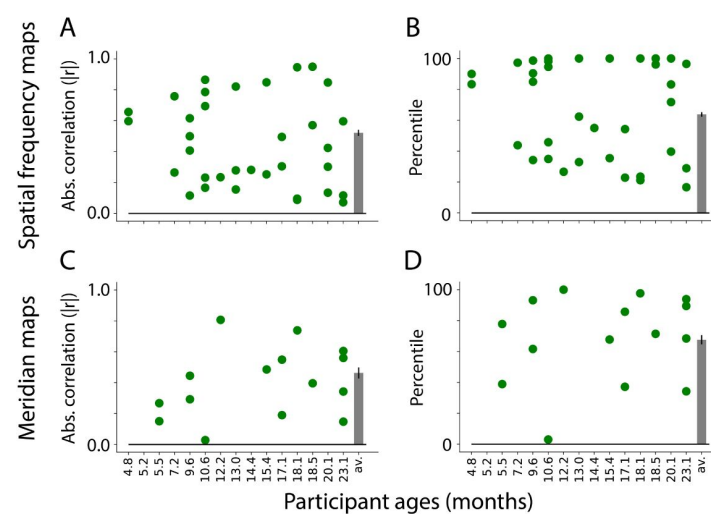


Figure 4

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.

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 alignment²⁸. 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 did not have usable retinotopy data (and thus were not included in the analyses above that compared movie and retinotopy data within participant). Second, we used data from adult participants, including 8 participants who completed the retinotopy task and saw a subset of the movies we showed infants, and 41 datasets from adults who had seen the movies shown to infants but did not have retinotopy data.

With this expanded dataset, we used shared response modeling (SRM)³⁴ 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 (**Figure 5A**). The number of features in the shared space ($K=10$) was determined via a cross-validation procedure on movie-watching data in adults (**Figure S9**). 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 (**Figure 5B**). We then mapped the held-out participant's movie data into the learned shared space without changing the shared space (**Figure 5C**). In other words, the shared response model was learned and frozen before the held-out participant's data was considered. This approach has been used and validated in prior SRM studies⁴⁵. 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 (**Figure 5D**).

This predicted visual organization was compared to the participant's actual visual map from the retinotopy task using the same methods as for ICA. In other words, the manually traced lines were used to measure the intensity gradients in the predicted maps, and these gradients were compared to the ground truth. 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-evoked retinotopy data (spatial frequency: $M=0.46$, range: $-0.06-0.78$; meridian: $M=0.24$, range: $-0.12-0.78$). Some participants were fit well using SRM (e.g., s2077 1 1, and s6687 1 5 for **Figures S10**, S11

To evaluate whether success was due to fitting the shared response, we flipped the held-out participant's movie data (i.e., the first timepoint became the last timepoint and vice versa) so that an appropriate fit is 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

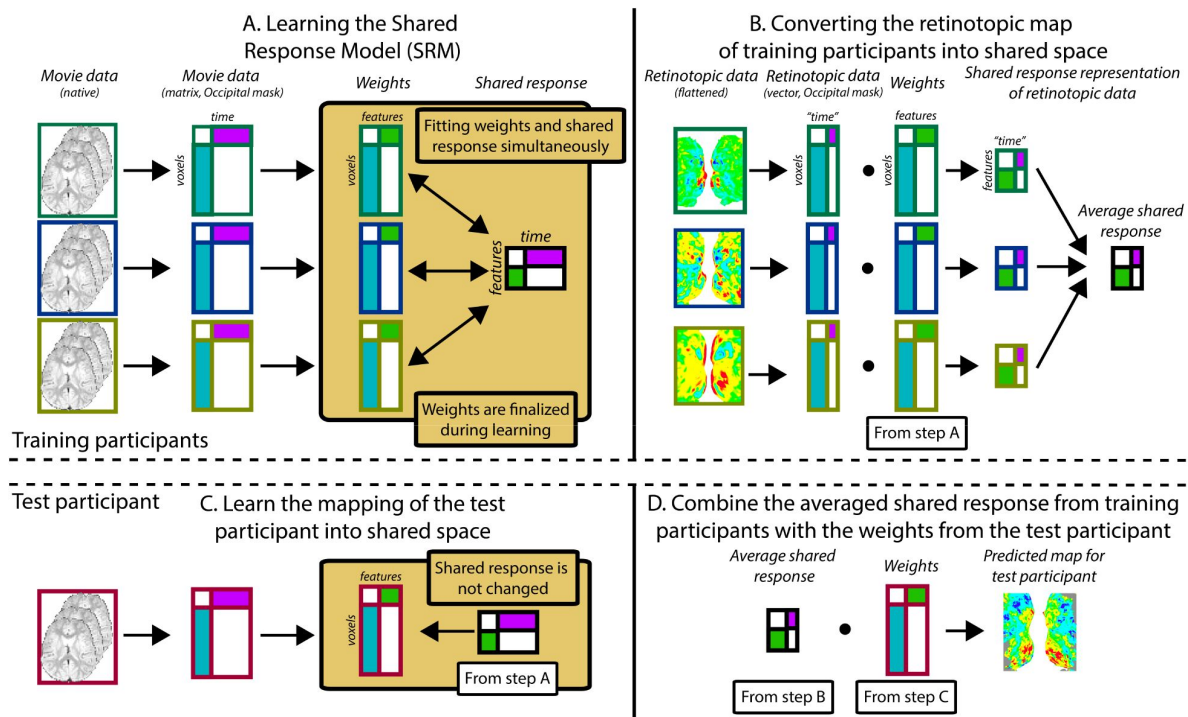


Figure 5

Pipeline for predicting visual maps from movie data. The figure divides the pipeline into 4 steps. All participants watched the same movie. To predict infant data from other infants (or adults), one participant was held out of the training and used as the test participant. Step A: The training participants' movie data (three color-coded participants shown in this schematic) is masked to include just Occipital voxels. The resulting matrix is run through shared response modeling (SRM)³⁴ to find a lower-dimensional embedding (i.e., a weight matrix) of their shared response. Step B: The training participants' retinotopic maps are transformed into the shared response space using the weight matrices determined in step A. Step C: Once steps A and B are finished, the test participant's movie data are mapped into the shared space that was fixed from step A. This creates a weight matrix for this test participant. Step D: The averaged shared response of the retinotopic maps from step B is combined with the test participant's weight matrix from step C to make a prediction of the retinotopic map in the test participant. This prediction can then be validated against their real map from the retinotopy task. Individual gradients for each participant are shown in **Figures S10**, **S11**, **S12**, **S13**.

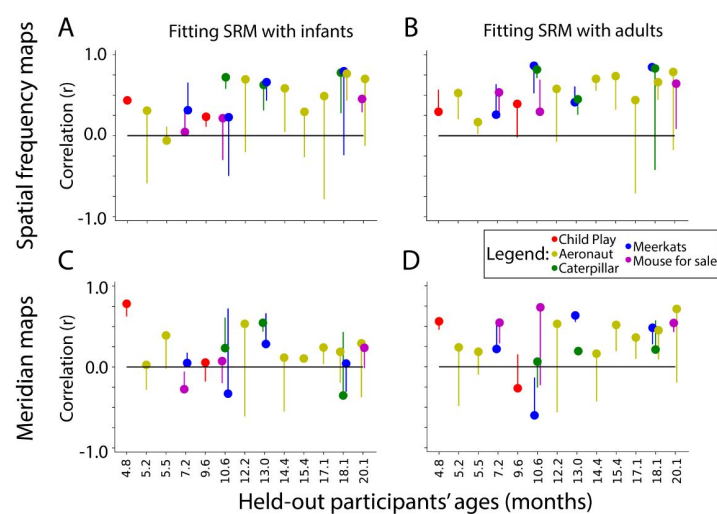


Figure 6

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.

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). Some participants were well predicted by these methods (e.g., s8037 1 2, and s6687 1 4 for **Figures S12** [↗](#), S13). Again, flipping the held-out participants movie 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.

We performed an anatomical alignment analog of the functional alignment (SRM) approach. This analysis serves as a benchmark for predicting visual maps using task-based data, rather than movie data, from other participants. For each infant participant, we aggregated all other infant or adult participants as a reference. The retinotopic maps from these reference participants were anatomically aligned to the standard surface template, and then averaged. These averages served as predictions of the maps in the test participant, akin to SRM, and were analyzed equivalently (i.e., correlating the gradients in the predicted map with the gradients in the task-based map). These correlations (**Table S4** [↗](#)) are significantly higher than for functional alignment (using infants to predict spatial frequency, anatomical alignment < functional alignment: $\Delta_{Fisher\ Z}$ M=0.44, CI=[0.32–0.58], $p<.001$; using infants to predict meridians, anatomical alignment < functional alignment: $\Delta_{Fisher\ Z}$ M=0.61, CI=[0.47–0.74], $p<.001$; using adults to predict spatial frequency, anatomical alignment < functional alignment: $\Delta_{Fisher\ Z}$ M=0.31, CI=[0.21–0.42], $p<.001$; using adults to predict meridians, anatomical alignment < functional alignment: $\Delta_{Fisher\ Z}$ M=0.49, CI=[0.39–0.60], $p<.001$). This suggests that even if SRM shows that movies *can* be used to produce retinotopic maps that are significantly similar to a participant, these maps are not as good as those that can be produced by anatomical alignment of the maps from other participants without any movie 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. This suggests that visual areas are functionally differentiated in infancy and that this function is shared across hemispheres³² [↗](#). By comparing across anatomically distant hemispheres, we reduced the impact of spatial autocorrelation and isolated the stimulus-driven signals in the brain activity³² [↗](#), ³⁶ [↗](#), ⁴⁶ [↗](#). 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 infants⁴⁷ [↗](#) but that the localization of these functions is immature⁴⁸ [↗](#). 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.

Not only do movies evoke differentiated activity in the infant visual cortex between areas, but movies also evoke fine-grained information about the organization of maps within areas. We used a data-driven approach (ICA) to discover maps that are similar to retinotopic maps in the infant visual cortex. We observed components that were highly similar to a spatial frequency map obtained from the same infant in a retinotopy task. This was also true for the meridian maps, to a lesser degree. This means that the retinotopic organization of the infant brain accounts for a detectable amount of variance in visual activity, otherwise components resembling these maps would not be discoverable. Importantly, the components could be identified without knowledge of these ground-truth maps; however, their moderate similarity to the task-defined maps makes them a poor replacement. 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^{23, 49–51} (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. Another limitation is that ICA does not provide a scale to the variation: although we find a correlation between gradients of spatial frequency in the ground truth and the selected component, we cannot use the component alone to infer the spatial frequency selectivity of any part of cortex. In other words, we cannot infer units of spatial frequency sensitivity from the components alone. Nonetheless, these results do show that it is possible to discover approximations of visual maps in infants and toddlers with movie-watching data and ICA.

We also asked whether functional alignment^{30, 34} could be used to detect visual maps in infants. Using a shared response model³⁴ 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 task²⁸. 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 data⁵². This is initial evidence that functional alignment may be useful for enhancing signal quality, like it has in adults^{28, 33, 34}, or revealing changing function over development⁴⁵, which may prove especially useful for infant fMRI⁵². 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.

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. Additionally, anatomical alignment (i.e., averaging the maps from other participants and anatomically aligning them to a held-out participant) resulted in maps that were highly similar to the ground truth. Indeed, we previously²⁴ found that adult-defined visual areas were moderately similar to infants. While functional alignment with adults can outperform anatomical alignment methods in similar analyses²⁸, here we find that functional alignment is inferior to anatomical alignment. Thus, if the goal is to define visual areas in an infant that lacks task-based retinotopy, anatomical alignment of other participants' retinotopic maps is superior to using movie-based analyses, at least as we tested it.

In conclusion, movies evoke activity in infants and toddlers 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 study²⁴. 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.

An adult sample was collected ($N=8$, 3 females) and used for validating the analyses and for supporting SRM analyses in infants. Each participant had both retinotopy and movie watching data. The adult participants saw the five most common movies that were seen by infants in our retinotopy 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 Kit⁵³; <https://brainiak.org/docs/>. Raw and preprocessed functional and anatomical data is available here: <https://doi.org/10.5061/dryad.jm63xsm3>.

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* gradientecho 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 paper⁷, 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 Psychtoolbox⁵⁴ for MATLAB. The details for the retinotopy task are explained fully elsewhere²⁴. 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 orientation⁵⁵. 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 frequency³⁶. 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^{29, 31}. 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 previously²⁴. The gaze coding for the movies was also the same as reported previously^{29, 31}. 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 data⁷. 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 SFNR⁵⁶ 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 mrTools⁵⁷.

The final step common across analyses created a transformation into surface space. Surfaces were reconstructed from iBEAT v2.0⁵⁸. These surfaces were then aligned into standard Buckner40 standard surface space⁵⁹ using FreeSurfer⁵⁹.

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 approximately⁶⁰ 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 ANTs⁶¹. For infants, an initial linear alignment with 12 DOF was used to align anatomical data to the age-specific infant template⁶², 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 atlas⁶³ in standard space – defined liberally to include any voxel with an above zero probability of being labelled as the occipital lobe – 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 paper²⁴ 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 AFNI⁶⁴, we demarcated the visual areas V1, V2, V3, V4, and V3A/B using traditional protocols based on the meridian map contrast⁶⁵. We traced lines perpendicular

and parallel to the area boundaries to quantify gradients in the visual areas. The anatomically-defined areas of interest³⁸ used in **Figure 2** were available in this standard surface space. The adult data were also traced using the same methods as infants (described previously²⁴) by one of the original infant coders (CE).

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 resampling⁶⁶. 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 adults³⁸ 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 separately³⁹), 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 MELODIC⁴². 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. For the adult participants who were analyzed, the components were selected *before* those participants were retinotopically traced, in order to minimize the potential contamination that could occur when performing these manual steps close in time.

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.

To evaluate whether components resembling retinotopic maps arise by chance, we misaligned the functional and anatomical data for a subset of participants and manually relabeled them. If retinotopic components are identified at the same rate in the misaligned data as the original data, this would support the concern that the selection process finds structure where there is none. For each participant, we aligned the components to standard surface space and then flipped the labels for left and right hemispheres. Loading these flipped files as if they were correctly aligned had the effect of rolling the functional signals with respect to the anatomy of the cortical surface. Specifically, because the image files are always read in the same order, but the hemispheres differ in the mosaic alignment of nodes in surface space, this flipping transposed voxels from early visual cortex laterally to the approximate position of the lateral occipital cortex and vice versa, while preserving smoothness. Of the 15 total participants, 9 were excluded from this analysis because they had partial volumes (e.g., missing the superior extent of the parietal lobe) such that their rolled data in surface space contained tell-tale signs (e.g., missing voxels were now in an unrealistic place) that precluded blind coding.

To set up the blind test for the coder, the rolled components from the 6 remaining participants were intermixed with an equal number of their original components and then the labels (as original or rolled) were hashed. A coder was given all of the original and rolled components with their hashed names and categorized each one as a spatial frequency component, meridian component, or neither. Once completed, these responses were cross-referenced against the unhashed names to determine whether the components the coder selected as retinotopic had been rolled. The proportion of selected components that were original (vs. rolled) was compared against chance (50%) with a binomial test.

Shared Response Modeling (SRM)

We based our SRM analyses on previous approaches using hyperalignment²⁸ and adapted them for our sample. 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 S9). 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 across movies. 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.

Anatomical alignment test

We performed a second type of between-participant analysis in addition to SRM. Specifically, we anatomically aligned the retinotopic maps from other participants to make a prediction of the map in a held-out participant. To achieve this, we first aligned all spatial frequency and meridian maps from infant and adult participants with retinotopy into the Buckner40 standard surface space⁵⁹. For each infant participant, we composed a map from the average of the *other* participants. The other participants were either all the other infants or all the adult participants. We then used the lines traced parallel to the area boundaries (for spatial frequency) or perpendicular to the area boundaries (for meridian) to extract gradients of response in the average maps. These gradients were then correlated with the ground-truth gradients (i.e., the alternations in sensitivity in the held-out infant using lines traced from that participant). These correlations were then compared to SRM results within participants using bootstrap resampling. If a participant had multiple movies worth of data, then they were averaged prior to this comparison.

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> )

Supplementary materials

Figure S1

Homotopic correlations between retinotopic areas in the adult sample, akin to [Figure 1](#). (A) Average correlation of the timecourse of activity evoked during movie watching for all areas. Correlation of homotopic areas: $M=0.83$ (range: 0.78–0.88). (B) Across-hemisphere similarity of the same visual area from the same stream and from different streams. Difference with bootstrap resampling: $\Delta_{Fisher} Z$ $M=0.24$, $p<0.001$. (C) Across-hemisphere similarity in the same stream when matching the same area, matching to an adjacent area, or matching to a distal area. Difference with bootstrap resampling: Same > Adjacent $\Delta_{Fisher} Z$ $M=0.10$, $p<0.001$; Adjacent > Distal $\Delta_{Fisher} Z$ $M=0.16$, $p<0.001$. Grey lines represent individual participants. *** = $p<0.001$ from bootstrap resampling.

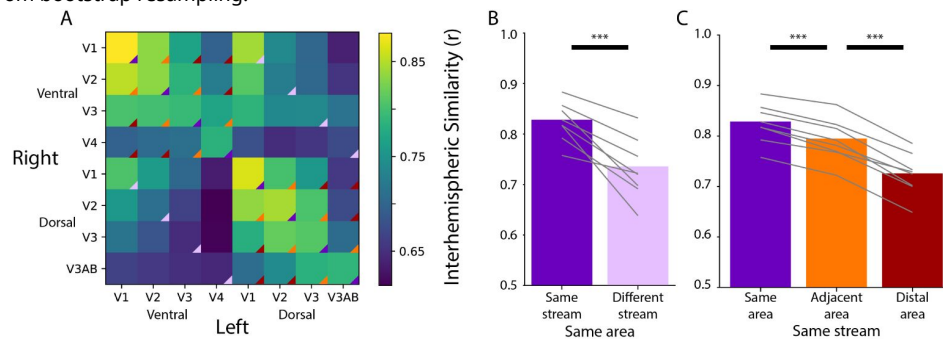


Figure S2

Multi-dimensional scaling of movie-evoked activity in adult visual cortex, akin to [Figure 2](#). A 2-dimensional embedding had inappropriately high stress – 0.87 – whereas a 3-dimensional embedding had appropriate stress: 0.105. This 3-dimensional scatter depicts the similarity of the functional timecourse of areas as a function of Euclidean distance. The plot shows a projection that emphasizes the similarity to the brain's organization.

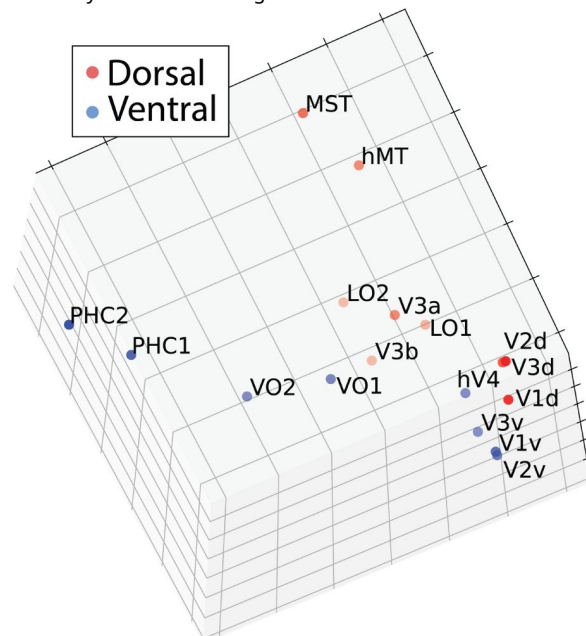


Figure S3

Similarity between visual maps from the adult retinotopy task and ICA applied to movies, akin to [Figure 4](#). 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 8 out of 8 adult 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. Percentile tests: $M=70.6$ percentile, range: 26.2–92.3, ΔM from chance=20.6, $CI=[4.2-34.9]$, $p=.014$. C) Same as A but for meridian maps. At least one component was identified in 6 out of 8 participants. D) Same as B but for meridian maps. Percentile tests: $M=74.6$ percentile, range: 40.3–98.0, ΔM from chance=24.6, $CI=[8.2-39.6]$, $p=.004$.

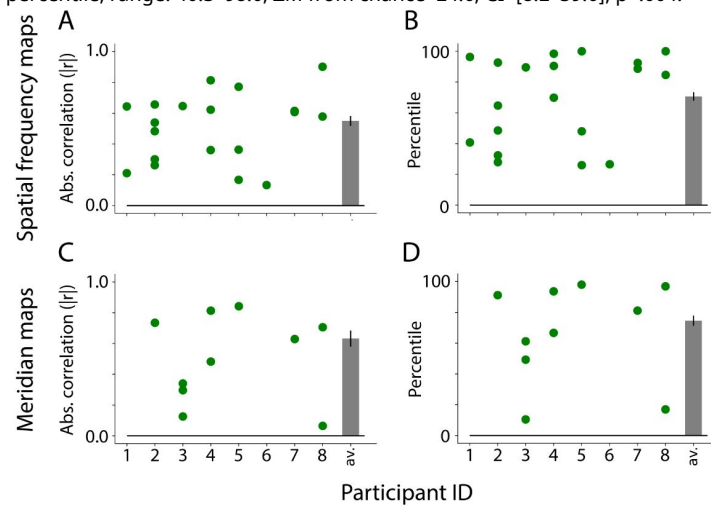
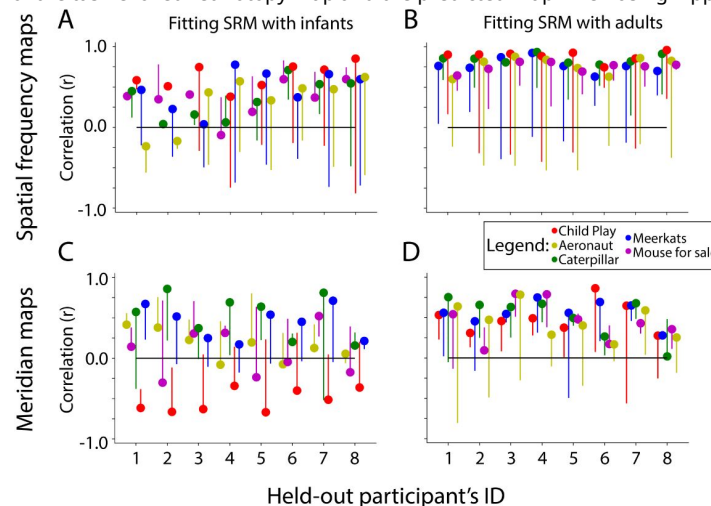


Figure S4

Similarity of SRM-predicted maps and task-evoked retinotopic maps in adults, akin to [Figure 6](#). Correlation between the gradients of the A) spatial frequency maps and C) meridian maps predicted with SRM from infants and their task-evoked retinotopy maps. Difference between real and flipped SRM fit: Spatial frequency= $\Delta_{Fisher} Z$ $M=0.59$, $CI=[0.36-0.83]$, $p<.001$. Meridian= $\Delta_{Fisher} Z$ $M=-0.07$, $CI=[-0.22-0.10]$, $p=.382$. Note: only two infants were used in the prediction with Child Play (red dots), hence why they likely show erratic behavior. B, D) Same as A, except using adult participants to train the SRM and predict maps. Difference between real and flipped SRM fit: Spatial frequency= $\Delta_{Fisher} Z$ $M=1.05$, $CI=[0.85-1.22]$, $p<.001$. Meridian= $\Delta_{Fisher} Z$ $M=0.49$, $CI=[0.36-0.64]$, $p<.001$. 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.



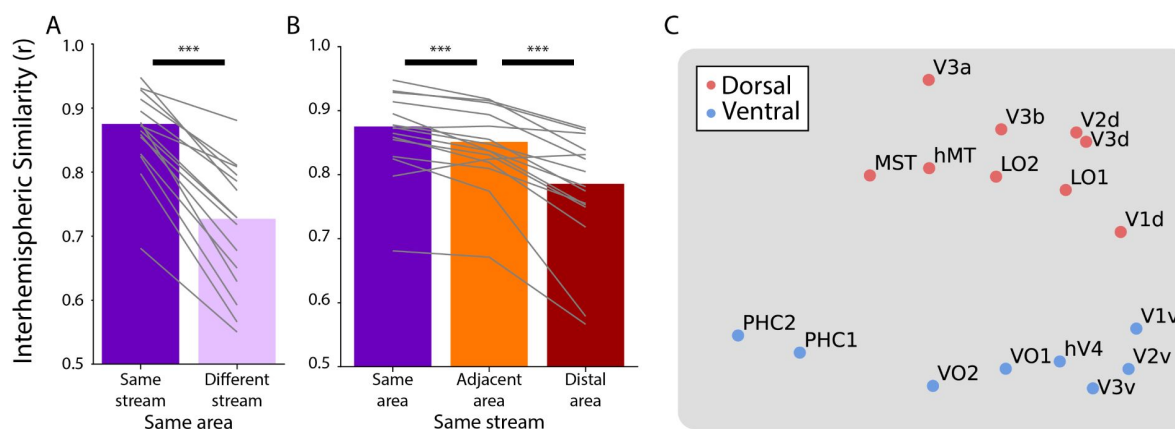


Figure S5

Homotopic correlations when controlling for motion. In this analysis, we computed correlations for all pairwise comparisons while partialing out our metric of motion: framewise displacement. In other words, if the functional timecourse in an area was correlated with the motion metric then this would decrease the correlation between that area and others. Subfigures A and B use task-evoked retinotopic definitions of areas (akin to [Figure 1](#)), whereas subfigure C uses anatomical definitions of areas (akin to [Figure 2](#)). Overall the results are qualitatively similar, suggesting that motion does not explain the effect observed here. A) 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: $\Delta_{Fisher Z} M=0.43$, $p<0.001$. B) 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 hV4). Difference with bootstrap resampling: Same > Adjacent $\Delta_{Fisher Z} M=0.09$, $p<0.001$; Adjacent > Distal $\Delta_{Fisher Z} M=0.20$, $p<0.001$. Grey lines represent individual participants. *** = $p<0.001$ from bootstrap resampling. C) Multidimensional scaling of the partial correlation between all anatomically defined areas. The timecourse of functional activity for each area was extracted and correlated across hemispheres, while partialing out framewise displacement. 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 (0.089). The plot shows a projection that emphasizes the similarity to the brain's organization.

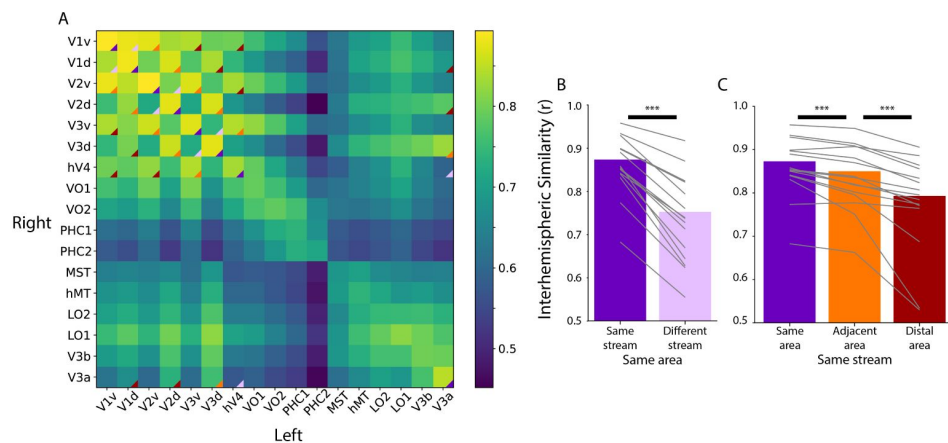


Figure S6

Homotopic correlations between anatomically defined areas corresponding to the data used in [Figure 2](#). A) Average correlation of the time course of activity evoked during movie watching for ventral and dorsal areas in an anatomical segmentation³⁸. 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 map (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 hV4). 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

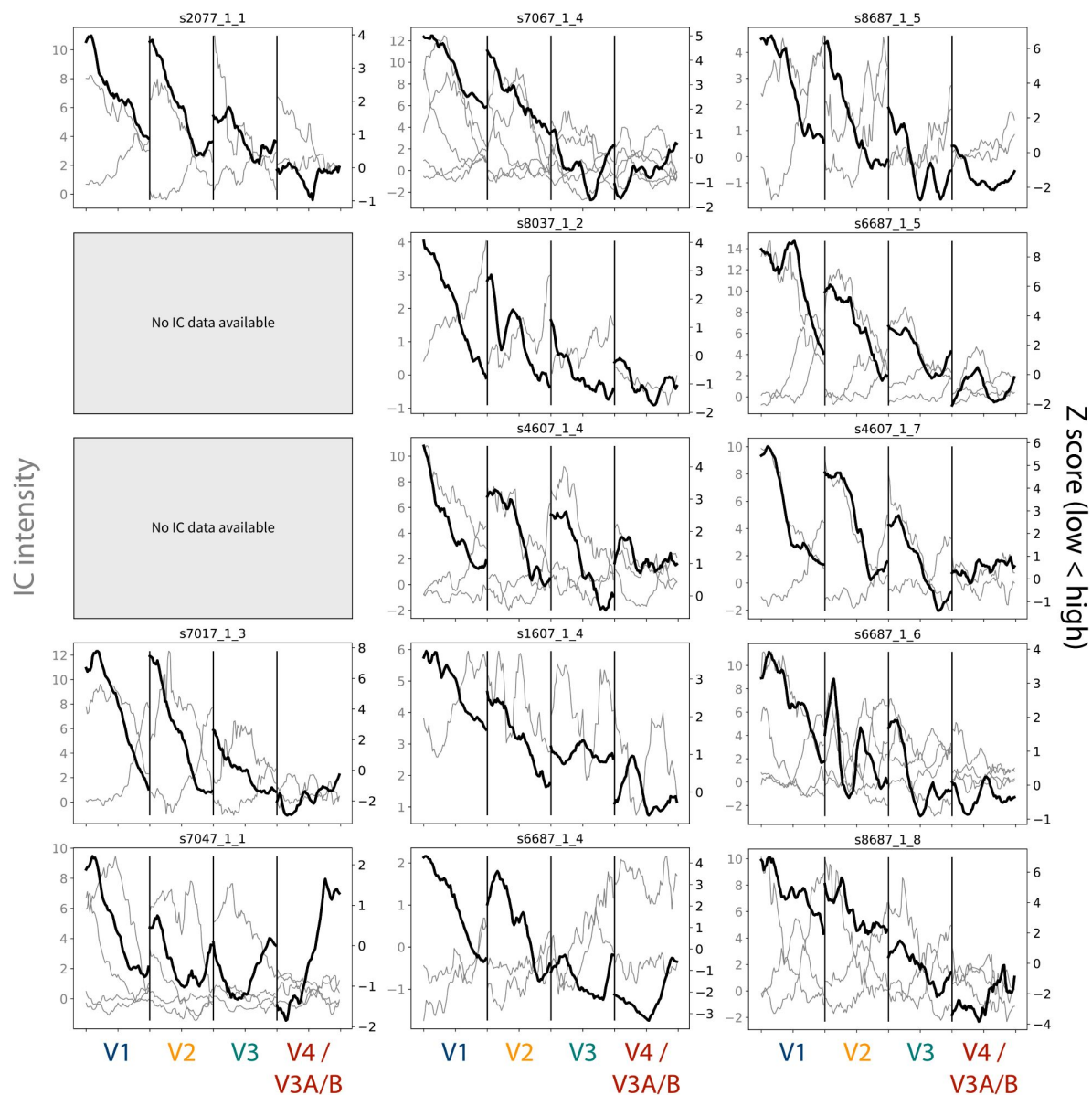


Figure S7

Gradients for the task-evoked and ICA-based spatial frequency maps. The grey lines depict the gradients from each chosen IC map, and their scale is indicated by the Y-axis on the left-hand side. The sign of the maps have not been edited, but it is arbitrary. The black line indicates the gradient from the task-evoked map, and their scale is indicated by the Y-axis on the right-hand side. Participants are listed in order of age. Participant data is not reported if no components were chosen for that participant.

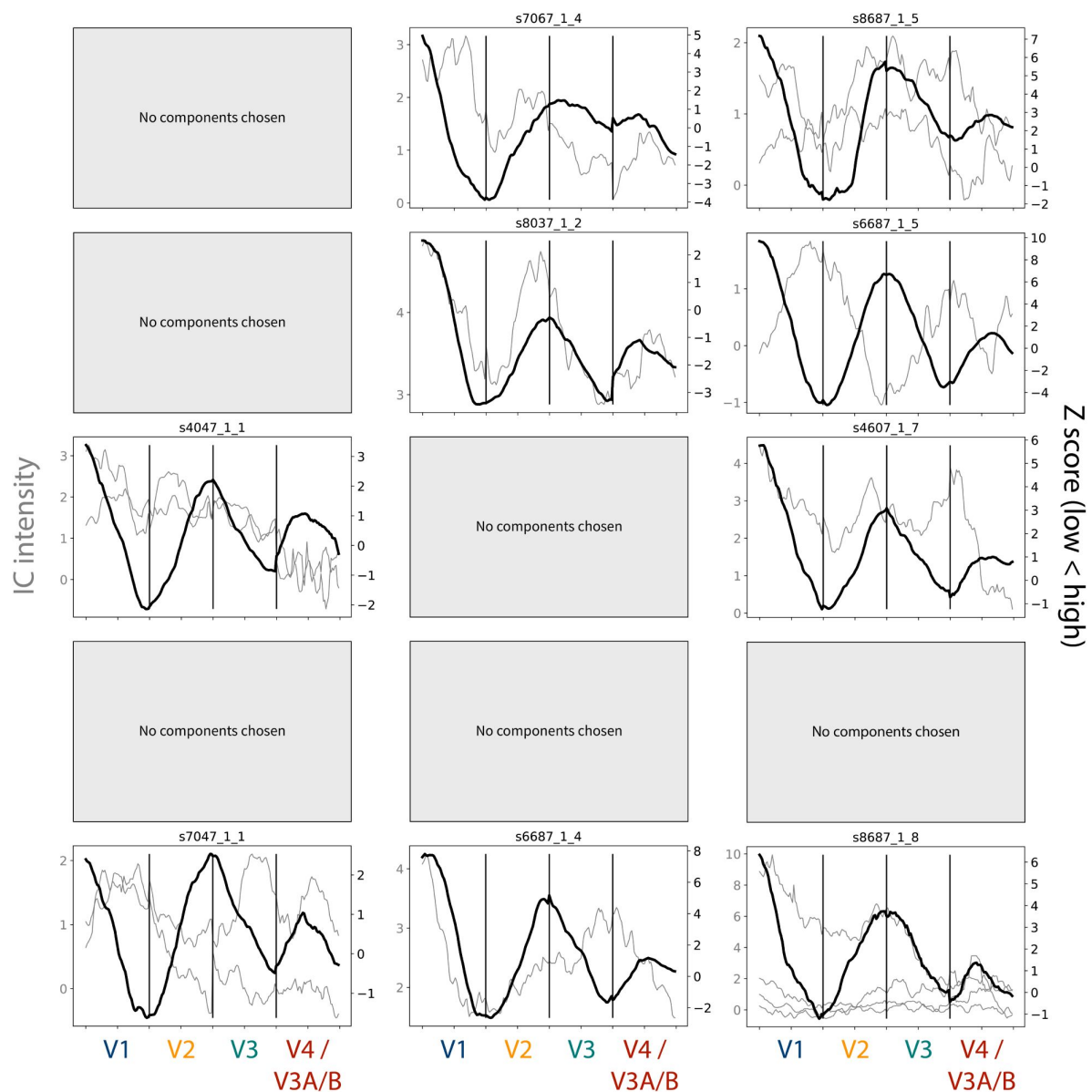


Figure S8

Gradients for the task-evoked and ICA-based meridian maps. The grey lines depict the gradients from each chosen IC map, and their scale is indicated by the Y-axis on the left-hand side. The sign of the maps have not been edited, but it is arbitrary. The black line indicates the gradient from the task-evoked map, and their scale is indicated by the Y-axis on the right-hand side. Participants are listed in order of age. Participant data is not reported if no components were chosen for that participant.

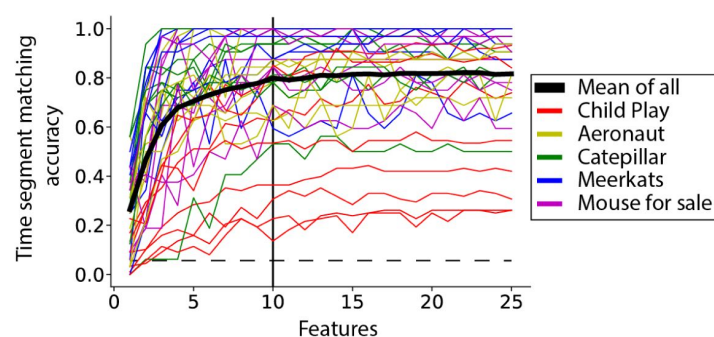


Figure S9

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^{30, 34}. 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.

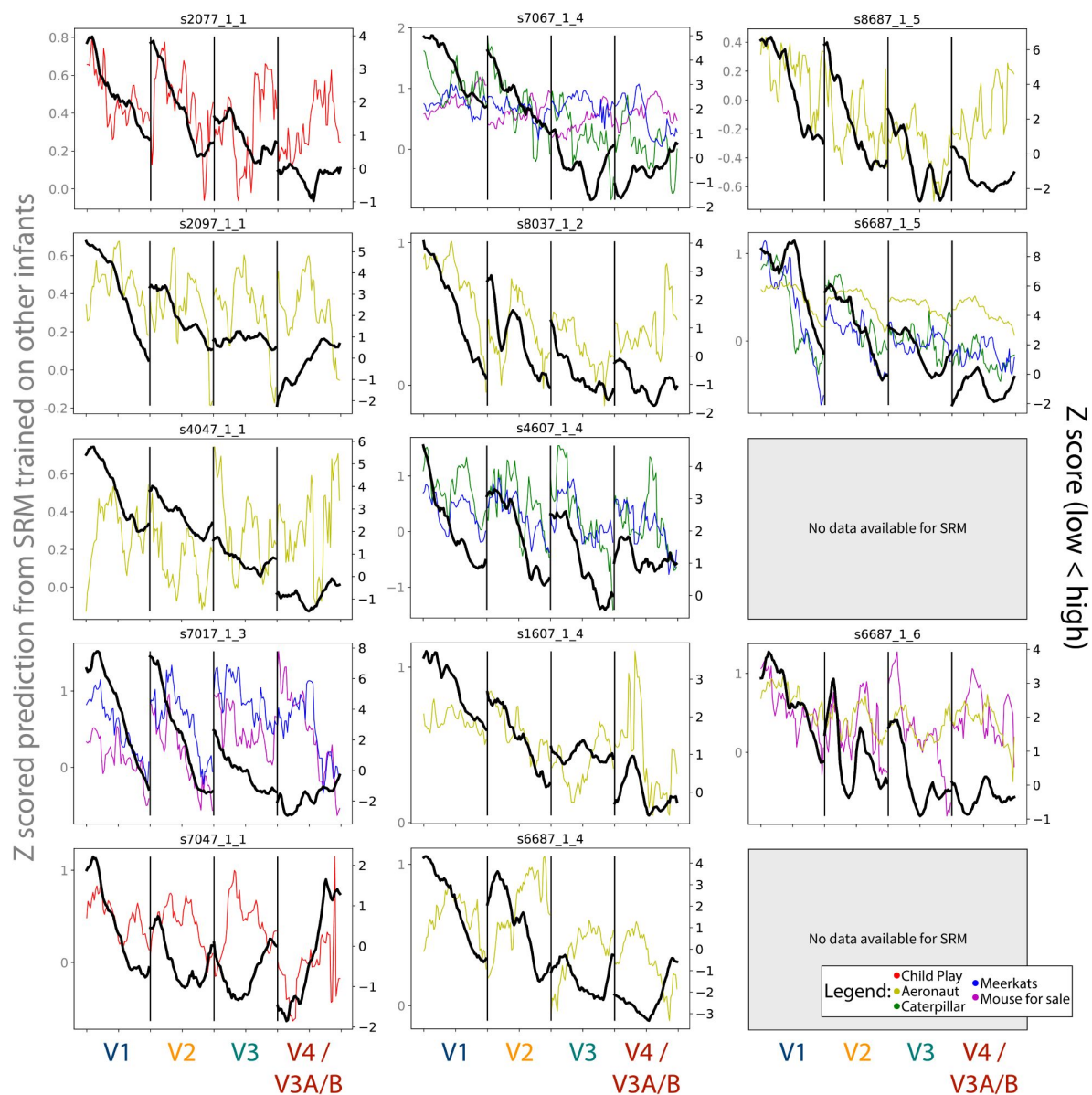


Figure S10

Gradients for the spatial frequency maps predicted using SRM from other infant participants, compared to the task-evoked gradients. The colored lines depict the gradients from each chosen movie that could be used, and their scale is indicated by the Y-axis on the left-hand side. The black line indicates the gradient from the task-evoked map, and their scale is indicated by the Y-axis on the right-hand side. Participants are listed in order of age. Participant data is not reported if the participant did not have SRM-compatible movie data.

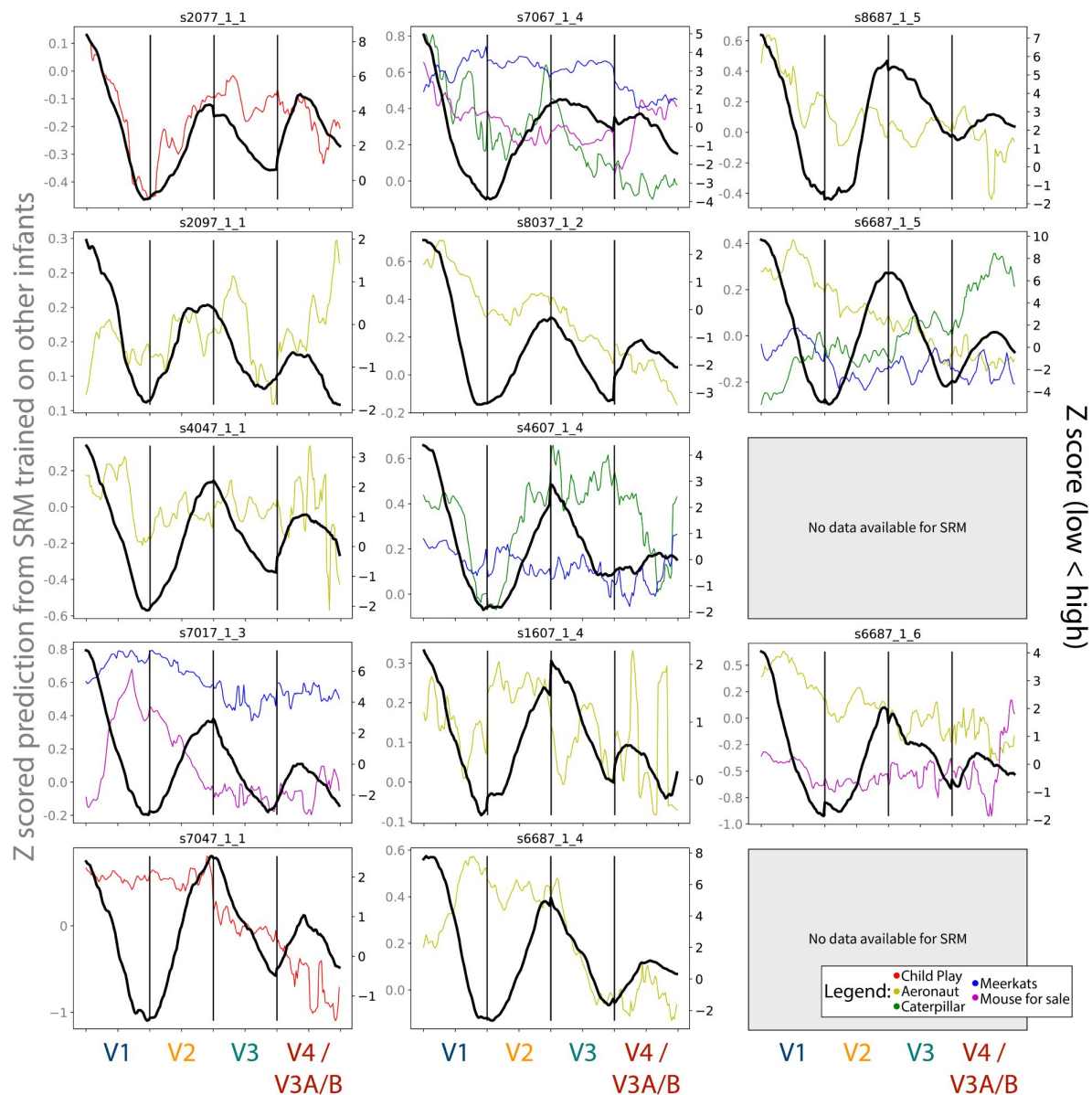


Figure S11

Gradients for the meridian maps predicted using SRM from other infant participants, compared to the task-evoked gradients. The colored lines depict the gradients from each chosen movie that could be used, and their scale is indicated by the Y-axis on the left-hand side. The black line indicates the gradient from the task-evoked map, and their scale is indicated by the Y-axis on the right-hand side. Participants are listed in order of age. Participant data is not reported if the participant did not have SRM-compatible movie data.

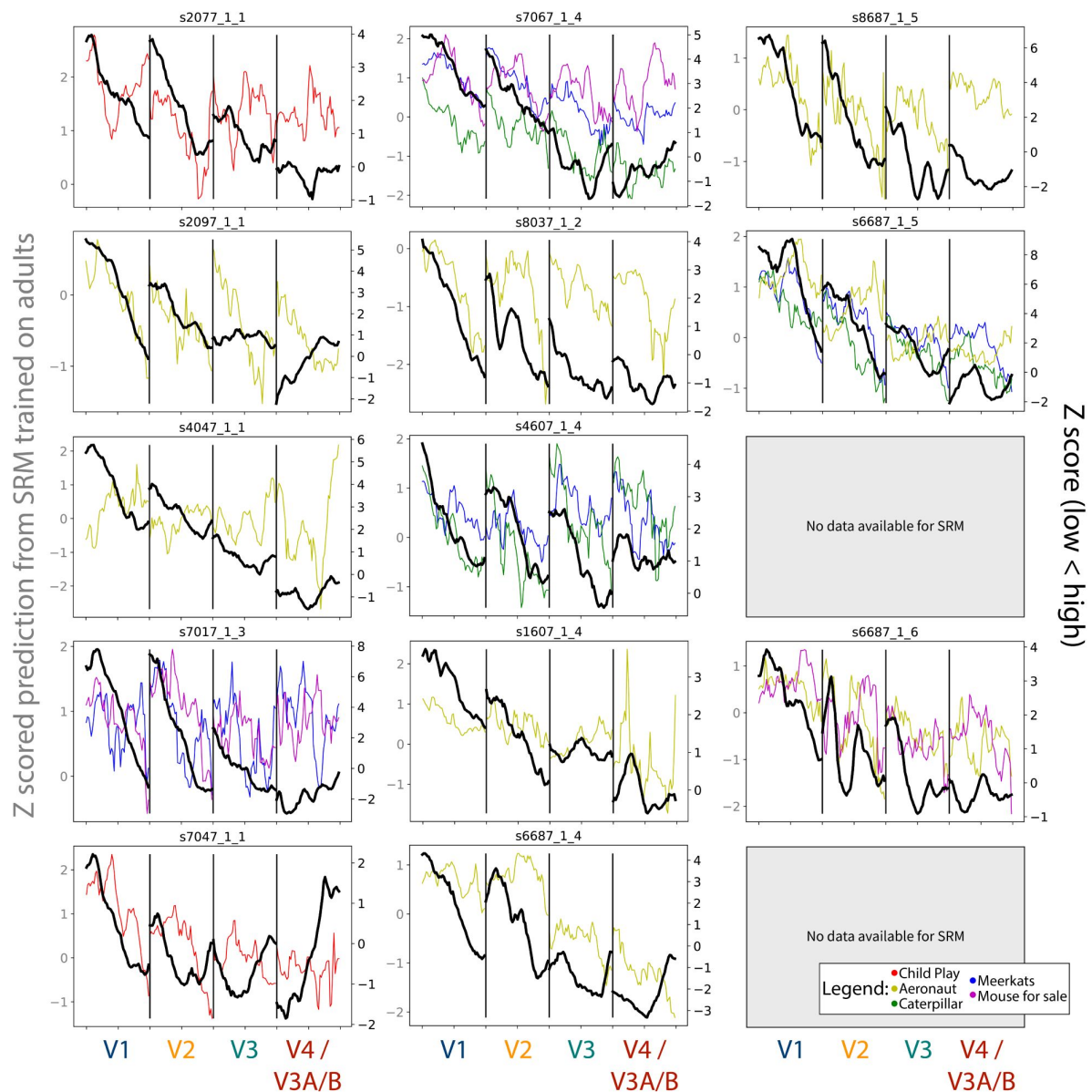


Figure S12

Gradients for the spatial frequency maps predicted using SRM from adult participants, compared to the task-evoked gradients. The colored lines depict the gradients from each chosen movie that could be used, and their scale is indicated by the Y-axis on the left-hand side. The black line indicates the gradient from the task-evoked map, and their scale is indicated by the Y-axis on the right-hand side. Participants are listed in order of age. Participant data is not reported if the participant did not have SRM-compatible movie data.

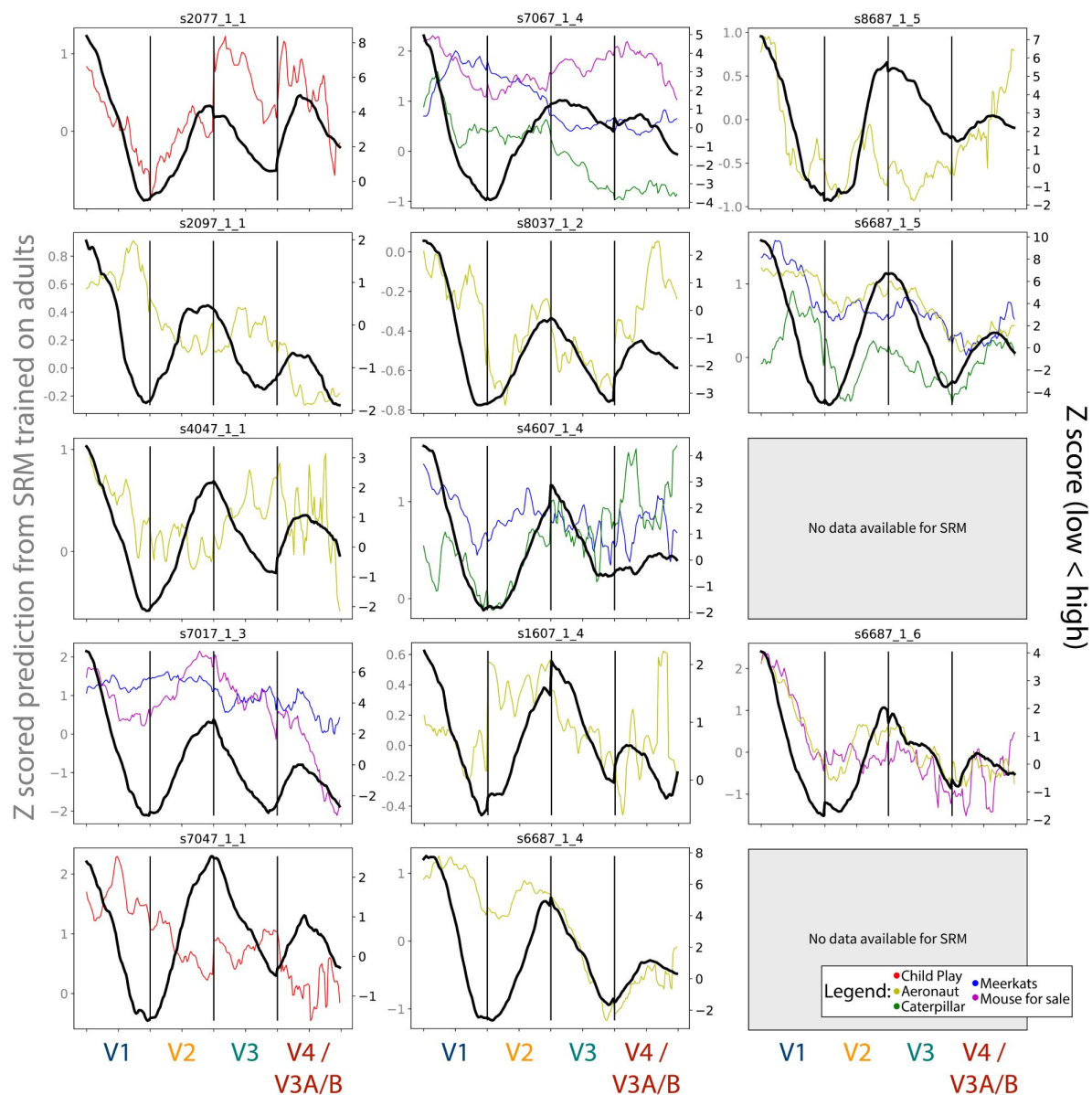


Figure S13

Gradients for the meridian maps predicted using SRM from adult participants, compared to the task-evoked gradients. The colored lines depict the gradients from each chosen movie that could be used, and their scale is indicated by the Y-axis on the left-hand side. The black line indicates the gradient from the task-evoked map, and their scale is indicated by the Y-axis on the right-hand side. Participants are listed in order of age. Participant data is not reported if the participant did not have SRM-compatible movie data.

ID	Age	Sex	Retinotopy Areas	Homotopy and ICA					SRM	
				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 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.

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 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.

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 ²⁹ .
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.

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.

ID	Adults		Infants	
	Spatial freq.	Meridians	Spatial freq.	Meridians
s2077_1_1	0.85	0.77	0.89	0.81
s2097_1_1	0.66	0.72	0.66	0.65
s4047_1_1	0.86	0.78	0.94	0.82
s7017_1_3	0.90	0.92	0.93	0.92
s7047_1_1	0.43	0.65	0.56	0.64
s7067_1_4	0.87	0.67	0.92	0.61
s8037_1_2	0.92	0.73	0.93	0.83
s4607_1_4	0.77	0.97	0.74	0.94
s1607_1_4	0.93	0.82	0.92	0.86
s6687_1_4	0.87	0.90	0.93	0.93
s8687_1_5	0.97	0.89	0.98	0.83
s6687_1_5	0.92	0.81	0.97	0.91
s4607_1_7	0.85	0.91	0.80	0.86
s6687_1_6	0.92	0.94	0.86	0.97
s8687_1_8	0.89	0.93	0.90	0.88
Mean	0.84	0.83	0.86	0.83

Table S4

Correlations between infant gradients and the spatial average of other infants or adults. For each participant, all other participants with retinotopy data (adults or infants) were aligned to standard surface space and averaged. The traced lines from the held-out participant were then applied to this average. The resulting gradients were correlated with the held-out participant and the correlation is reported here. This was done separately for meridian maps and spatial frequency maps

References

1. Biagi L., Crespi S. A., Tosetti M., Morrone M. C (2015) **Bold response selective to flow-motion in very young infants** *PLoS Biology* **13**
2. Biagi L., Tosetti M., Crespi S. A., Morrone M. C (2023) **Development of bold response to motion in human infants** *Journal of Neuroscience* **43**:3825–3837
3. Cabral L., Zubiaurre-Elorza L., Wild C. J., Linke A., Cusack R (2022) **Anatomical correlates of category-selective visual regions have distinctive signatures of connectivity in neonates** *Developmental Cognitive Neuroscience* **58**
4. Deen B., et al. (2017) **Organization of high-level visual cortex in human infants** *Nature Communications* **8**
5. Kosakowski H., et al. (2021) **Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants** *psyarxiv*
6. Truzzi A., Cusack R (2023) **The development of intrinsic timescales: A comparison between the neonate and adult brain** *NeuroImage* **275**
7. Ellis C. T., et al. (2020) **Re-imagining fmri for awake behaving infants** *Nature Communications* **11**
8. Vanderwal T., Eilbott J., Castellanos F. X (2019) **Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging** *Developmental Cognitive Neuroscience* **36**
9. Vanderwal T., Kelly C., Eilbott J., Mayes L. C., Castellanos F. X (2015) **Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging** *NeuroImage* **122**:222–232
10. Richardson H., Lisandrelli G., Riobueno-Naylor A., Saxe R (2018) **Development of the social brain from age three to twelve years** *Nature Communications* **9**
11. Alexander L. M., et al. (2017) **An open resource for transdiagnostic research in pediatric mental health and learning disorders** *Scientific Data* **4**:1–26
12. Nastase S. A., Goldstein A., Hasson U (2020) **Keep it real: rethinking the primacy of experimental control in cognitive neuroscience** *NeuroImage* **222**
13. Finn E. S., Glerean E., Hasson U., Vanderwal T (2022) **Naturalistic imaging: The use of ecologically valid conditions to study brain function** *NeuroImage* **247**
14. Franchak J. M., Heeger D. J., Hasson U., Adolph K. E (2016) **Free viewing gaze behavior in infants and adults** *Infancy* **21**:262–287
15. Tran M., Cabral L., Patel R., Cusack R (2017) **Online recruitment and testing of infants with mechanical turk** *Journal of Experimental Child Psychology* **156**:168–178

16. Brodmann K. (1909) **Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues** Barth
17. Felleman D. J., Van Essen D. C. (1991) **Distributed hierarchical processing in the primate cerebral cortex** *Cerebral Cortex (New York, NY: 1991)* **1**:1–47
18. Ungerleider L. G., Mishkin M., Ingle D. M. R., Goodale MA (1982) **Two cortical visual systems** *Analysis of visual behavior* Cambridge: MIT Press :549–586
19. Kaas J. H (1997) **Topographic maps are fundamental to sensory processing** *Brain research bulletin* **44**:107–112
20. White L. E., Fitzpatrick D. (2007) **Vision and cortical map development** *Neuron* :327–338
21. Fox P. T., Miezin F. M., Allman J. M., Van Essen D. C., Raichle M. E (1987) **Retinotopic organization of human visual cortex mapped with positron-emission tomography** *Journal of Neuroscience* **7**:913–922
22. Schneider W., Noll D. C., Cohen J. D (1993) **Functional topographic mapping of the cortical ribbon in human vision with conventional mri scanners** *Nature* **365**:150–153
23. Henriksson L., Nurminen L., Hyvärinen A., Vanni S. (2008) **Spatial frequency tuning in human retinotopic visual areas** *Journal of Vision* **8**:1–13
24. Ellis C. T., et al. (2021) **Retinotopic organization of visual cortex in human infants** *Neuron* **109**:2616–2626
25. Loiotile R. E., Cusack R., Bedny M (2019) **Naturalistic audio-movies and narrative synchronize “visual” cortices across congenitally blind but not sighted individuals** *Journal of Neuroscience* **39**:8940–8948
26. Knapen T (2021) **Topographic connectivity reveals task-dependent retinotopic processing throughout the human brain** *Proceedings of the National Academy of Sciences* **118**
27. Lu K.-H., Jeong J. Y., Wen H., Liu Z (2017) **Spontaneous activity in the visual cortex is organized by visual streams** *Human Brain Mapping* **38**:4613–4630
28. Guntupalli J. S., et al. (2016) **A model of representational spaces in human cortex** *Cerebral cortex* **26**:2919–2934
29. Yates T. S., et al. (2022) **Neural event segmentation of continuous experience in human infants** *Proceedings of the National Academy of Sciences* **119**
30. Turek J. S., Ellis C. T., Skalaban L. J., Turk-Browne N. B., Willke T. L (2018) **Capturing shared and individual information in fmri data** *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* IEEE :826–830
31. Yates T. S., Ellis C. T., Turk-Browne N. B (2023) **Functional networks in the infant brain during sleep and wake states** *bioRxiv* :2023–2
32. Li M., et al. (2022) **Development of visual cortex in human neonates is selectively modified by postnatal experience** *eLife* **11**

33. Busch E. L., et al. (2021) **Hybrid hyperalignment: A single high-dimensional model of shared information embedded in cortical patterns of response and functional connectivity** *NeuroImage* **233**
34. Chen P.-H., et al. (2015) **A reduced-dimension fmri shared response model** *NIPS* :460–468
35. Kumar S., Ellis C. T., O’Connell T. P., Chun M. M., Turk-Browne N. B (2020) **Searching through functional space reveals distributed visual, auditory, and semantic coding in the human brain** *PLOS Computational Biology* **16**
36. Arcaro M. J., Livingstone M. S (2017) **A hierarchical, retinotopic proto-organization of the primate visual system at birth** *eLife* **6**
37. Butt O. H., Benson N. C., Datta R., Aguirre G. K (2015) **Hierarchical and homotopic correlations of spontaneous neural activity within the visual cortex of the sighted and blind** *Frontiers in human neuroscience* **9**
38. Wang L., Mruczek R. E., Arcaro M. J., Kastner S (2014) **Probabilistic maps of visual topography in human cortex** *Cerebral Cortex* **25**:3911–3931
39. Haak K. V., Beckmann C. F (2018) **Objective analysis of the topological organization of the human cortical visual connectome suggests three visual pathways** *Cortex* **98**:73–83
40. Weiner K. S., Gomez J (2021) **Third visual pathway, anatomy, and cognition across species** *Trends in Cognitive Sciences* **25**:548–549
41. Haak K. V., et al. (2013) **Connective field modeling** *Neuroimage* :376–384
42. Beckmann C. F., DeLuca M., Devlin J. T., Smith S. M (2005) **Investigations into resting-state connectivity using independent component analysis** *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**:1001–1013
43. Moeller S., Nallasamy N., Tsao D. Y., Freiwald W. A (2009) **Functional connectivity of the macaque brain across stimulus and arousal states** *Journal of Neuroscience* **29**:5897–5909
44. Arcaro M. J., McMains S. A., Singer B. D., Kastner S (2009) **Retinotopic organization of human ventral visual cortex** *Journal of Neuroscience* **29**:10638–10652
45. Yates T. S., Ellis C. T., Turk-Browne N. B (2021) **Emergence and organization of adult brain function throughout child development** *NeuroImage* **226**
46. Smyser C. D., et al. (2010) **Longitudinal analysis of neural network development in preterm infants** *Cerebral Cortex* **20**:2852–2862
47. Wattam-Bell J., et al. (2010) **Reorganization of global form and motion processing during human visual development** *Current Biology* **20**:411–415
48. Braddick O., Atkinson J (2011) **Development of human visual function** *Vision Research* **51**:1588–1609
49. Smith A. T., Singh K. D., Williams A., Greenlee M. W (2001) **Estimating receptive field size from fmri data in human striate and extrastriate visual cortex** *Cerebral Cortex* **11**:1182–1190

50. Srihasam K., Vincent J. L., Livingstone M. S (2014) **Novel domain formation reveals proto-architecture in inferotemporal cortex** *Nature Neuroscience* **17**:1776–1783
51. Tolhurst D., Thompson I (1981) **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
52. Ellis C. T., Turk-Browne N. B (2018) **Infant fmri: A model system for cognitive neuroscience** *Trends in Cognitive Sciences* **22**:375–387
53. Kumar M., et al. (2020) **BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis** *PLOS Computational Biology* **16**
54. Kleiner M., et al. (2007) **What's new in psychtoolbox-3** *Perception* **36**
55. Tootell R., et al. (1995) **Functional analysis of human mt and related visual cortical areas using magnetic resonance imaging** *Journal of Neuroscience* **15**:3215–3230
56. Friedman L., Glover G. H (2006) **Report on a multicenter fmri quality assurance protocol** *Journal of Magnetic Resonance Imaging* **23**:827–839
57. Gardner J., Merriam E., Schluppeck D., Besle J., Heeger D. (2018) **mrtools: Analysis and visualization package for functional magnetic resonance imaging data** *Zenodo* **10**
58. Wang L., et al. (2023) **ibeat v2. 0: a multisite-applicable, deep learning-based pipeline for infant cerebral cortical surface reconstruction** *Nature Protocols* :1–32
59. Dale A. M., Fischl B., Sereno M. I (1999) **Cortical surface-based analysis: I. segmentation and surface reconstruction** *NeuroImage* **9**:179–194
60. Poppe T., Moore J. W., Arichi T (2021) **Individual focused studies of functional brain development in early human infancy** *Current Opinion in Behavioral Sciences* **40**:137–143
61. Avants B. B., et al. (2011) **A reproducible evaluation of ants similarity metric performance in brain image registration** *NeuroImage* **54**:2033–2044
62. Fonov V., et al. (2011) **Unbiased average age-appropriate atlases for pediatric studies** *NeuroImage* **54**:313–327
63. Mazziotta J., et al. (2001) **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
64. Cox R. W (1996) **Afni: software for analysis and visualization of functional magnetic resonance neuroimages** *Computers and Biomedical research* **29**:162–173
65. Wandell B. A., Dumoulin S. O., Brewer A. A (2007) **Visual field maps in human cortex** *Neuron* **56**:366–383
66. Efron B., Tibshirani R (1986) **Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy** *Statistical Science* :54–75

Author information

Cameron T Ellis

Department of Psychology, Stanford University, Palo Alto, United States

For correspondence: cte@stanford.edu

Tristan S Yates

Department of Psychology, Yale University, New Haven, United States

Michael J Arcaro

Department of Psychology, University of Pennsylvania, Philadelphia, United States

ORCID iD: [0000-0002-4612-9921](https://orcid.org/0000-0002-4612-9921)

Nicholas B Turk-Browne

Department of Psychology, Yale University, New Haven, United States, Wu Tsai Institute, Yale University, New Haven, United States

Editors

Reviewing Editor

Jessica Dubois

Inserm Unité NeuroDiderot, Université Paris Cité, Paris, France

Senior Editor

Timothy Behrens

University of Oxford, Oxford, United Kingdom

Reviewer #2 (Public review):

Summary:

This manuscript reports analyses of fMRI data from infants and toddlers watching naturalistic movies. Visual areas in the infant brain show distinct functions, consistent with previous studies using resting state and awake task-based infant fMRI. The pattern of activity in visual regions contains some features predicted by the regions' retinotopic responses. The revised version of the manuscript provides additional validation of the methodology, and clarifies the claims. As a result, the data provide clear support for the claims.

Strengths:

The authors have collected a unique dataset: the same individual infants both watched naturalistic animations and a specific retinotopy task. Using these data position the authors show that activity evoked by movies, in infants' visual areas, is correlated with the regions' retinotopic response. The revised manuscript validates this methodology, using adult data. The revised manuscript also shows that an infant's movie watching data is not sufficient or optimal to predict their visual areas' retinotopic responses; anatomical alignment with a group of previous participants provides more accurate prediction of a new participant's retinotopic response.

Weaknesses:

A key step in the analysis of the movie-watching data is the selection of independent components of the movie evoked response that resemble retinotopic spatial patterns. While the trained researcher was unlikely to be biased by this infant's own retinotopy, he/she was actively looking for ICs that resemble average patterns of retinotopic response. To show that these ICs didn't arise by chance (i.e. in noise), the authors proposed an additional analysis in the revised manuscript, by misaligning the functional and anatomical data for a subset of participants. This only partially confirms the reliability of the original components, since when the (new) coder tried to be conservative to avoid false components, he/she identified just over half of the 'true' components (13 vs 22 estimated over the group of 6 infants).

<https://doi.org/10.7554/eLife.92119.3.sa1>

Author response:

The following is the authors' response to the previous reviews.

eLife Assessment

This study presents valuable findings on the potential of short-movie viewing fMRI protocol to explore the functional and topographical organization of the visual system in awake infants and toddlers. Although the data are compelling given the difficulty of studying this population, the evidence presented is incomplete and would be strengthened by additional analyses to support the authors' claims. This study will be of interest to cognitive neuroscientists and developmental psychologists, especially those interested in using fMRI to investigate brain organisation in pediatric and clinical populations with limited fMRI tolerance.

We are grateful for the thorough and thoughtful reviews. We have provided point-by-point responses to the reviewers' comments, but first, we summarize the major revisions here. We believe these revisions have substantially improved the clarity of the writing and impact of the results.

Regarding the framing of the paper, we have made the following major changes in response to the reviews:

- (1) We have clarified that our goal in this paper was to show that movie data contains topographic, fine-grained details of the infant visual cortex. In the revision, we now state clearly that our results should not be taken as evidence that movies could replace retinotopy and have reworded parts of the manuscript that could mislead the reader in this regard.
- (2) We have added extensive details to the (admittedly) complex methods to make them more approachable. An example of this change is that we have reorganized the figure explaining the Shared Response Modelling methods to divide the analytic steps more clearly.
- (3) We have clarified the intermediate products contributing to the results by adding 6 supplementary figures that show the gradients for each IC or SRM movie and each infant participant.

In response to the reviews, we have conducted several major analyses to support our findings further:

- (1) To verify that our analyses can identify fine-grained organization, we have manually traced and labeled adult data, and then performed the same analyses on them. The results from this additional dataset validate that these analyses can recover fine-grained organization of the visual cortex from movie data.

(2) To further explore how visual maps derived from movies compare to alternative methods, we performed an anatomical alignment control analysis. We show that high-quality maps can be predicted from other participants using anatomical alignment.

(3) To test the contribution of motion to the homotopy analyses, we regressed out the motion effects in these analyses. We found qualitatively similar results to our main analyses, suggesting motion did not play a substantial role.

(4) To test the contribution of data quantity to the homotopy analyses, we correlated the amount of movie data collected from each participant with the homotopy results. We did not find a relationship between data quantity and the homotopy results.

Public Reviews:

Reviewer #1 (Public Review):

Summary:

Ellis et al. investigated the functional and topographical organization of the visual cortex in infants and toddlers, as evidenced by movie-viewing data. They build directly on prior research that revealed topographic maps in infants who completed a retinotopy task, claiming that even a limited amount of rich, naturalistic movie-viewing data is sufficient to reveal this organization, within and across participants. Generating this evidence required methodological innovations to acquire high-quality fMRI data from awake infants (which have been described by this group, and elsewhere) and analytical creativity. The authors provide evidence for structured functional responses in infant visual cortex at multiple levels of analyses; homotopic brain regions (defined based on a retinotopy task) responded more similarly to one another than to other brain regions in visual cortex during movie-viewing; ICA applied to movie-viewing data revealed components that were identifiable as spatial frequency, and to a lesser degree, meridian maps, and shared response modeling analyses suggested that visual cortex responses were similar across infants/toddlers, as well as across infants/toddlers and adults. These results are suggestive of fairly mature functional response profiles in the visual cortex in infants/toddlers and highlight the potential of movie-viewing data for studying finer-grained aspects of functional brain responses, but further evidence is necessary to support their claims and the study motivation needs refining, in light of prior research.

Strengths:

- This study links the authors' prior evidence for retinotopic organization of visual cortex in human infants (Ellis et al., 2021) and research by others using movie-viewing fMRI experiments with adults to reveal retinotopic organization (Knapen, 2021).*
- Awake infant fMRI data are rare, time-consuming, and expensive to collect; they are therefore of high value to the community. The raw and preprocessed fMRI and anatomical data analyzed will be made publicly available.*

We are grateful to the reviewer for their clear and thoughtful description of the strengths of the paper, as well as their helpful outlining of areas we could improve.

Weaknesses:

- The Methods are at times difficult to understand and in some cases seem inappropriate for the conclusions drawn. For example, I believe that the movie-defined ICA components were validated using independent data from the retinotopy task, but this was a point of confusion among reviewers.*

We acknowledge the complexity of the methods and wish to clarify them as best as possible for the reviewers and the readers. We have extensively revised the methods and results sections to help avoid potential misunderstandings. For instance, we have revamped the figure and caption describing the SRM pipeline (Figure 5).

To answer the stated confusion directly, the ICA components were derived from the movie data and validated on the (completely independent) retinotopy data. There were no additional tasks. The following text in the paper explains this point:

“To assess the selected component maps, we correlated the gradients (described above) of the task-evoked and component maps. This test uses independent data: the components were defined based on movie data and validated against task-evoked retinotopic maps.” Pg. 11

In either case: more analyses should be done to support the conclusion that the components identified from the movie reproduce retinotopic maps (for example, by comparing the performance of movie-viewing maps to available alternatives (anatomical ROIs, group-defined ROIs).

Before addressing this suggestion, we want to restate our conclusions: features of the retinotopic organization of infant visual cortex could be predicted from movie data. We did not conclude that movie data could ‘reproduce’ retinotopic maps in the sense that they would be a replacement. We recognize that this was not clear in our original manuscript and have clarified this point throughout, including in this section of the discussion:

“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. Additionally, anatomical alignment (i.e., averaging the maps from other participants and anatomically aligning them to a held-out participant) resulted in maps that were highly similar to the ground truth. Indeed, we previously[23] found that adult-defined visual areas were moderately similar to infants. While functional alignment with adults can outperform anatomical alignment methods in similar analyses[27], here we find that functional alignment is inferior to anatomical alignment. Thus, if the goal is to define visual areas in an infant that lacks task-based retinotopy, anatomical alignment of other participants’ retinotopic maps is superior to using movie-based analyses, at least as we tested it.” Pg. 21

As per the reviewer’s suggestion and alluded to in the paragraph above, we have created anatomically aligned visual maps, providing an analogous test to the betweenparticipant analyses like SRM. We find that these maps are highly similar to the ground truth. We describe this result in a new section of the results:

“We performed an anatomical alignment analog of the functional alignment (SRM) approach. This analysis serves as a benchmark for predicting visual maps using taskbased data, rather than movie data, from other participants. For each infant participant, we aggregated all other infant or adult participants as a reference. The retinotopic maps from these reference participants were anatomically aligned to the standard surface template, and then averaged. These averages served as predictions of the maps in the test participant, akin to SRM, and were analyzed equivalently (i.e., correlating the gradients in the predicted map with the gradients in the task-based map). These correlations (Table S4) are significantly higher than for functional alignment (using infants to predict spatial frequency, anatomical alignment > functional alignment: $\Delta_{\text{Fisher Z}}$ M=0.44, CI=[0.32–0.58], $p<.001$; using infants to predict meridians, anatomical alignment > functional alignment: $\Delta_{\text{Fisher Z}}$ M=0.61, CI=[0.47–0.74], $p<.001$; using adults to predict spatial frequency, anatomical alignment > functional

alignment: $\Delta_{\text{Fisher } Z}$ $M=0.31$, $CI=[0.21-0.42]$, $p<.001$; using adults to predict meridians, anatomical alignment > functional alignment: $\Delta_{\text{Fisher } Z}$ $M=0.49$, $CI=[0.39-0.60]$, $p<.001$). This suggests that even if SRM shows that movies can be used to produce retinotopic maps that are significantly similar to a participant, these maps are not as good as those that can be produced by anatomical alignment of the maps from other participants without any movie data.” Pg. 16–17

Also, the ROIs used for the homotopy analyses were defined based on the retinotopic task rather than based on movie-viewing data alone - leaving it unclear whether movie-viewing data alone can be used to recover functionally distinct regions within the visual cortex.

We agree with the reviewer that our approach does not test whether movie-viewing data alone can be used to recover functionally distinct regions. The goal of the homotopy analyses was to identify whether there was functional differentiation of visual areas in the infant brain while they watch movies. This was a novel question that provides positive evidence that these regions are functionally distinct. In subsequent analyses, we show that when these areas are defined anatomically, rather than functionally, they also show differentiated function (e.g., Figure 2). Nonetheless, our intention was not to use the homotopy analyses to define the regions. We have added text to clarify the goal and novelty of this analysis.

“Although these analyses cannot define visual maps, they test whether visual areas have different functional signatures.” Pg. 6

Additionally, even if the goal were to define areas based on homotopy, we believe the power of that analysis would be questionable. We would need to use a large amount of the movie data to define the areas, leaving a low-powered dataset to test whether their function is differentiated by these movie-based areas.

- The authors previously reported on retinotopic organization of the visual cortex in human infants (Ellis et al., 2021) and suggest that the feasibility of using movie-viewing experiments to recover these topographic maps is still in question. They point out that movies may not fully sample the stimulus parameters necessary for revealing topographic maps/areas in the visual cortex, or the time-resolution constraints of fMRI might limit the use of movie stimuli, or the rich, uncontrolled nature of movies might make them inferior to stimuli that are designed for retinotopic mapping, or might lead to variable attention between participants that makes measuring the structure of visual responses across individuals challenging. This motivation doesn't sufficiently highlight the importance or value of testing this question in infants. Further, it's unclear if/how this motivation takes into account prior research using movie-viewing fMRI experiments to reveal retinotopic organization in adults (e.g., Knapen, 2021). Given the evidence for retinotopic organization in infants and evidence for the use of movie-viewing experiments in adults, an alternative framing of the novel contribution of this study is that it tests whether retinotopic organization is measurable using a limited amount of movie-viewing data (i.e., a methodological stress test). The study motivation and discussion could be strengthened by more attention to relevant work with adults and/or more explanation of the importance of testing this question in infants (is the reason to test this question in infants purely methodological - i.e., as a way to negate the need for retinotopic tasks in subsequent research, given the time constraints of scanning human infants?).

We are grateful to the reviewer for giving us the opportunity to clarify the innovations of this research. We believe that this research contributes to our understanding of how infants process dynamic stimuli, demonstrates the viability and utility of movie experiments in

infants, and highlights the potential for new movie-based analyses (e.g., SRM). We have now consolidated these motivations in the introduction to more clearly motivate this work:

“The primary goal of the current study is to investigate whether movie-watching data recapitulates the organization of visual cortex. Movies drive strong and naturalistic responses in sensory regions while minimizing task demands[12, 13, 24] and thus are a proxy for typical experience. In adults, movies and resting-state data have been used to characterize the visual cortex in a data-driven fashion[25–27]. Movies have been useful in awake infant fMRI for studying event segmentation[28], functional alignment[29], and brain networks[30]. However, this past work did not address the granularity and specificity of cortical organization that movies evoke. For example, movies evoke similar activity in infants in anatomically aligned visual areas[28], but it remains unclear whether responses to movie content differ between visual areas (e.g., is there more similarity of function within visual areas than between31). Moreover, it is unknown whether structure within visual areas, namely visual maps, contributes substantially to visual evoked activity. Additionally, we wish to test whether methods for functional alignment can be used with infants. Functional alignment finds a mapping between participants using functional activity – rather than anatomy – and in adults can improve signal-to-noise, enhance across participant prediction, and enable unique analyses[27, 32–34].” Pg. 3-4

Furthermore, the introduction culminates in the following statement on what the analyses will tell us about the nature of movie-driven activity in infants:

“These three analyses assess key indicators of the mature visual system: functional specialization between areas, organization within areas, and consistency between individuals.” Pg. 5

Furthermore, in the discussion we revisit these motivations and elaborate on them further:

[Regarding homotopy:] “This suggests that visual areas are functionally differentiated in infancy and that this function is shared across hemispheres[31].” Pg. 19

[Regarding ICA:] “This means that the retinotopic organization of the infant brain accounts for a detectable amount of variance in visual activity, otherwise components resembling these maps would not be discoverable.” Pg. 19–20

[Regarding SRM:] “This is initial evidence that functional alignment may be useful for enhancing signal quality, like it has in adults[27,32,33], or revealing changing function over development[45].” Pg. 21

Additionally, we have expanded our discussion of relevant work that uses similar methods such as the excellent research from Knapen (2021) and others:

“In adults, movies and resting-state data have been used to characterize the visual cortex in a data-driven fashion[25-27].” Pg. 4

“We next explored whether movies can reveal fine-grained organization within visual areas by using independent components analysis (ICA) to propose visual maps in individual infant brains[25,26,35,42,43].” Pg. 9

Reviewer #2 (Public Review):

Summary:

This manuscript shows evidence from a dataset with awake movie-watching in infants, that the infant brain contains areas with distinct functions, consistent with previous studies using resting state and awake task-based infant fMRI. However, substantial new analyses would be required to support the novel claim that movie-watching data in

infants can be used to identify retinotopic areas or to capture within-area functional organization.

Strengths:

The authors have collected a unique dataset: the same individual infants both watched naturalistic animations and a specific retinotopy task. These data position the authors to test their novel claim, that movie-watching data in infants can be used to identify retinotopic areas.

Weaknesses:

To claim that movie-watching data can identify retinotopic regions, the authors should provide evidence for two claims:

- Retinotopic areas defined based only on movie-watching data, predict retinotopic responses in independent retinotopy-task-driven data.*
- Defining retinotopic areas based on the infant's own movie-watching response is more accurate than alternative approaches that don't require any movie-watching data, like anatomical parcellations or shared response activation from independent groups of participants.*

We thank the reviewer for their comments. Before addressing their suggestions, we wish to clarify that we do not claim that movie data can be used to identify retinotopic areas, but instead that movie data captures components of the within and between visual area organization as defined by retinotopic mapping. We recognize that this was not clear in our original manuscript and have clarified this point throughout, including in this section of the discussion:

“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. Additionally, anatomical alignment (i.e., averaging the maps from other participants and anatomically aligning them to a held-out participant) resulted in maps that were highly similar to the ground truth. Indeed, we previously[23] found that adult-defined visual areas were moderately similar to infants. While functional alignment with adults can outperform anatomical alignment methods in similar analyses[27], here we find that functional alignment with infants is inferior to anatomical alignment. Thus, if the goal is to define visual areas in an infant that lacks task-based retinotopy, anatomical alignment of other participants’ retinotopic maps is superior to using movie-based analyses, at least as we tested it.” Pg. 21

In response to the reviewer’s suggestion, we compare the maps identified by SRM to the averaged, anatomically aligned maps from infants. We find that these maps are highly similar to the task-based ground truth and we describe this result in a new section:

“We performed an anatomical alignment analog of the functional alignment (SRM) approach. This analysis serves as a benchmark for predicting visual maps using taskbased data, rather than movie data, from other participants. For each infant participant, we aggregated all other infant or adult participants as a reference. The retinotopic maps from these reference participants were anatomically aligned to the standard surface template, and then averaged. These averages served as predictions of the maps in the test participant, akin to SRM, and were analyzed equivalently (i.e., correlating the gradients in the predicted map with the gradients in the task-based map). These correlations (Table S4) are significantly higher than for functional alignment (using infants to predict spatial frequency, anatomical alignment <

functional alignment: $\Delta_{\text{Fisher } Z}$ $M=0.44$, $CI=[0.32-0.58]$, $p<.001$; using infants to predict meridians, anatomical alignment < functional alignment: $\Delta_{\text{Fisher } Z}$ $M=0.61$, $CI=[0.47-0.74]$, $p<.001$; using adults to predict spatial frequency, anatomical alignment < functional alignment: $\Delta_{\text{Fisher } Z}$ $M=0.31$, $CI=[0.21-0.42]$, $p<.001$; using adults to predict meridians, anatomical alignment < functional alignment: $\Delta_{\text{Fisher } Z}$ $M=0.49$, $CI=[0.39-0.60]$, $p<.001$). This suggests that even if SRM shows that movies can be used to produce retinotopic maps that are significantly similar to a participant, these maps are not as good as those that can be produced by anatomical alignment of the maps from other participants without any movie data.” Pg. 16–17

Note that we do not compare the anatomically aligned maps with the ICA maps statistically. This is because these analyses are not comparable: ICA is run within-participant whereas anatomical alignment is necessarily between-participant — either infant or adults. Nonetheless, an interested reader can refer to the Table where we report the results of anatomical alignment and see that anatomical alignment outperforms ICA in terms of the correlation between the predicted and task-based maps.

Both of these analyses are possible, using the (valuable!) data that these authors have collected, but these are not the analyses that the authors have done so far. Instead, the authors report the inverse of (1): regions identified by the retinotopy task can be used to predict responses in the movies. The authors report one part of (2), shared responses from other participants can be used to predict individual infants' responses in the movies, but they do not test whether movie data from the same individual infant can be used to make better predictions of the retinotopy task data, than the shared response maps.

So to be clear, to support the claims of this paper, I recommend that the authors use the retinotopic task responses in each individual infant as the independent "Test" data, and compare the accuracy in predicting those responses, based on:

- *The same infant's movie-watching data, analysed with MELODIC, when blind experimenters select components for the SF and meridian boundaries with no access to the ground-truth retinotopy data.*
- *Anatomical parcellations in the same infant.*
- *Shared response maps from groups of other infants or adults.*
- *(If possible, ICA of resting state data, in the same infant, or from independent groups of infants).*

Or, possibly, combinations of these techniques.

If the infant's own movie-watching data leads to improved predictions of the infant's retinotopic task-driven response, relative to these existing alternatives that don't require movie-watching data from the same infant, then the authors' main claim will be supported.

These are excellent suggestions for additional analyses to test the suitability for moviebased maps to replace task-based maps. We hope it is now clear that it was never our intention to claim that movie-based data could replace task-based methods. We want to emphasize that the discoveries made in this paper — that movies evoke fine-grained organization in infant visual cortex — do not rely on movie-based maps being better than alternative methods for producing maps, such as the newly added anatomical alignment.

The proposed analysis above solves a critical problem with the analyses presented in the current manuscript: the data used to generate maps is identical to the data used to

validate those maps. For the task-evoked maps, the same data are used to draw the lines along gradients and then test for gradient organization. For the component maps, the maps are manually selected to show the clearest gradients among many noisy options, and then the same data are tested for gradient organization. This is a double-dipping error. To fix this problem, the data must be split into independent train and test subsets.

We appreciate the reviewer's concern; however, we believe it is a result of a miscommunication in our analytic strategy. We have now provided more details on the analyses to clarify how double-dipping was avoided.

To summarize, a retinotopy task produced visual maps that were used to trace both area boundaries and gradients across the areas. These data were then fixed and unchanged, and we make no claims about the nature of these maps in this paper, other than to treat them as the ground truth to be used as a benchmark in our analyses. The movie data, which are collected independently from the same infant in the session, used the boundaries from the retinotopy task (in the case of homotopy) or were compared with the maps from the retinotopy task (in the case of ICA and SRM). In other words, the statement that "the data used to generate maps is identical to the data used to validate those maps" is incorrect because we generated the maps with a retinotopy task and validated the maps with the movie data. This means no double dipping occurred.

Perhaps a cause of the reviewer's interpretation is that the gradients used in the analysis are not clearly described. We now provide this additional description: "Using the same manually traced lines from the retinotopy task, we measured the intensity gradients in each component from the movie-watching data. We can then use the gradients of intensity in the retinotopy task-defined maps as a benchmark for comparison with the ICA-derived maps." Pg. 10

Regarding the SRM analyses, we take great pains to avoid the possibility of data contamination. To emphasize how independent the SRM analysis is, the prediction of the retinotopic map from the test participant does not use their retinotopy data at all; in fact, the predicted maps could be made before that participant's retinotopy data were ever collected. To make this prediction for a test participant, we need to learn the inversion of the SRM, but this only uses the movie data of the test participant. Hence, there is no double-dipping in the SRM analyses. We have elaborated on this point in the revision, and we remade the figure and its caption to clarify this point:

We also have updated the description of these results to emphasize how double-dipping was avoided:

"We then mapped the held-out participant's movie data into the learned shared space without changing the shared space (Figure 5c). In other words, the shared response model was learned and frozen before the held-out participant's data was considered.

This approach has been used and validated in prior SRM studies[45]." Pg. 14

The reviewer suggests that manually choosing components from ICA is double-dipping. Although the reviewer is correct that the manual selection of components in ICA means that the components chosen ought to be good candidates, we are testing whether those choices were good by evaluating those components against the task-based maps that were not used for the ICA. Our statistical analyses evaluate whether the components chosen were better than the components that would have been chosen by random chance. Critically: all decisions about selecting the components happen before the components are compared to the retinotopic maps. Hence there is no double-dipping in the selection of components, as the choice of candidate ICA maps is not informed by the ground-truth retinotopic maps. We now clarify what the goal of this process is in the results:

“Success in this process requires that 1) retinotopic organization accounts for sufficient variance in visual activity to be identified by ICA and 2) experimenters can accurately identify these components.” Pg. 10

The reviewer also alludes to a concern that the researcher selecting the maps was not blind to the ground-truth retinotopic maps from participants and this could have influenced the results. In such a scenario, the researcher could have selected components that have the gradients of activity in the places that the infant has as ground truth. The researcher who made the selection of components (CTE) is one of the researchers who originally traced the areas in the participants approximately a year prior to the identification of ICs. The researcher selecting the components didn't use the ground-truth retinotopic maps as reference, nor did they pay attention to the participant IDs when sorting the IC components. Indeed, they weren't trying to find participant specific maps per se, but rather aimed to find good candidate retinotopic maps in general. In the case of the newly added adult analyses, the ICs were selected before the retinotopic mapping was reviewed or traced; hence, no knowledge about the participant-specific ground truth could have influenced the selection of ICs. Even with this process from adults, we find results of comparable strength as we found in infants, as shown below. Nonetheless, there is a possibility that this researcher's previous experience of tracing the infant maps could have influenced their choice of components at the participant-specific level. If so, it was a small effect since the components the researcher selected were far from the best possible options (i.e., rankings of the selected components averaged in the 64th percentile for spatial frequency maps and the 68th percentile for meridian maps). We believe all reasonable steps were taken to mitigate bias in the selection of ICs.

Reviewer #3 (Public Review):

The manuscript reports data collected in awake toddlers recording BOLD while watching videos. The authors analyse the BOLD time series using two different statistical approaches, both very complex but do not require any a priori determination of the movie features or contents to be associated with regressors. The two main messages are that 1) toddlers have occipital visual areas very similar to adults, given that an SRM model derived from adult BOLD is consistent with the infant brains as well; 2) the retinotopic organization and the spatial frequency selectivity of the occipital maps derived by applying correlation analysis are consistent with the maps obtained by standard and conventional mapping.

Clearly, the data are important, and the author has achieved important and original results. However, the manuscript is totally unclear and very difficult to follow; the figures are not informative; the reader needs to trust the authors because no data to verify the output of the statistical analysis are presented (localization maps with proper statistics) nor so any validation of the statistical analysis provided. Indeed what I think that manuscript means, or better what I understood, may be very far from what the authors want to present, given how obscure the methods and the result presentation are.

In the present form, this reviewer considers that the manuscript needs to be totally rewritten, the results presented each technique with appropriate validation or comparison that the reader can evaluate.

We are grateful to the reviewer for the chance to improve the paper. We have broken their review into three parts: clarification of the methods, validation of the analyses, and enhancing the visualization.

Clarification of the methods

We acknowledge that the methods we employed are complex and uncommon in many fields of neuroimaging. That said, numerous papers have conducted these analyses on adults (Beckman et al., 2005; Butt et al., 2015; Guntupalli et al., 2016; Haak & Beckman, 2018; Knapen, 2021; Lu et al., 2017) and non-human primates (Arcaro & Livingstone, 2017; Moeller et al., 2009). We have redoubled our efforts in the revision to make the methods as clear as possible, expanding on the original text and providing intuitions where possible. These changes have been added throughout and are too vast in number to repeat here, especially without context, but we hope that readers will have an easier time following the analyses now.

Additionally, we updated Figures 3 and 5 in which the main ICA and SRM analyses are described. For instance, in Figure 3's caption we now add details about how the gradient analyses were performed on the components:

"We used the same lines that were manually traced on the task-evoked map to assess the change in the component's response. We found a monotonic trend within area from medial to lateral, just like we see in the ground truth." Pg. 11

Regarding Figure 5, we reconsidered the best way to explain the SRM analyses and decided it would be helpful to partition the diagram into steps, reflecting the analytic process. These updates have been added to Figure 5, and the caption has been updated accordingly.

We hope that these changes have improved the clarity of the methods. For readers interested in learning more, we encourage them to either read the methods-focused papers that debut the analyses (e.g., Chen et al., 2015), read the papers applying the methods (e.g., Guntupalli et al., 2016), or read the annotated code we publicly release which implements these pipelines and can be used to replicate the findings.

Validation of the analyses

One of the requests the reviewer makes is to validate our analyses. Our initial approach was to lean on papers that have used these methods in adults or primates (e.g., Arcaro, & Livingstone, 2017; Beckman et al., 2005; Butt et al., 2015; Guntupalli et al., 2016; Haak & Beckman, 2018; Knapen, 2021; Moeller et al., 2009) where the underlying organization and neurophysiology is established. However, we have made changes to these methods that differ from their original usage (e.g., we used SRM rather than hyperalignment, we use meridian mapping rather than traveling wave retinotopy, we use movie-watching data rather than rest). Hence, the specifics of our design and pipeline warrant validation.

To add further validation, we have rerun the main analyses on an adult sample. We collected 8 adult participants who completed the same retinotopy task and a large subset of the movies that infants saw. These participants were run under maximally similar conditions to infants (i.e., scanned using the same parameters and without the top of the head-coil) and were preprocessed using the same pipeline. Given that the relationship between adult visual maps and movie-driven (or resting-state) analyses has been shown in many studies (Beckman et al., 2005; Butt et al., 2015; Guntupalli et al., 2016; Haak & Beckman, 2018; Knapen, 2021; Lu et al., 2017), these adult data serve as a validation of our analysis pipeline. These adult participants were included in the original manuscript; however, they were previously only used to support the SRM analyses (i.e., can adults be used to predict infant visual maps). The adult results are described before any results with infants, as a way to engender confidence. Moreover, we have provided new supplementary figures of the adult results that we hope will be integrated with the article when viewing it online, such that it will be easy to compare infant and adult results, as per the reviewer's request.

As per the figures and captions below, the analyses were all successful with the adult participants: 1) Homotopic correlations are higher than correlations between comparable

areas in other streams or areas that are more distant within stream. 2) A multidimensional scaling depiction of the data shows that areas in the dorsal and ventral stream are dissimilar. 3) Using independent components analysis on the movie data, we identified components that are highly correlated with the retinotopy task-based spatial frequency and meridian maps. 4) Using shared response modeling on the movie data, we predicted maps that are highly correlated with the retinotopy task-based spatial frequency and meridian maps.

These supplementary analyses are underpowered for between-group comparisons, so we do not statistically compare the results between infants and adults. Nonetheless, the pattern of adult results is comparable overall to the infant results.

We believe these adult results provide a useful validation that the infant analyses we performed can recover fine-grained organization.

Enhancing the visualization

The reviewer raises an additional concern about the lack of visualization of the results. We recognize that the plots of the summary statistics do not provide information about the intermediate analyses. Indeed, we think the summary statistics can understate the degree of similarity between the components or predicted visual maps and the ground truth. Hence, we have added 6 new supplementary figures showing the intensity gradients for the following analyses: 1. spatial frequency prediction using ICA, 2. meridian prediction using ICA, 3. spatial frequency prediction using infant SRM, 4. meridian prediction using infant SRM, 5. spatial frequency prediction using adult SRM, and 6. meridian prediction using adult SRM.

We hope that these visualizations are helpful. It is possible that the reviewer wishes us to also visually present the raw maps from the ICA and SRM, akin to what we show in Figure 3A and 3B. We believe this is out of scope of this paper: of the 1140 components that were identified by ICA, we selected 36 for spatial frequency and 17 for meridian maps. We also created 20 predicted maps for spatial frequency and 20 predicted meridian maps using SRM. This would result in the depiction of 93 subfigures, requiring at least 15 new full-page supplementary figures to display with adequate resolution. Instead, we encourage the reader to access this content themselves: we have made the code to recreate the analyses publicly available, as well as both the raw and preprocessed data for these analyses, including the data for each of these selected maps.

Recommendations for the authors:

Reviewer #1 (Recommendations For The Authors):

(1) As mentioned in the public review, the authors should consider incorporating relevant adult fMRI research into the Introduction and explain the importance of testing this question in infants.

Our public response describes the several citations to relevant adult research we have added, and have provided further motivation for the project.

(2) The authors should conduct additional analyses to support their conclusion that movie data alone can generate accurate retinotopic maps (i.e., by comparing this approach to other available alternatives).

We have clarified in our public response that we did not wish to conclude that movie data alone can generate accurate retinotopic maps, and have made substantial edits to the text to emphasize this. Thus, because this claim is already not supported by our analyses, we do not think it is necessary to test it further.

(3) The authors should re-do the homotopy analyses using movie-defined ROIs (i.e., by splitting the movie-viewing data into independent folds for functional ROI definition and analyses).

As stated above, defining ROIs based on the movie content is not the intended goal of this project. Even if that were the general goal, we do not believe that it would be appropriate to run this specific analysis with the data we collected. Firstly, halving the data for ROI definition (e.g., using half the movie data to identify and trace areas, and then use those areas in the homotopy analysis to run on the other half of data) would qualitatively change the power of the analyses described here. Secondly, we would be unable to define areas beyond hV4/V3AB with confidence, since our retinotopic mapping only affords specification of early visual cortex. Thus we could not conduct the MDS analyses shown in Figure 2.

(4) If the authors agree that a primary contribution of this study and paper is to showcase what is possible to do with a limited amount of movie-viewing data, then they should make it clearer, sooner, how much usable movie data they have from infants. They could also consider conducting additional analyses to determine the minimum amount of fMRI data necessary to reveal the same detailed characteristics of functional responses in the visual cortex.

We agree it would be good to highlight the amount of movie data used. When the infant data is first introduced in the results section, we now state the durations:

“All available movies from each session were included (Table S2), with an average duration of 540.7s (range: 186–1116s).” Pg. 5

Additionally, we have added a homotopy analysis that describes the contribution of data quantity to the results observed. We compare the amount of data collected with the magnitude of same vs. different stream effect (Figure 1B) and within stream distance effect (Figure 1C). We find no effect of movie duration in the sample we tested, as reported below:

“We found no evidence that the variability in movie duration per participant correlated with this difference [of same stream vs. different stream] ($r=0.08$, $p=.700$).” Pg. 6-7

“There was no correlation between movie duration and the effect (Same > Adjacent: $r=-0.01$, $p=.965$, Adjacent > Distal: $r=-0.09$, $p=.740$).” Pg. 7

(5) If any of the methodological approaches are novel, the authors should make this clear. In particular, has the approach of visually inspecting and categorizing components generated from ICA and movie data been done before, in adults/other contexts?

The methods we employed are similar to others, as described in the public review.

However, changes were necessary to apply them to infant samples. For instance, Guntupalli et al. (2016) used hyperalignment to predict the visual maps of adult participants, whereas we use SRM. SRM and hyperalignment have the same goal — find a maximally aligned representation between participants based on brain function — but their implementation is different. The application of functional alignment to infants is novel, as is their use in movie data that is relatively short by comparison to standard adult data. Indeed, this is the most thorough demonstration that SRM — or any functional alignment procedure — can be usefully applied to infant data, awake or sleeping. We have clarified this point in the discussion.

“This is initial evidence that functional alignment may be useful for enhancing signal quality, like it has in adults[27,32,33], or revealing changing function over development[45], which

may prove especially useful for infant fMRI[52].” Pg. 21

(6) The authors found that meridian maps were less identifiable from ICA and movie data and suggest that this may be because these maps are more susceptible to noise or gaze variability. If this is the case, you might predict that these maps are more identifiable in adult data. The authors could consider running additional analyses with their adult participants to better understand this result.

As described in the manuscript, we hypothesize that meridian maps are more difficult to identify than spatial frequency maps because meridian maps are a less smooth, more fine-grained map than spatial frequency. Indeed, it has previously been reported (Moeller et al., 2009) that similar procedures can result in meridian maps that are constituted by multiple independent components (e.g., a component sensitive to horizontal orientations, and a separate component sensitive to vertical components). Nonetheless, we have now conducted the ICA procedure on adult participants and again find it is easier to identify spatial frequency components compared to meridian maps, as reported in the public review.

Minor corrections:

(1) Typo: Figure 3 title: "Example retinotopic task vs. ICA-based spatial frequency maps."

Fixed

(2) Given the age range of the participants, consider using "infants and toddlers"? (Not to diminish the results at all; on the contrary, I think it is perhaps even more impressive to obtain awake fMRI data from ~1-2-year-olds). Example: Figure 3 legend: "A) Spatial frequency map of a 17.1-monthold infant."

We agree with the reviewer that there is disagreement about the age range at which a child starts being considered a toddler. We have changed the terms in places where we refer to a toddler in particular (e.g., the figure caption the reviewer highlights) and added the phrase “infants and toddlers” in places where appropriate. Nonetheless, we have kept “infants” in some places, particularly those where we are comparing the sample to adults. Adding “and toddlers” could imply three samples being compared which would confuse the reader.

(3) Figure 6 legend: The following text should be omitted as there is no bar plot in this figure: "The bar plot is the average across participants. The error bar is the standard error across participants."

Fixed

(4) Table S1 legend: Missing first single quote: Runs'.

Fixed

Reviewer #2 (Recommendations For The Authors):

I request that this paper cite more of the existing literature on the fMRI of human infants and toddlers using task-driven and resting-state data. For example, early studies by (first authors) Biagi, Dehaene-Lambertz, Cusack, and Fransson, and more recent studies by Chen, Cabral, Truzzi, Deen, and Kosakowski.

We have added several new citations of recent task-based and resting state studies to the second sentence of the main text:

“Despite the recent growth in infant fMRI[1-6], one of 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[7].”

Reviewer #3 (Recommendations For The Authors):

In the following, I report some of my main perplexities, but many more may arise when the material is presented more clearly.

The age of the children varies from 5 months to about 2 years. While the developmental literature suggests that between 1 and 2 years children have a visual system nearly adult-like, below that age some areas may be very immature. I would split the sample and perhaps attempt to validate the adult SRM model with the youngest children (and those can be called infants).

We recognize the substantial age variability in our sample, which is why we report participant-specific data in our figures. While splitting up the data into age bins might reveal age effects, we do not think we can perform adequately powered null hypothesis testing of the age trend. In order to investigate the contribution of age, larger samples will be needed. That said, we can see from the data that we have reported that any effect of age is likely small. To elaborate: Figures 4 and 6 report the participant-specific data points and order the participants by age. There are no clear linear trends in these plots, thus there are no strong age effects.

More broadly, we do not think there is a principled way to divide the participants by age. The reviewer suggests that the visual system is immature before the first year of life and mature afterward; however, such claims are the exact motivation for the type of work we are doing here, and the verdict is still out. Indeed, the conclusion of our earlier work reporting retinotopy in infants (Ellis et al., 2021) suggests that the organization of the early visual cortex in infants as young as 5 months — the youngest infant in our sample — is surprisingly adult-like.

The title cannot refer to infants given the age span.

There is disagreement in the field about the age at which it is appropriate to refer to children as infants. In this paper, and in our prior work, we followed the practice of the most attended infant cognition conference and society, the International Congress of Infant Studies (ICIS), which considers infants as those aged between 0-3 years old, for the purposes of their conference. Indeed, we have never received this concern across dozens of prior reviews for previous papers covering a similar age range. That said, we understand the spirit of the reviewer’s comment and now refer to the sample as “infants and toddlers” and to older individuals in our sample as “toddlers” wherever it is appropriate (the younger individuals would fairly be considered “infants” under any definition).

Figure 1 is clear and an interesting approach. Please also show the average correlation maps on the cortical surface.

While we would like to create a figure as requested, we are unsure how to depict an area-by-area correlation map on the cortical surface. One option would be to generate a seed-based map in which we take an area and depict the correlation of that seed (e.g., vV1) with all other voxels. This approach would result in 8 maps for just the task-defined areas, and 17 maps for anatomically-defined areas. Hence, we believe this is out of scope of this paper, but an interested reader could easily generate these maps from the data we have released.

Figure 2 results are not easily interpretable. Ventral and dorsal V1-V3 areas represent upper or lower VF respectively. Higher dorsal and ventral areas represent both upper and lower VF, so we should predict an equal distance between the two streams. Again, how can we verify that it is not a result of some artifacts?

In adults, visual areas differ in their functional response properties along multiple dimensions, including spatial coding. The dorsal/ventral stream hypothesis is derived from the idea that areas in each stream support different functions, independent of spatial coding. The MDS analysis did not attempt to isolate the specific contribution of spatial representations of each area but instead tested the similarity of function that is evoked in naturalistic viewing. Other covariance-based analyses specifically isolate the contribution of spatial representations (Haak et al., 2013); however, they use a much more constrained analysis than what was implemented here. The fact that we find broad differentiation of dorsal and ventral visual areas in infants is consistent with adults (Haak & Beckman, 2018) and neonate non-human primates (Arcaro & Livingstone, 2017).

Nonetheless, we recognize that we did not mention the differences in visual field properties across areas and what that means. If visual field properties alone drove the functional response then we would expect to see a clustering of areas based on the visual field they represent (e.g., hV4 and V3AB should have similar representations). Since we did not see that, and instead saw organization by visual stream, the result is interesting and thus warrants reporting. We now mention this difference in visual fields in the manuscript to highlight the surprising nature of the result.

“This separation between streams is striking when considering that it happens *despite* differences in visual field representations across areas: while dorsal V1 and ventral V1 represent the lower and upper visual field, respectively, V3A/B and hV4 both have full visual field maps. These visual field representations can be detected in adults[41]; however, they are often not the primary driver of function[39]. We see that in infants too: hV4 and V3A/B represent the same visual space yet have distinct functional profiles.” Pg. 8

The reviewer raises a concern that the MDS result may be spurious and caused by noise. Below, we present three reasons why we believe these results are not accounted for by artifacts but instead reflect real functional differentiation in the visual cortex.

(1) Figure 2 is a visualization of the similarity matrix presented in Figure S1. In Figure S1, we report the significance testing we performed to confirm that the patterns differentiating dorsal and ventral streams — as well as adjacent areas from distal areas — are statistically reliable across participants. If an artifact accounted for the result then it would have to be a kind of systematic noise that is consistent across participants.

(2) One of the main sources of noise (both systematic and non-systematic) with infant fMRI is motion. Homotopy is a within-participant analysis that could be biased by motion. To assess whether motion accounts for the results, we took a conservative approach of regressing out the framewise motion (i.e., how much movement there is between fMRI volumes) from the comparisons of the functional activity in regions. Although the correlations numerically decreased with this procedure, they were qualitatively similar to the analysis that does not regress out motion:

“Additionally, if we control for motion in the correlation between areas --- in case motion transients drive consistent activity across areas --- then the effects described here are negligibly different (Figure S5).” Pg. 7

(3) We recognize that despite these analyses, it would be helpful to see what this pattern looks like in adults where we know more about the visual field properties and the function of

dorsal and ventral streams. This has been done previously (e.g., Haak & Beckman, 2018), but we have now run those analyses on adults in our sample, as described in the public review. As with infants, there are reliable differences in the homotopy between streams (Figure S1). The MDS results show that the adult data was more complex than the infant data, since it was best described by 3 dimensions rather than 2. Nonetheless, there is a rotation of the MDS such that the structure of the ventral and dorsal streams is also dissociable.

Figure 3 also raises several alternative interpretations. The spatial frequency component in B has strong activity ONLY at the extreme border of the VF and this is probably the origin of the strong correlation. I understand that it is only one subject, but this brings the need to show all subjects and to report the correlation. Also, it is important to show the putative average ICA for retinotopy and spatial frequencies across subjects and for adults. All methods should be validated on adults where we have clear data for retinotopy and spatial frequency.

The reviewer notes that the component in Figure 3 shows strong negative response in the periphery. It is often the case, as reported elsewhere (Moeller et al., 2009), that ICA extracts portions of visual maps. To make a full visual map would require combining components into a composite (e.g., a component that has a high response in the periphery and another component that has a high response in the fovea). If we were to claim that this component, or others like it, could replace the need for retinotopic mapping, then we would want to produce these composite maps; however, our conclusion in this project is that the topographic information of retinotopic maps manifest in individual components of ICA. For this purpose, the analysis we perform adequately assesses this topography.

Regarding the request to show the results for all subjects, we address this in the public response and repeat it here briefly: we have added 6 new figures to show results akin to Figure 3C and D. It is impractical to show the equivalent of Figure 3A and B for all participants, yet we do release the data necessary to see to visualize these maps easily.

Finally, the reviewer suggests that we validate the analyses on adult participants. As shown in Figure S3 and reported in the public response, we now run these analyses on adult participants and observe qualitatively similar results to infants.

How much was the variation in the presumed spatial frequency map? Is it consistent with the acuity range? 5-month-old infants should have an acuity of around 10c/deg, depending on the mean luminance of the scene.

The reviewer highlights an important weakness of conducting ICA: we cannot put units on the degree of variation we see in components. We now highlight this weakness in the discussion:

“Another limitation is that ICA does not provide a scale to the variation: although we find a correlation between gradients of spatial frequency in the ground truth and the selected component, we cannot use the component alone to infer the spatial frequency selectivity of any part of cortex. In other words, we cannot infer units of spatial frequency sensitivity from the components alone.” Pg. 20

Figure 5 pipeline is totally obscure. I presumed that I understood, but as it is it is useless. All methods should be clearly described, and the intermediate results should be illustrated in figures and appropriately discussed. Using such blind analyses in infants in principle may not be appropriate and this needs to be verified. Overall all these techniques rely on correlation activities that are all biased by head movement, eye movement, and probably the dummy sucking. All those movements need to be estimated

and correlated with the variability of the results. It is a strong assumption that the techniques should work in infants, given the presence of movements.

We recognize that the SRM methods are complex. Given this feedback, we remade Figure 5 with explicit steps for the process and updated the caption (as reported in the public review).

Regarding the validation of these methods, we have added SRM analyses from adults and find comparable results. This means that using these methods on adults with comparable amounts of data as what we collected from infants can predict maps that are highly similar to the real maps. Even so, it is not a given that these methods are valid in infants. We present two considerations in this regard.

First, as part of the SRM analyses reported in the manuscript, we show that control analyses are significantly worse than the real analyses (indicated by the lines on Figure 6). To clarify the control analysis: we break the mapping (i.e., flip the order of the data so that it is backwards) between the test participant and the training participants used to create the SRM. The fact that this control analysis is significantly worse indicates that SRM is learning meaningful representations that matter for retinotopy.

Second, we believe that this paper is a validation of SRM for infants. Infant fMRI is a nascent field and SRM has the potential to increase the signal quality in this population. We hope that readers will see these analyses as a proof of concept that SRM can be used in their work with infants. We have stated this contribution in the paper now.

“Additionally, we wish to test whether methods for functional alignment can be used with infants. Functional alignment finds a mapping between participants using functional activity – rather than anatomy – and in adults can improve signal-to-noise, enhance across participant prediction, and enable unique analyses[27,32-34].” Pg. 4

“This is initial evidence that functional alignment may be useful for enhancing signal quality, like it has in adults[27,32,33], or revealing changing function over development[45].” Pg. 21

Regarding the reviewer’s concern that motion may bias the results, we wish to emphasize the nature of the analyses being conducted here: we are using data from a group of participants to predict the neural responses in a held-out participant. For motion to explain consistency between participants, the motion would need to be timelocked across participants. Even if motion was time-locked during movie watching, motion will impair the formation of an adequate model that can contain retinotopic information. Thus, motion should only hurt the ability for a shared response to be found that can be used for predicting retinotopic maps. Hence, the results we observed are despite motion and other sources of noise.

What is M??? is it simply the mean value??? If not, how it is estimated?

M is an abbreviation for mean. We have now expanded the abbreviation the first time we use it.

Figure 6 should be integrated with map activity where the individual area correlation should be illustrated. Probably fitting SMR adult works well for early cortical areas, but not for more ventral and associative, and the correlation should be evaluated for the different masks.

With the addition of plots showing the gradients for each participant and each movie (Figures S10–S13) we hope we have addressed this concern. We additionally want to clarify that the regions we tested in the analysis in Figure 6 are only the early visual areas V1, V2, V3, V3A/B, and hV4. The adult validation analyses show that SRM works well for predicting the visual maps in these areas. Nonetheless, it is an interesting question for future research with more extensive retinotopic mapping in infants to see if SRM can predict maps beyond extrastriate cortex.

| *Occipital masks have never been described or shown.*

The occipital mask is from the MNI probabilistic structural atlas (Mazziotta et al., 2001), as reported in the original version and is shared with the public data release. We have added the additional detail that the probabilistic atlas is thresholded at 0% in order to be liberally inclusive.

“We used the occipital mask from the MNI structural atlas[63] in standard space -- defined liberally to include any voxel with an above zero probability of being labelled as the occipital lobe -- and used the inverted transform to put it into native functional space.” Pg. 27–28

| *Methods lack the main explanation of the procedures and software description.*

We hope that the additions we have made to address this reviewer’s concerns have provided better explanations for our procedures. Additionally, as part of the data and code release, we thoroughly explain all of the software needed to recreate the results we have observed here.

<https://doi.org/10.7554/eLife.92119.3.sa0>