

Current Biology

Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants

Highlights

- Category-selective responses are present within the first year of life
- FFA, PPA, and EBA are face, scene, and body selective, respectively, in infancy
- Selective responses did not reach significance in infant OFA, OPA, RSC, or FBA
- Category-selective responses are not explained by visual “protomaps”

Authors

Heather L. Kosakowski,
Michael A. Cohen, Atsushi Takahashi,
Boris Keil, Nancy Kanwisher,
Rebecca Saxe

Correspondence

heatherlkosakowski@gmail.com

In brief

Portions of the ventral visual pathway in adults selectively respond to faces, scenes, and bodies. Here, Kosakowski et al. show that 2- to 9-month-old human infants have face-, scene- and body-selective responses in FFA, PPA, and EBA, respectively, constraining theories of cortical development.

Article

Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants

Heather L. Kosakowski,^{1,4,5,*} Michael A. Cohen,^{1,2} Atsushi Takahashi,¹ Boris Keil,³ Nancy Kanwisher,¹ and Rebecca Saxe¹

¹Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, USA

²Department of Psychology and Program in Neuroscience, Amherst College, 220 South Pleasant Street, Amherst, MA, USA

³Institute of Medical Physics and Radiation Protection, Department of Life Science Engineering, Mittelhessen University of Applied Science, Giessen, Germany

⁴Twitter: @heatherlkos

⁵Lead contact

*Correspondence: heatherlkosakowski@gmail.com

<https://doi.org/10.1016/j.cub.2021.10.064>

SUMMARY

Three of the most robust functional landmarks in the human brain are the selective responses to faces in the fusiform face area (FFA), scenes in the parahippocampal place area (PPA), and bodies in the extrastriate body area (EBA). Are the selective responses of these regions present early in development or do they require many years to develop? Prior evidence leaves this question unresolved. We designed a new 32-channel infant magnetic resonance imaging (MRI) coil and collected high-quality functional MRI (fMRI) data from infants (2–9 months of age) while they viewed stimuli from four conditions—faces, bodies, objects, and scenes. We find that infants have face-, scene-, and body-selective responses in the location of the adult FFA, PPA, and EBA, respectively, powerfully constraining accounts of cortical development.

INTRODUCTION

The human mind is not homogeneous and equipotential but structured, containing a set of highly specialized mechanisms for processing particular domains of information, from perceiving faces to understanding language to thinking about other people's thoughts.¹ Much of this domain-specific structure of the mind is mirrored in the functional organization of the brain, which features cortical regions selectively engaged in processing these content domains.¹ How does this functional organization arise over development? Is our cognitive and neural machinery built slowly, over years of experience, or is much of that structure already present early in development? Here, we approach this question by asking whether the domain-specific organization of high-level visual cortex is already present in young infants.

A number of prior fMRI studies in children have argued that category-selective regions of the ventral visual pathway (VVP), including the fusiform face area (FFA),² parahippocampal place area (PPA),³ and extrastriate body area (EBA),⁴ develop very slowly, increasing in size throughout childhood and into adolescence.^{5–11} But those studies have not tested children younger than 3 years of age, focusing primarily instead on children ages 7 and up. To understand development, we need to know whether category-selective responses are present in infants. Research using electroencephalography (EEG), magnetoencephalography (MEG), and functional near-infrared spectroscopy (fNIRS) has reported neural responses in the infant brain to faces,^{12–14} scenes,¹⁴ and bodies,^{15,16} but these studies have used only a single control stimulus type, providing only weak tests of selectivity. Further, these methods lack the spatial

resolution to determine whether the recorded responses are coming from the same cortical locations as the adult FFA, PPA, and EBA. Only two publications have reported fMRI responses in the VVP of awake infants: one in humans ($n = 6$)¹⁷ and one in macaques ($n = 3$).¹⁸ These studies observed an adult-like spatial organization of responses to faces in both species^{17,18} and scenes in human infants.¹⁷ However, the activated regions did not show selective responses, instead responding as strongly to objects as to faces or scenes. Thus, no prior studies have reported selective responses in infants to faces, bodies, or scenes that can be localized to the FFA, PPA, or EBA, respectively.

Taken together, these results have led to a hypothesis that infant cortex is organized by feature-based protomaps that extend from primary sensory areas to higher level cortical regions^{18–22} and that selective responses to specific visual categories, such as faces, scenes, and bodies, develop later based on extensive visual experience with that category. This hypothesis predicts that category-selective responses will emerge in regions of cortex that correspond to overlapping protomap features. For example, because faces tend to be foveated and have substantial low spatial frequency and curvilinear content, the FFA will develop in a region of the protomap that responds preferentially to foveal, low spatial frequency, and curvilinear input. Conversely, scenes extend into the periphery and tend to have high spatial frequency and rectilinear content, so the PPA will develop in a peripheral region of the protomap that overlaps with preferences for high spatial frequency and rectilinear content. Importantly, this hypothesis predicts that responses to features of the protomaps, such as retinotopy, curvilinearity, rectilinearity, and spatial frequency, would arise earlier in

development than responses to high-level visual categories, such as faces, scenes, and bodies.²⁰

Much of the evidence for this hypothesis comes from the absence of category-selective responses in VVP as measured with fMRI.^{17,18} We reasoned to the contrary that the FFA, PPA, and EBA might be category selective but have gone undetected in prior infant studies due to limitations in the quality and amount of data obtained. It is extremely difficult to obtain high-quality fMRI data from awake infants because they have short attention spans,²³ cannot follow verbal instructions, and tend to move in the scanner. Further, the tools used for collecting and processing infant data are less advanced than those for adults. To test our hypothesis, we scanned a much larger number of infants and devised a variety of technical innovations to increase both the quantity and quality of data we could obtain (Figure 1A; STAR Methods). To sustain infants' attention, we used engaging, colorful stimuli depicting faces, bodies, objects, and scenes (Figure 1B). We recruited 87 infants and were able to collect usable fMRI data from 52 of them (2.1–9.7 months; STAR Methods). The primary results reported here reflect data from 26 infants (STAR Methods) scanned on a second-generation custom 32-channel infant head coil (referred to here as *Coil 2021*) with a higher signal-to-noise ratio (SNR) than an equivalent adult coil²⁴ and MR-safe headphones that provided better hearing protection, enabling the collection of infant fMRI data at a resolution commonly used for adult studies.

RESULTS

Do human infants have face-, scene-, or body-selective responses in the VVP? As an initial exploratory analysis, we first conducted voxelwise whole-brain analyses testing for higher responses to each condition of interest (i.e., faces, scenes, and bodies) compared to objects. Individual subject activation maps and group random effects analyses ($n = 23$) found higher responses to faces than objects in the fusiform gyrus, higher responses to scenes than objects in the parahippocampal gyrus, and higher responses to bodies than objects in lateral-occipital cortex (Figure 2), areas that correspond to the location of the FFA, PPA, and EBA in adults. While suggestive, these responses did not reach thresholds for statistical significance when correcting for multiple comparisons across the whole brain.

FFA, PPA, and EBA are category selective in infancy

To test the anatomical location and selectivity of these regions more directly and stringently, we next conducted a standard fROI analysis using previously described anatomical parcels that constrain localization of the FFA, PPA, and EBA in adults.²⁵ To accommodate the imperfect registration of infant functional data to adult anatomical templates, we uniformly increased the size of each parcel by a small amount (see similar fROI results for unexpanded parcels in Table S1). We then selected the top 5% of voxels in each parcel, in each participant that had a numerically greater response to the preferred category corresponding to that parcel (e.g., faces for the FFA) than to objects, in one subset of the data. Then, in independent data in the same participant, we quantified the response in those voxels to each of the four conditions using linear mixed effects models (STAR

Methods). In the main dataset, 19 infants had enough data to be included in this analysis.

This fROI analysis identified voxels in the FFA parcel where responses to faces were significantly greater than the response to each of the other three categories (all $p < 0.001$; Figure 3A; Table S1), scene responses in the PPA parcel that were significantly greater than each of the other three categories (all $p < 0.01$; Figure 4A; Table S1; though not higher than baseline), and body responses in the EBA parcel that were significantly greater than each of the other three categories (all $p < 0.05$; Figure 5A; Table S1).

How much of each cortical parcel contains the predicted pattern of selectivity? We conducted the same fROI analysis in each parcel, varying the number of voxels selected from the top 1% of voxels to 100% of voxels. In the FFA, we observe a selective response to faces in held-out data when selecting anywhere from 1% to 30% of voxels (Figure 3A), suggesting that face-selective responses in infant FFA spans about 217.5 mm³ of cortical territory. In the PPA, the scene response is greater than the response to all other conditions, even when 100% of voxels are included in the analysis (Figure 4A), suggesting that scene responses in infant PPA covers at least 1,320 mm³. Finally, in the EBA, the response to bodies is selective only when the top 1%, 2%, or 5% of voxels are selected (Figure 5A) or approximately 131.6 mm³ (note that these areas are estimates in the volume, not on the cortical surface). Thus, selective responses in infant fROIs are robust to analytic parameters for voxel selection; and face-, scene-, and body-selective responses are present in the locations predicted from prior studies of adults.

Are face-, scene-, and body-selective responses found distinctively in their predicted locations? In each of the three parcels (FFA, PPA, and EBA), we selected voxels using all three contrasts (faces > objects, scenes > objects, and bodies > objects) at each threshold for voxel selection (top 1%–100%) and then computed the selectivity of these voxels in held-out data. In the FFA parcel, there were many face-selective voxels and no scene- or body-selective voxels (Figure 6A). In the PPA, there were many scene-selective voxels and no body- or face-selective voxels (Figure 6B). In the EBA, both the bodies > objects and faces > objects contrasts identified reliable voxels (Figure 6C); however, while the top 1%–5% of body-responding voxels were body selective (significantly higher than faces, objects, and scenes each independently), most face-responding voxels in the EBA parcel responded to both faces and bodies. In the EBA parcel as a whole, the majority of voxels responded to faces, bodies, and objects more than to scenes (Figure 5A).

Taken together, these results indicate that the FFA, PPA, and EBA are present in infancy, in the same anatomical location and with a selective response to the same preferred category as adults.

Higher quality fMRI data led to observation of selective responses

Why do we find face-, scene-, and body-selective responses when Deen et al.¹⁷ did not? The face, body, object, and scene videos used in the present study were a subset of the the same videos used in Deen et al.,¹⁷ and the average age of the infants across the studies is similar. However, in contrast to Deen

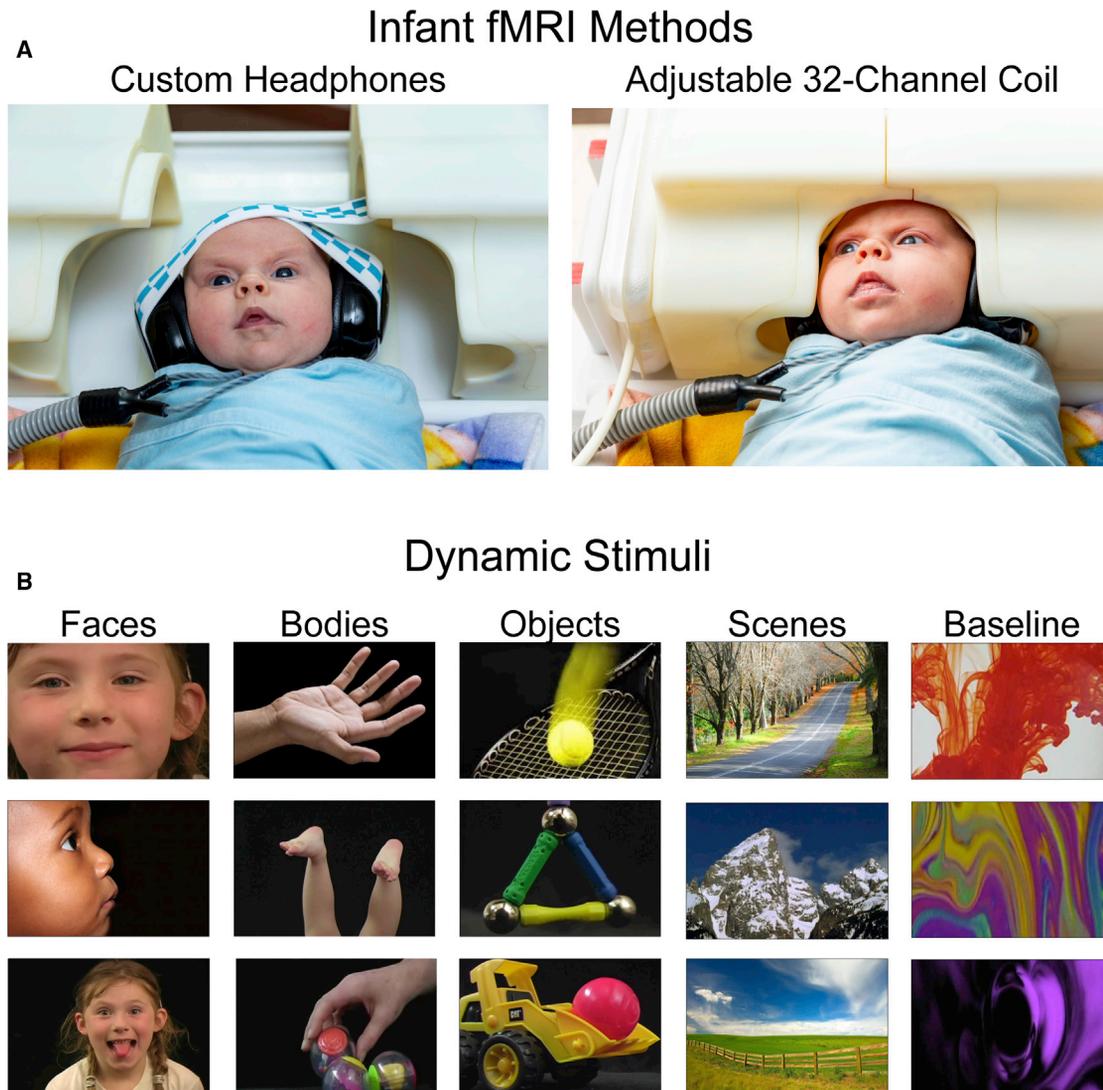


Figure 1. Infant scanning procedures and stimuli

(A) Infants were swaddled, wore custom MR-safe headphones to protect their hearing, and were scanned in a newly designed 32-channel head coil (*Coil 2021*)²⁴ that was shaped like a cradle and had adjustable frontal coils.

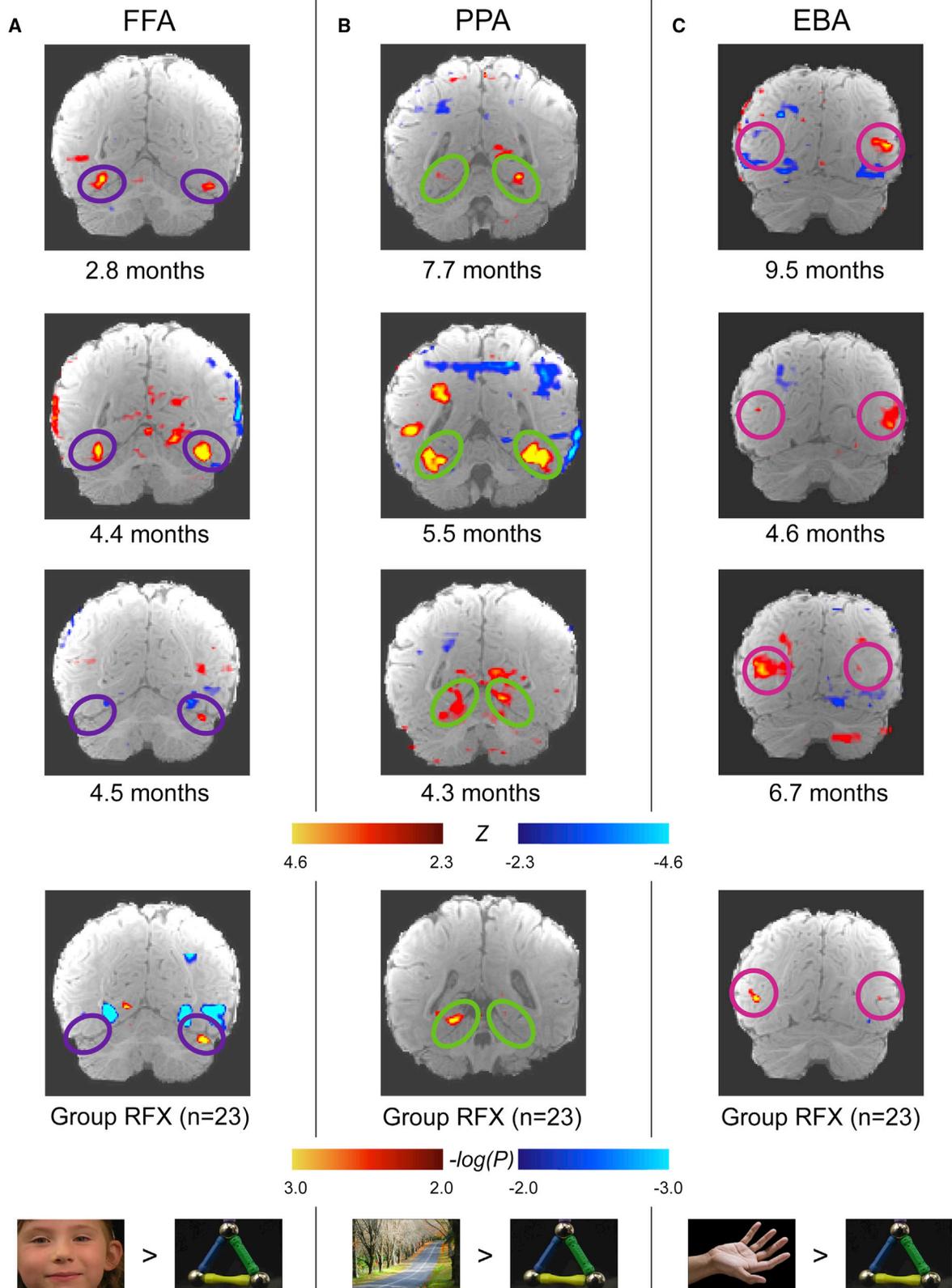
(B) Examples of face, body, object, scene, and baseline stimuli. Videos were 2.7 s long. To enhance infant attention, each video was followed by a 300-ms presentation of a still image from the same condition. Stimuli were displayed on a mirror over the infant's eyes and played continuously while the infant was awake, content, and attending.

et al.,¹⁷ the current study used a new coil (*Coil 2021*), a different pulse sequence producing higher resolution data, and included data from more subjects; there are also some differences in the analyses. To test which of these factors explained the difference in results between the present study and those of Deen et al.,¹⁷ we analyzed a third dataset we had collected on the first-generation custom 32-channel infant head coil (referred to as *Coil 2011*), before *Coil 2021* was built. For all subjects in the *Coil 2011* dataset (which did not overlap with the current data or the data published by Deen et al.¹⁷), we used the same pulse sequence and coil used by Deen et al.¹⁷ This enabled us to test whether the use of *Coil 2021* and/or the higher resolution acquisition sequence could explain the observed difference in results. Indeed, in matched fROI analyses of the *Coil 2021* data and the

Coil 2011 data ($n = 39$), we found significant selectivity for faces, places, and bodies in *Coil 2021* data, but not *Coil 2011* data, and a significant experiment (*Coil 2011* versus *Coil 2021*) by stimulus condition by fROI interaction (Figure S3C; $F(6,108) = 2.38$; $\eta^2 = 0.013$; $p = 0.013$). These results indicate that it is the new coil and pulse sequence of the current study that enabled us to find selectivity when Deen et al.¹⁷ did not.

Visual statistics are not sufficient to explain category-selective responses

Are face- and scene-selective responses in infants driven by low-level visual features of the stimuli? The protomap hypothesis predicts that the region that will develop into FFA initially responds preferentially to low spatial frequency and curvilinear



(legend on next page)

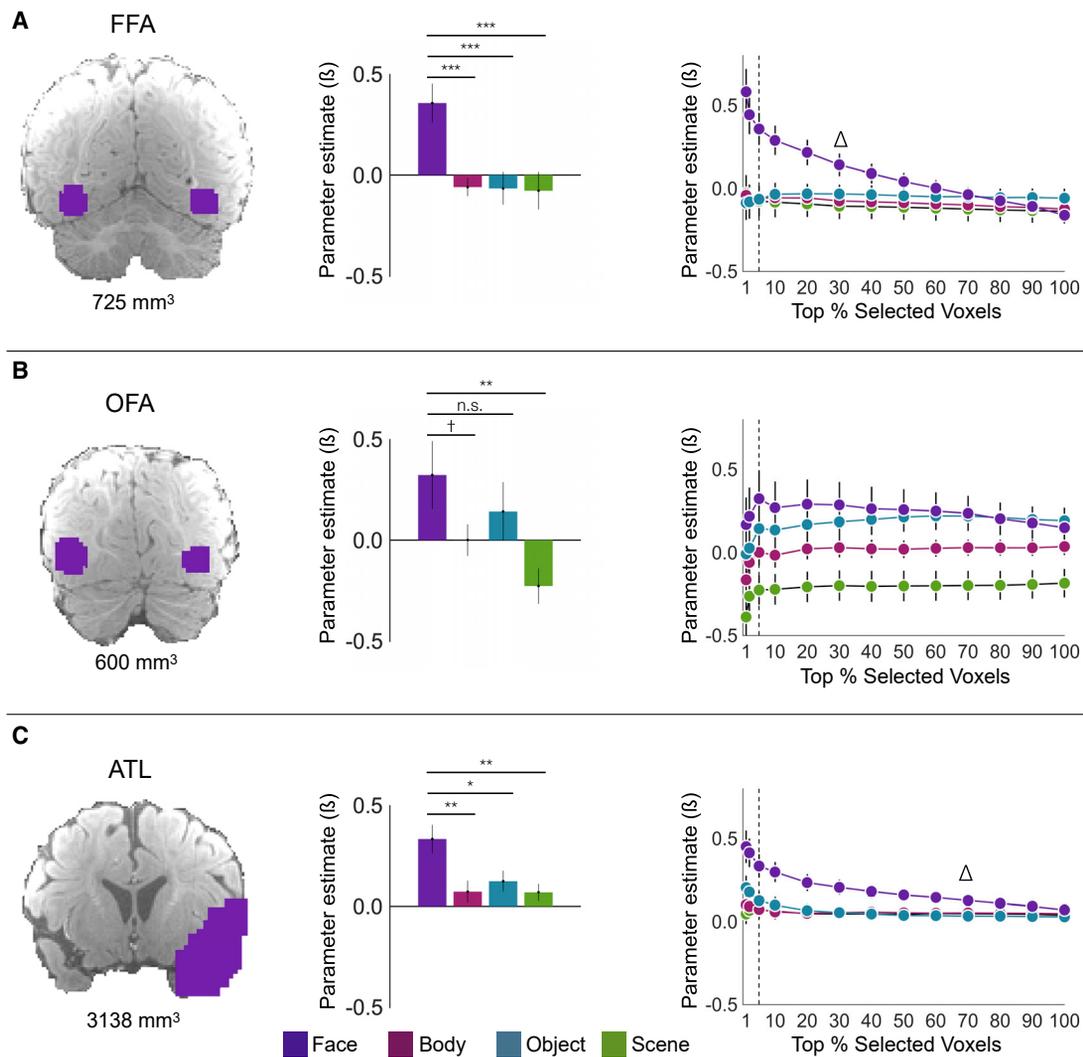


Figure 3. Face selectivity in the infant brain

An fROI analysis ($n = 20$) with Coil 2021 data was conducted with adult-constrained parcels of the FFA,²⁵ OFA,²⁵ and ATL²⁶ that were uniformly enlarged. All parcels are displayed on an anatomical image of an infant brain with face parcels shaded purple. Bar charts show the average response across participants in each fROI to each stimulus category (compared to baseline) in data independent of that used to define the fROI. Error bars are standard error of the mean accounting for within-subject variability.²⁷ Symbols used to report one-tailed statistics from linear mixed effect models: $\dagger p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Statistics are reported in Table S1. Line graphs show selectivity analysis with different proportions of voxels selected with the average response in independent data for faces plotted in purple, bodies plotted in pink, objects plotted in teal, and scenes plotted in green. The vertical dashed line marks the top 5% and corresponds to the bar charts. The highest proportion of voxels with face-selective responses is indicated with a black triangle. Error bars are standard error of the mean. See Figure S3 and Table S2 for comparison of *Coil 2011* and *Coil 2021* datasets.

input. In our paradigm, the stimuli with most curvilinear and low spatial frequency input were the object videos and the swirly baseline (Figure S2). Yet the response to faces in FFA is significantly greater than the response to both baseline and object videos ($p < 0.05$; Table S1). Therefore, curvilinearity and low

spatial frequency are unlikely to account for the selective response to faces in FFA.

Similarly, the protomap hypothesis predicts that the region that will develop into PPA initially responds preferentially to high spatial frequency and rectilinear input. In our stimuli,

Figure 2. Stereotyped location of category responses in the infant brain

Representative individual subject statistical maps (rows 1–3, threshold $p < 0.01$ transformed to Z for visualization) and group random effects analyses ($n = 23$, bottom row; threshold $-\log(p) = 2.0$) revealed higher responses to faces > objects in the fusiform gyrus (A; location of adult FFA outlined in purple), to scenes > objects in the parahippocampal gyrus (B; location of adult PPA outlined in green), and to bodies > objects in lateral-occipital cortex (C; location of adult EBA outlined in pink). Individual and group maps displayed on representative infant anatomical image collected on the same coil are shown (left hemisphere shown on the left and right on right). See Figure S1 for multiple slices through each region from the group maps.

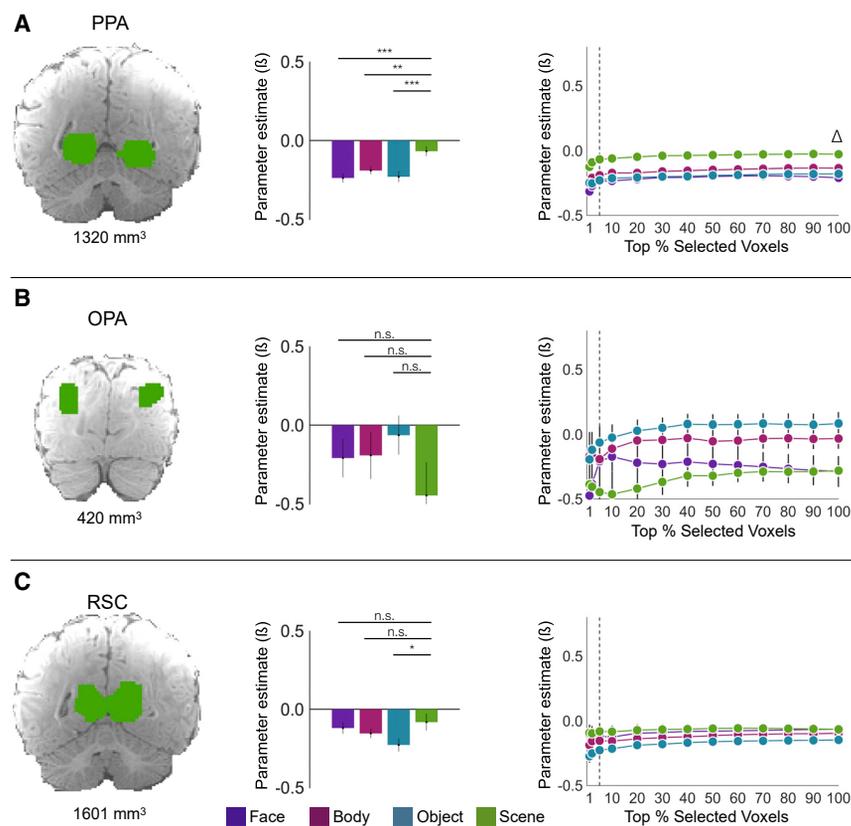


Figure 4. Scene selectivity in the infant brain

An fROI analysis ($n = 20$) with *Coil 2021* data was conducted with adult-constrained parcels of the PPA,²⁵ OPA,²⁵ and RSC²⁵ that were uniformly enlarged. All parcels are displayed on an anatomical image of an infant brain with scene parcels shaded green. Bar charts show the average response across participants in each fROI to each stimulus category (compared to baseline) in data independent of that used to define the fROI. Error bars are standard error of the mean accounting for within-subject variability.²⁷ Symbols used to report one-tailed statistics from linear mixed effect models: [†] $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Statistics are reported in [Table S1](#). Line graphs show selectivity analysis with different proportions of voxels selected with the average response in independent data for faces plotted in purple, bodies plotted in pink, objects plotted in teal, and scenes plotted in green. The vertical dashed line marks the top 5% and corresponds to the bar charts. The highest proportion of voxels with scene-selective responses is indicated with a black triangle. Error bars are standard error of the mean. See [Figure S3](#) and [Table S2](#) for comparison of *Coil 2011* and *Coil 2021* datasets.

rectilinear input was highest in the object movies ([Figure S2](#)), yet the PPA responded significantly more to scenes than objects. High spatial frequency input was highest in the scene movies and so might explain early PPA responses. However, one unexpected feature of the PPA response could not be explained by either a response to high spatial frequencies or category selectivity: the response in PPA was similar for scene movies and for the swirly baseline ($p > 0.2$; [Table S1](#)), despite scenes containing substantially more high spatial frequency (as well as semantic scene) content. It is unclear why the scene and swirly baseline stimuli evoked similar responses in PPA; future research should investigate the optimal stimuli for infant PPA.

In summary, the responses of the FFA and PPA in infants cannot be easily accounted for in terms of lower-level visual features or “protomaps” (see [Discussion](#)).^{18,20–22,28–32}

Selectivity observed in some, but not all, high-level visual areas that are selective in adults

In addition to the FFA, PPA, and EBA, other occipitotemporal regions exist in adults that respond selectively to faces, scenes, and bodies. Are these regions also present in infants? The occipital face area (OFA) and occipital place area (OPA) are located in lateral occipital cortex (with EBA between them) and are thought to reside earlier in the visual hierarchy than FFA³³ and PPA,³⁴ respectively. Thus, a posterior to anterior development of face and scene selectivity would predict that OFA and OPA would acquire characteristic functional specificity prior to FFA and PPA.²¹ However, fROI analyses using

adult parcels for OFA and OPA, with the same method just described, did not find significant face selectivity in the OFA or scene selectivity in the OPA (all $p > 0.05$ except OFA $F > S$ $p = 0.004$; [Figures 3B](#) and [4B](#); [Table S1](#)).

What about regions thought to reside later in the visual hierarchy, such as the face-selective response in anterior temporal lobe (ATL)? An fROI analysis with an adult ATL parcel²⁶ found that face responses in infant ATL were selective (all $p \leq 0.01$; [Figure 3C](#); [Table S1](#)). However, an fROI analysis using an adult parcel for the fusiform body area (FBA), a body-selective region thought to arise later in the neural hierarchy than EBA,³⁵ found only numerically, but not significantly, greater responses to bodies than the other three conditions in the FBA parcel (all $p > 0.05$; [Figure 5B](#); [Table S1](#)). Similarly, the scene response in the retrosplenial cortex (RSC) was numerically, but not significantly, greater than the response to all of the other conditions ($S > F$ $p = 0.3$; $S > B$ $p = 0.1$; $S > O$ $p = 0.02$; [Figure 4C](#); [Table S1](#)).

Thus, we find significantly category-selective responses in infant FFA and ATL for faces, PPA for scenes, and EBA for bodies, but not in other regions.

DISCUSSION

Here, we report that the FFA, PPA, and EBA are present in infants, in the same location, and with qualitatively similar selectivities, as adults. Continued development may expand and refine these regions over subsequent years,^{5–11} but the existence of category-selective mechanisms in the brain evidently does not require years of visual experience and maturation.

Our finding of face, scene, and body selectivity in infant FFA, PPA, and EBA differs from the earlier findings of Deen et al.,¹⁷

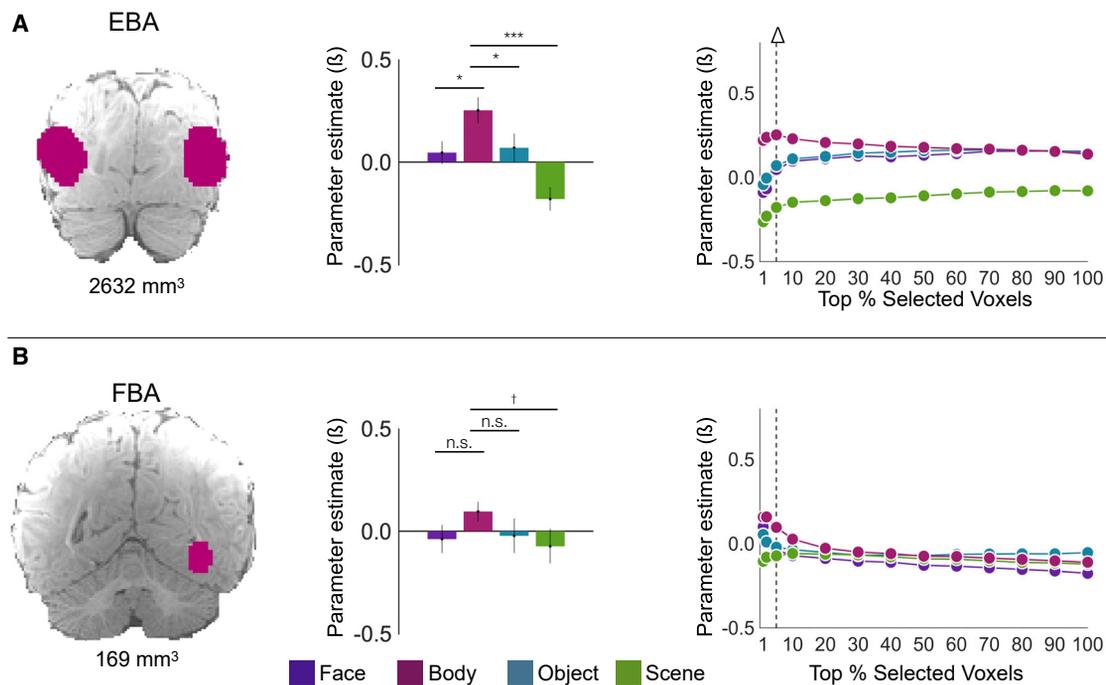


Figure 5. Body selectivity in the infant brain

An fROI analysis ($n = 20$) with *Coil 2021* data was conducted with adult-constrained parcels of the FBA²⁵ and EBA²⁵ that were uniformly enlarged. All parcels are displayed on an anatomical image of an infant brain with body parcels shaded pink. Bar charts show the average response across participants in each fROI to each stimulus category (compared to baseline) in data independent of that used to define the fROI. Error bars are standard error of the mean accounting for within-subject variability.²⁷ Symbols used to report one-tailed statistics from linear mixed effect models: † $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Statistics are reported in Table S1. Line graphs show selectivity analysis with different proportions of voxels selected with the average response in independent data for faces plotted in purple, bodies plotted in pink, objects plotted in teal, and scenes plotted in green. The vertical dashed line marks the top 5% and corresponds to the bar charts. The highest proportion of voxels with body-selective responses is indicated with a black triangle. Error bars are standard error of the mean. See Figure S3 and Table S2 for comparison of *Coil 2011* and *Coil 2021* datasets.

who found preferential, but not fully category-selective, responses in these regions in infants. The most likely account of this difference is that the current study benefited from the higher quality data afforded by our new infant coil²⁴ and the higher resolution pulse sequence.

Our findings inform ongoing debates about cortical development, including the question of how visual experience influences the emergence of functionally selective regions in VVP. These regions appear in approximately the same cortical location in every individual, implying that patches of cortex are predisposed for specific functions. However, the nature of these predispositions is unknown and disputed. On one view, visual cortex is organized at birth into “protomaps”: smooth, orthogonal spatial gradients of neurons tuned along low-level visual feature dimensions like retinotopy, spatial frequency, and curvilinearity.^{18,20–22,28–32} The natural statistics of early visual experience cause frequent co-activation of neurons with certain responses (e.g., faces are low frequency and curvy and typically foveated; scenes are high frequency and boxy and experienced in the periphery). Extensive visual experience thus leads to the slow emergence of cortical regions selectively responsive to these stimulus categories in cortical regions already tuned for their low-level correlates. Evidence for this view comes from fMRI studies with macaques. In early infancy (~ 30 days), V1 and regions that later become face-selective respond similarly to intact faces and

pixelated faces.¹⁸ In juvenile macaques (~ 200 days of age) the response profiles between V1 and face-selective regions begin to differentiate as face-selective regions (but not V1) begin to respond less to pixelated faces.¹⁸

A key prediction of the protomap hypothesis is that protomaps for spatial frequency and curvilinearity are present across the primate brain prior to functionally specialized responses.²⁰ However, our infant FFA responses do not have a pattern of response that corresponds to either low spatial frequency or curvilinearity, and our infant PPA responses do not have a pattern of response that corresponds to either high spatial frequency or rectilinearity. Thus, our data place important constraints on the protomap framework. If an underlying protomap combined with visual experience drives selectivity of domain specific regions like the FFA, PPA, and EBA, only a few months of visual experience must be sufficient to acquire that selectivity.

Relatedly, some theorists have proposed that selectivity in cortical regions might emerge in sequence, following the hierarchy of bottom-up visual processing in VVP. On this view, low-level representations in early visual cortex might emerge first, followed by mid-level representations in lateral-occipital cortical regions, culminating in higher level visual representations in the VVP. For example, fMRI studies of infant macaques found that blood-oxygen-level-dependent (BOLD) responses to motion emerged first in V1, followed by V4, and then MT.³⁶

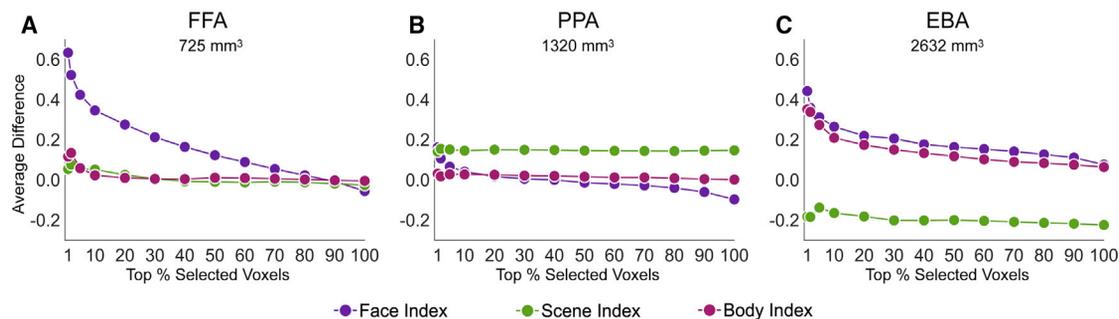


Figure 6. Predicted selectivities are the highest in predicted regions

The top n% of voxels for three contrasts (faces > objects, scenes > objects, and bodies > objects) were selected in the FFA (A), PPA (B), and EBA (C) parcels. Responses were measured in held-out data. Face index: faces – mean(bodies, scenes, objects); scene index: scenes – mean(faces, bodies, objects); body index: bodies – mean(faces, scenes, objects).

(A) In the FFA parcel, voxels selected by the faces > objects contrast, up to the top 30% of the parcel, had a significant face index in held-out data. No other contrast identified significant category preferences at any threshold.

(B) In the PPA parcel, voxels selected by the scenes > objects contrast had a significant scene index in held-out data. No other contrast identified significant category preferences at any threshold.

(C) In the EBA parcel, voxels selected by bodies > objects had a significant body index, and voxels selected by faces > objects had a significant face index. For all charts, error bars are standard error of the mean.

A feedforward pattern of development has similarly been suggested for VVP such that OFA and OPA would develop prior to FFA and PPA, respectively, because they are presumed to lie earlier in the visual hierarchy (the hierarchical relationship of FBA and EBA is unclear).^{37–44} In contrast to this prediction, we observed scene selectivity in PPA, but not its “precursor” OPA, and we observed face selectivity in FFA, but this selectivity did not reach significance in its precursor OFA. These results should be interpreted with caution, though, given that null results in infant fMRI are always suspect and the OPA and OFA are smaller and less robust regions, even in adults. Nevertheless, it is intriguing that our data show no evidence of feedforward hierarchical emergence of category selectivity in VVP in infants.

If not entirely driven by low-level feature protomaps and bottom-up input from earlier visual areas, what else might constrain the development of functionally selective regions in infants? One possibility is that, in infants, VVP regions have distinctive pre-existing long-range connectivity not only to early visual regions but also to non-visual regions (e.g., in parietal and frontal cortices). These patterns of connectivity can be used to accurately predict the location of functionally selective regions in infants⁴⁵ and children.⁴⁶ Further, these connections exist prior to the emergence of selective responses in at least some cases: for example, patterns of connectivity in pre-readers can be used to predict the future location of the visual word form area when the same child learns to read.⁴⁶ As further evidence for connectivity from non-visual areas, the putative VWFA has stronger functional correlations with putative language areas compared to functional correlations between nearby FFA and putative language areas.⁴⁷ Finally, congenitally blind humans have auditory and tactile responses in the FFA that are selective for faces^{48,49} and distinctive patterns of resting functional correlation between the “blind FFA” and both visual and non-visual regions.⁴⁸ Thus, we speculate that the location of FFA, PPA, and EBA may be influenced by pre-existing long-range connectivity to both early visual regions and also regions in parietal and frontal cortex.

Many fundamental open questions remain. First, although we found clear evidence of category selectivity at a much earlier age than previously reported, 5-month-old infants already have hundreds of hours of visual experience. What, if any, visual experience is necessary for the construction or maintenance of category-selective regions? Would purely visual exposure to faces, places, and bodies be sufficient to evoke category-selective responses?⁵⁰ Or must infants experience faces paired with social interaction, bodies paired with goal-directed actions, and places paired with self-motion through the environment? Second, are the representations extracted and computations conducted in these regions similar to those of adults or do these regions undergo major functional change between infancy and adulthood, and if so, what experience is necessary to produce that change? Research into these questions will ultimately provide scientific answers to the long-standing philosophical question of the origins of the human mind.

STAR★METHODS

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

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
- [METHOD DETAILS](#)
 - Stimuli
 - Data Collection
 - Custom Head Coils
 - Data Selection (subrun creation)
 - Preprocessing
 - Data registration
 - Parcels / Search Spaces

● **QUANTIFICATION AND STATISTICAL ANALYSIS**

- Subject-level Beta and Contrast Maps
- Group Random Effects Analyses
- Functional Region of Interest
- Category-Selectivity
- Models of stimulus visual statistics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2021.10.064>.

ACKNOWLEDGMENTS

This research was carried out at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. The authors thank Lyneé Herrera for help collecting the Coil 2011 dataset and Isabel Nicholson for help collecting the Coil 2021 dataset, as well as Nayanika Das, Isabel Nicholson, and Somaia Saba for help with registrations. We also thank Steven Shannon, members of Saxe Lab, and members of the Kanwisher Lab for help during recruitment and data collection; Stefano Anzellotti, Dana Boebinger, Tyler Bonnen, Katharina Dobs, Anna Ivanova, Frederik Kamps, Owen Kunhardt, and Somaia Saba for helpful comments on the manuscript; Kirsten Lydic for code and statistics transposition review; Hannah A. LeBlanc for everything; and all the infants and their families. We gratefully acknowledge support of this project by National Science Foundation (graduate fellowship to H.L.K. and Collaborative Research Award no. 1829470 to M.A.C.); NIH (no. 1F99NS124175 to H.L.K.; no. R21-HD090346-02 to R.S.; and no. DP1HD091947 to N.K.); shared instrumentation grant S10OD021569 for the MRI scanner; the McGovern Institute for Brain Research at MIT; and the Center for Brains, Minds and Machines (CBMM), funded by an NSF STC award (CCF-1231216).

AUTHOR CONTRIBUTIONS

H.L.K. and R.S. designed the study. H.L.K. collected the data. A.T. and B.K. provided technical support in the development of the coil and sequence parameters. H.L.K. analyzed the data with input from and supervision by M.A.C., N.K., and R.S. H.L.K., M.A.C., N.K., and R.S. wrote the manuscript. All authors provided feedback on the final version.

DECLARATION OF INTERESTS

N.K. served on the advisory board of *Current Biology*.

Received: August 4, 2021

Revised: September 27, 2021

Accepted: October 28, 2021

Published: November 15, 2021

REFERENCES

1. Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. USA* *107*, 11163–11170.
2. Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* *17*, 4302–4311.
3. Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* *392*, 598–601.
4. Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science* *293*, 2470–2473.
5. Grill-Spector, K., Golarai, G., and Gabrieli, J. (2008). Developmental neuroimaging of the human ventral visual cortex. *Trends Cogn. Sci.* *12*, 152–162.
6. Golarai, G., Liberman, A., and Grill-Spector, K. (2017). Experience shapes the development of neural substrates of face processing in human ventral temporal cortex. *Cereb. Cortex* *27*, 1229–1244.
7. Golarai, G., Ghahremani, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D.E., and Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.* *10*, 512–522.
8. Meissner, T.W., Nordt, M., and Weigelt, S. (2019). Prolonged functional development of the parahippocampal place area and occipital place area. *Neuroimage* *191*, 104–115.
9. Cohen, M.A., Dilks, D.D., Koldewyn, K., Weigelt, S., Feather, J., Kell, A.J., Keil, B., Fischl, B., Zöllei, L., Wald, L., et al. (2019). Representational similarity precedes category selectivity in the developing ventral visual pathway. *Neuroimage* *197*, 565–574.
10. Peelen, M.V., Glaser, B., Vuilleumier, P., and Eliez, S. (2009). Differential development of selectivity for faces and bodies in the fusiform gyrus. *Dev. Sci.* *12*, F16–F25.
11. Kamps, F.S., Hendrix, C.L., Brennan, P.A., and Dilks, D.D. (2020). Connectivity at the origins of domain specificity in the cortical face and place networks. *Proc. Natl. Acad. Sci. USA* *117*, 6163–6169.
12. de Heering, A., and Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife* *4*, e06564.
13. Kobayashi, M., Otsuka, Y., Kanazawa, S., Yamaguchi, M.K., and Kakigi, R. (2014). The processing of faces across non-rigid facial transformation develops at 7 month of age: a fNIRS-adaptation study. *BMC Neurosci.* *15*, 81.
14. Powell, L.J., Deen, B., and Saxe, R. (2018). Using individual functional channels of interest to study cortical development with fNIRS. *Dev. Sci.* *21*, e12595.
15. Meltzoff, A.N., Ramírez, R.R., Saby, J.N., Larson, E., Taulu, S., and Marshall, P.J. (2018). Infant brain responses to felt and observed touch of hands and feet: an MEG study. *Dev. Sci.* *21*, e12651.
16. Lisboa, I.C., Queirós, S., Miguel, H., Sampaio, A., Santos, J.A., and Pereira, A.F. (2020). Infants' cortical processing of biological motion configuration - a fNIRS study. *Infant Behav. Dev.* *60*, 101450.
17. Deen, B., Richardson, H., Dilks, D.D., Takahashi, A., Keil, B., Wald, L.L., Kanwisher, N., and Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nat. Commun.* *8*, 13995.
18. Livingstone, M.S., Vincent, J.L., Arcaro, M.J., Srihasam, K., Schade, P.F., and Savage, T. (2017). Development of the macaque face-patch system. *Nat. Commun.* *8*, 14897.
19. Powell, L.J., Kosakowski, H.L., and Saxe, R. (2018). Social origins of cortical face areas. *Trends Cogn. Sci.* *22*, 752–763.
20. Arcaro, M.J., and Livingstone, M.S. (2021). On the relationship between maps and domains in inferotemporal cortex. *Nat. Rev. Neurosci.* *22*, 573–583.
21. Arcaro, M.J., Schade, P.F., and Livingstone, M.S. (2019). Universal mechanisms and the development of the face network: what you see is what you get. *Annu. Rev. Vis. Sci.* *5*, 341–372.
22. Arcaro, M.J., and Livingstone, M.S. (2017). A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife* *6*, 1–24.
23. Ellis, C.T., Skalaban, L.J., Yates, T.S., Bejjanki, V.R., Córdova, N.I., and Turk-Browne, N.B. (2020). Re-imagining fMRI for awake behaving infants. *Nat. Commun.* *11*, 4523.
24. Ghotra, A., Kosakowski, H.L., Takahashi, A., Etzel, R., May, M.W., Scholz, A., Jansen, A., Wald, L.L., Kanwisher, N., Saxe, R., and Keil, B. (2021). A size-adaptive 32-channel array coil for awake infant neuroimaging at 3 Tesla MRI. *Magn. Reson. Med.* *86*, 1773–1785.
25. Julian, J.B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* *60*, 2357–2364.
26. Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl. Acad. Sci. USA* *104*, 20600–20605.

27. Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* *1*, 42–45.
28. Hasson, U., Levy, I., Behrmann, M., Hendler, T., and Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* *34*, 479–490.
29. Nasr, S., Echavarria, C.E., and Tootell, R.B.H. (2014). Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *J. Neurosci.* *34*, 6721–6735.
30. Yue, X., Pourladian, I.S., Tootell, R.B.H., and Ungerleider, L.G. (2014). Curvature-processing network in macaque visual cortex. *Proc. Natl. Acad. Sci. USA* *111*, E3467–E3475.
31. Yue, X., Robert, S., and Ungerleider, L.G. (2020). Curvature processing in human visual cortical areas. *Neuroimage* *222*, 117295.
32. Livingstone, M.S., Arcaro, M.J., and Schade, P.F. (2019). Cortex is cortex: ubiquitous principles drive face-domain development. *Trends Cogn. Sci.* *23*, 3–4.
33. Pitcher, D., Walsh, V., and Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Exp. Brain Res.* *209*, 481–493.
34. Kamps, F.S., Julian, J.B., Kubilius, J., Kanwisher, N., and Dilks, D.D. (2016). The occipital place area represents the local elements of scenes. *Neuroimage* *132*, 417–424.
35. Bratch, A., Engel, S., Burton, P., and Kersten, D. (2018). The fusiform body area represents spatial relationships between pairs of body parts. *J. Vis.* *18*, 408.
36. Van Grootel, T.J., Meeson, A., Munk, M.H.J., Kourtzi, Z., Movshon, J.A., Logothetis, N.K., and Kiorpes, L. (2017). Development of visual cortical function in infant macaques: a BOLD fMRI study. *PLoS ONE* *12*, e0187942.
37. Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A.L., McGettigan, C., and Garrido, L. (2021). FFA and OFA encode distinct types of face identity information. *J. Neurosci.* *41*, 1952–1969.
38. Henriksson, L., Mur, M., and Kriegeskorte, N. (2015). Faciotopy—a face-feature map with face-like topology in the human occipital face area. *Cortex* *72*, 156–167.
39. Henriksson, L., Mur, M., and Kriegeskorte, N. (2019). Rapid invariant encoding of scene layout in human OPA. *Neuron* *103*, 161–171.e3.
40. Taylor, J.C., Wiggett, A.J., and Downing, P.E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J. Neurophysiol.* *98*, 1626–1633.
41. Brandman, T., and Yovel, G. (2016). Bodies are represented as wholes rather than their sum of parts in the occipital-temporal cortex. *Cereb. Cortex* *26*, 530–543.
42. Pitcher, D., Ianni, G., and Ungerleider, L.G. (2019). A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Sci. Rep.* *9*, 8242.
43. Pitcher, D., and Ungerleider, L.G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends Cogn. Sci.* *25*, 100–110.
44. Ewbank, M.P., Lawson, R.P., Henson, R.N., Rowe, J.B., Passamonti, L., and Calder, A.J. (2011). Changes in “top-down” connectivity underlie repetition suppression in the ventral visual pathway. *J. Neurosci.* *31*, 5635–5642.
45. Cabral, L., Zubiaurre, L., Wild, C., Linke, A., and Cusack, R. (2019). Category-selective visual regions have distinctive signatures of connectivity in early infancy. *bioRxiv*. <https://doi.org/10.1101/675421>.
46. Saygin, Z.M., Osher, D.E., Norton, E.S., Youssoufian, D.A., Beach, S.D., Feather, J., Gaab, N., Gabrieli, J.D., and Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* *19*, 1250–1255.
47. Li, J., Osher, D.E., Hansen, H.A., and Saygin, Z.M. (2020). Innate connectivity patterns drive the development of the visual word form area. *Sci. Rep.* *10*, 18039.
48. Murty, N.A.R., Teng, S., Beeler, D., Mynick, A., Oliva, A., and Kanwisher, N.G. (2020). Visual experience is not necessary for the development of face selectivity in the lateral fusiform gyrus. *bioRxiv*. <https://doi.org/10.1101/2020.02.25.964890>.
49. van den Hurk, J., Van Baelen, M., and Op de Beeck, H.P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proc. Natl. Acad. Sci. USA* *114*, E4501–E4510.
50. Dobs, K., Martinez, J., Kell, A.J.E., and Kanwisher, N. (2021). Brain-like functional specialization emerges spontaneously in deep neural networks. *bioRxiv*. <https://doi.org/10.1101/2021.07.05.451192>.
51. Keil, B., Alagappan, V., Mareyam, A., McNab, J.A., Fujimoto, K., Tountcheva, V., Triantafyllou, C., Dilks, D.D., Kanwisher, N., Lin, W., et al. (2011). Size-optimized 32-channel brain arrays for 3 T pediatric imaging. *Magn. Reson. Med.* *66*, 1777–1787.
52. Kouider, S., Stahlhut, C., Gelskov, S.V., Barbosa, L.S., Dutat, M., de Gardelle, V., Christophe, A., Dehaene, S., and Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science* *340*, 376–380.
53. Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., and Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* *56*, 2356–2363.
54. Zapp, J., Schmitter, S., and Schad, L.R. (2012). Sinusoidal echo-planar imaging with parallel acquisition technique for reduced acoustic noise in auditory fMRI. *J. Magn. Reson. Imaging* *36*, 581–588.
55. Dufford, A.J., Hahn, C.A., Peterson, H., Gini, S., Mehta, S., Alfano, A., and Scheinost, D. (Un)common space in infant neuroimaging studies: a systematic review of infant templated. *bioRxiv*, 10.1101/2021.09.08.459462.
56. Arichi, T., Fagiolo, G., Varela, M., Melendez-Calderon, A., Allievi, A., Merchant, N., Tumor, N., Counsell, S.J., Burdet, E., Beckmann, C.F., and Edwards, A.D. (2012). Development of BOLD signal hemodynamic responses in the human brain. *Neuroimage* *63*, 663–673.
57. Kay, K.N., Rokem, A., Winawer, J., Dougherty, R.F., and Wandell, B.A. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front. Neurosci.* *7*, 247.
58. Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., and Tootell, R.B.H. (2011). The “parahippocampal place area” responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol.* *9*, e1000608.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Group random effects analyses	This paper; http://dx.doi.org/10.17605/OSF.IO/JNX5A	http://dx.doi.org/10.17605/OSF.IO/JNX5A
Functional region of interest analyses	This paper; http://dx.doi.org/10.17605/OSF.IO/JNX5A	http://dx.doi.org/10.17605/OSF.IO/JNX5A
Visual statistics of stimuli	This paper; http://dx.doi.org/10.17605/OSF.IO/JNX5A	http://dx.doi.org/10.17605/OSF.IO/JNX5A
Software and algorithms		
All pre-processing software	This paper (http://dx.doi.org/10.17605/OSF.IO/JNX5A), previous work, ¹⁷ FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/), Freesurfer (https://surfer.nmr.mgh.harvard.edu/)	http://dx.doi.org/10.17605/OSF.IO/JNX5A
All data analysis software	This paper (http://dx.doi.org/10.17605/OSF.IO/JNX5A), FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/), Freesurfer (https://surfer.nmr.mgh.harvard.edu/)	http://dx.doi.org/10.17605/OSF.IO/JNX5A
Visual statistics of stimuli	This paper (http://dx.doi.org/10.17605/OSF.IO/JNX5A); rectilinearity, ²⁹ curvilinearity ^{30,31}	http://dx.doi.org/10.17605/OSF.IO/JNX5A
Stimuli presentation	This paper (http://dx.doi.org/10.17605/OSF.IO/JNX5A) and previous work ¹⁷	http://dx.doi.org/10.17605/OSF.IO/JNX5A
MATLAB 2020a/2020b	MathWorks	https://www.mathworks.com/products/matlab.html
Freesurfer 6.0.0	Massachusetts General Hospital	https://surfer.nmr.mgh.harvard.edu
FSL 5.0.9	FMRIB Software Library	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Heather L. Kosakowski (heatherkosakowski@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- De-identified results from the functional region of interest analyses have been deposited at osf.io and are publicly available as of the date of publication. Accession numbers are listed in the [Key resources table](#).
- Group maps for the three published contrasts have been deposited at osf.io and are publicly available as of the date of publication. Accession numbers are listed in the [Key resources table](#).
- All original code has been deposited at osf.io and is publicly available as of the date of publication. DOIs are listed in the [Key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Infants were recruited from the Boston metro area through word-of-mouth, fliers, and social media. Parents of participants were provided parking or, when travel was a constraint for participation, reimbursed travel expenses. Participants received a \$50 Amazon gift card for each visit and, whenever possible, printed images of their brain. We recruited 87 infants (2.1–11.9 months; mean age = 5.0 months; 48 female) that participated in a total of 162 visits (the scanner was booked for two hours for each visit). Forty-two of these infants (2.1–9.7 months, mean = 5.5 months, 21 female) were scanned on the *Coil 2021*²⁴ and we recovered usable data (see [Data Selection \(subrun creation\)](#)) from 26 infants (2.1–9.7 months, 11 female). The other 47 of these infants (2.5–11.9 months, mean = 4.5 months, 26 female) were scanned on the *Coil 2011*⁵¹ and we recovered usable data from 26 infants (2.5–8.7 months,

mean = 4.8 months, 12 female). (For full subject, data, and inclusion details, see [Table S3](#)). We were unable to collect data from 15 subjects (included in the full count) due to technical error or lack of subject compliance. One subject (included in the full count) was excluded due to experimenter error. Infants were scanned with *Coil 2011* until *Coil 2021* was available. We stopped all data collection due to the COVID-19 pandemic. Ethical approval for this research was obtained through the Internal Review Board at MIT.

METHOD DETAILS

Stimuli

Infants watched blocks of videos and still images from five categories (faces, objects, bodies, scenes, curvy baseline; [Figure 1B](#)). There are four important ways this paradigm is different than what was used previously.¹⁷ First, all infants saw videos from all four conditions. Second, we selected a subset of videos that would make each block more visually and categorically homogeneous within blocks and heterogeneous across blocks (see below for a description). Third, each block included continuous videos that were 2.7 s and interleaved with still images from the same category (but not drawn from the videos) presented for 300 ms. We added interleaved 300ms still images with the hopes of 1) increasing the BOLD response by mitigating effects of adaptation and 2) increasing infant attention. One previous study indicated that 300 ms still image presentations led to the greatest neural response using EEG,⁵² thus we chose a 300 ms still image presentation for our stimuli. We additionally wanted to maximize looking duration to each video so we shortened the 3 s videos from previous research^{17,53} to 2.7 s. Finally, the baseline condition in previous research¹⁷ was scrambled blocks of scenes whereas in this study the baseline consists of curvy abstract scenes. Scrambled objects and scenes are a good control for a variety of low-level features of the object and scene blocks but not the face blocks. Thus, we chose to use curvy abstract scenes rather than scrambled objects or scrambled scenes to control for curvilinear visual features (see below for visual features of stimuli). As such, if the face condition is substantially greater than baseline, we can have some confidence that the response to faces is not merely due to curvilinear features.

All blocks were 18 s and included 6 videos and 6 images. Face videos featured a single face of a child on a black background. Body stimuli focused on hands and feet of children wearing shorts, t-shirts, jeans, or dresses, with or without socks and shoes. Object blocks displayed toys (e.g., magnetic toys, plastic cars and planes, balls) on a black background, and scene blocks mountain or pastoral scenes. Baseline blocks were also 18 s and consisted of six 2.7 s videos and six still images that featured abstract color scenes such as liquid bubbles or tie-dyed patterns. Video and image order were randomized within blocks and block order was pseudo-random by category. There were 3 face blocks (frontal, averted, expressive), 2 body blocks, 3 object blocks, 3 scene blocks, and 1 block depicting hand-object interactions (not included in analysis).

The first 31 infants of the *Coil 2011* dataset (8 of whom were included in the *Coil 2011* fROI analysis) viewed face, body, object, scene, and abstract curvy videos. Video blocks were 18 s long, each video was 3 s long, and was no still image presentation. The videos played continuously for the full time the infant was in the scanner, awake, and attending. Additionally, instead of 12 condition blocks, there were 10 condition blocks – 3 face blocks, 3 scene blocks, 2 body blocks, and 2 object blocks – and the blocks played in a random order. Though the contents of each block was slightly different than the main paradigm, the videos were from the same stimulus set.⁵³

A run began as soon as the infant was in the scanner and videos played continuously for as long as the infant was content, paying attention, and awake. All stimuli available on OSF (osf.io).

Data Collection

Infants were swaddled if possible. A parent or researcher went into the scanner with the infant while a second adult “scanner buddy” stood outside the bore of the scanner. The scanner buddy monitored infant attention and the TRs in which the infant was not looking at the stimuli were recorded for exclusion during preprocessing. Infants heard lullabies (<https://store.jammyjams.net/products/pop-goes-lullaby-10>) played through custom infant headphones for the duration of the scan.

Custom Head Coils

Forty-two infants were scanned in a newly designed 32-channel coil (*Coil 2021*) designed to comfortably cradle infants and wore custom infant MR-safe headphones²⁴ ([Figure 1A](#)). Infant headphones attenuated scanner noises (attenuation statistics reported in [Ghotra et al.²⁴](#)) and allowed infants to listen to music at a comfortable volume for the duration of the scan. An adjustable coil design increased infant comfort, accommodated headphones, and suited a variety of infant head sizes ([Figure 1A](#)). *Coil 2021* and infant headphones were designed for a Siemens Prisma 3T scanner and enabled the use of an EPI with standard trajectory with 44 near-axial slices (repetition time, TR = 3 s, echo time, TE = 30ms, flip angle = 90°, field of view, FOV = 160 mm, matrix = 80x80, slice thickness = 2 mm, slice gap = 0 mm). A few subjects were scanned on a different EPI with standard trajectory with near-axial slices (repetition time, TR = 3 s, echo time, TE = 30ms, flip angle = 90°, field of view, FOV = 208 mm, matrix = 104x104, slice thickness = 2 mm, slice gap = 0 mm). Four of these subjects were included in the fROI analysis.

Data from 47 infants were collected with *Coil 2011* with the same data collection methods reported by [Deen et al.¹⁷](#) We used a custom 32-channel infant coil designed for a Siemens Trio 3T scanner⁵¹ (*Coil 2011*) and a quiet EPI with sinusoidal trajectory⁵⁴ with 22 near-axial slices (repetition time, TR = 3 s, echo time, TE = 43 ms, flip angle = 90°, field of view, FOV = 192 mm, matrix = 64x64, slice thickness = 3 mm, slice gap = 0.6 mm). The sinusoidal acquisition sequence caused substantial distortions in the functional images ([Figure S3B](#)).

Data Selection (subrun creation)

To be included in the analysis, data had to meet the criteria for low head motion that was reported by Deen et al.¹⁷ Data were cleaved between consecutive time points having more than 2 degrees or 2 mm of motion, creating “subruns,” which contained at least 24 consecutive low-motion volumes. All volumes included in a subrun were extracted from the original “run” data (a “run” began when the subject went into the scanner and ended when the subject fell asleep or was fussy) and combined to create a new Nifti file for each subrun. Paradigm files were similarly updated for each subrun. Volumes with greater than 0.5 degrees or mm of motion between volumes were scrubbed. We used highly conservative motion thresholds²³ that were identical to the thresholds we used previously,¹⁷ to ensure inclusion of only the highest quality data in our analysis. If a subject had more than one dataset (data collected within a 30-day window were analyzed as a single dataset), we included only the dataset with more volumes that met our criteria for inclusion. One subject had two datasets but the dataset with more volumes was missing a large portion the temporal lobe due to motion. For this subject we used the dataset with fewer volumes but an intact temporal lobe. This procedure resulted in 514.3 minutes of useable data (mean = 5.9 minutes, s.d. = 11.3) from the *Coil 2021* dataset and 566.00 minutes of useable data (mean = 6.5 minutes, s.d. = 12.6) from the *Coil 2011* dataset.

To be included in the group random effects analysis, subjects had to have at least 96 volumes that met the above motion threshold criteria. We only included subjects from each dataset that were collected with the same acquisition sequence. This resulted in 23 subjects from the *Coil 2021* dataset (2.1–9.7 months, mean = 5.7 months) and 29 subjects from the *Coil 2011* dataset (2.5–8.8 months, mean = 4.8 months) met these criteria.

To be included in the fROI analysis, subjects had to have at least two subruns with at least 96 volumes in each subrun (one to choose voxels showing the relevant contrast, and the other to independently extract response magnitudes from the selected voxels). This resulted in a final fROI dataset of 20 datasets (3.0–9.7 months; mean = 5.8 months) for the *Coil 2021* dataset (16 with the standard acquisition sequence and 3 with a different acquisition sequence and 1 with both sequences) and 19 datasets (18 unique subjects) for the *Coil 2011* dataset (2.5–8.8 months, mean = 4.8 months). Due to the variable amount of data in each subrun for each subject and the impact this could have on reliable parameter estimates from the GLM, we first combined or split subruns to approximately equate the amount of data across subruns within subjects. For example, if a subject had three subruns and the first was 35 volumes, the second was 57 volumes and third was 220 volumes, then we concatenated the first two subruns to create one subrun and we split the third subrun into 2 resulting in a total of three subruns with approximately 100 volumes per subrun.

Preprocessing

Each subrun was processed individually. First, an individual functional image was extracted from the middle of the subrun to be used for registering the subruns to one another for further analysis. Then, each subrun was motion corrected using FSL MCFLIRT. If more than 3 consecutive images had more than 0.5 mm or 0.5 degrees of motion, there had to be at least 7 consecutive low-motion volumes following the last high-motion volume in order for those volumes to be included in the analysis. Additionally, each subrun had to have at least 24 volumes after accounting for motion and sleep TRs. Functional data were skull-stripped (FSL BET2), intensity normalized, and spatially smoothed with a 3mm FWHM Gaussian kernel (FSL SUSAN).

Data registration

Due to the substantial distortion in these data and the lack of an anatomical image from many subjects, we registered all functional data to a representative functional image collected with the same coil and acquisition parameters (referred to as the functional template image). As there is no standard approach to the use of common space in infant MRI research,⁵⁵ we elected to use a template image collected with the same acquisition parameters as each subject. For *Coil 2011* data, the functional template image was the same image used by Deen and colleagues.¹⁷ For *Coil 2021* data, we selected a representative functional image from a representative subject for whom we also had a high-quality anatomical image. All subruns were registered within subjects to a target image from that subject. Then each subject target image was registered to a template functional image collected using the same coil and acquisition parameters. First, the middle image of each subrun was extracted and used as an example image for registration. If the middle image was corrupted by motion or distortion, a better image was selected to be the example image. The example image from the middle subrun of the first visit was used as the target image and all other subruns from that subject were registered to that subject’s target image. The target image for each subject was registered to a template image collected using the same acquisition parameters. Subrun and target image registrations were concatenated so that each subrun was individually registered to template space. We attempted to register each image using a rigid, an affine, and a partial affine registration with FSL FLIRT. The best image was selected by eye from the three registration options and manually tuned with the Freesurfer GUI for the best possible data alignment. To display group results, images were transformed to the anatomical space of the template image.

Parcels / Search Spaces

Group constrained parcels were acquired from previous adult research to localize areas of selectivity.^{25,26} Face-selective parcels included OFA, FFA, and ATL. Scene-selective parcels included OPA, PPA, and RSC. Body-selective parcels included EBA and FBA. After transforming the FFA parcel to individual subject space, we noticed that the peak face activation in the fusiform gyrus often fell just anterior to the FFA parcel. We attributed this difference to difficulties in obtaining reliable registrations with infant functional data and used the `-dillm` function in `fslmaths` to uniformly increase the size of all parcels. As this was an unplanned change in our analysis, we also report fROI results from the original parcels (Table S1) and the dilated parcels (Figures 3, 4, and 5; Table S1).

The large ventral and lateral parcels are the same ones we used in Deen et al.¹⁷ and results are reported in [Figure S3C](#) and [Table S2](#). All parcels will be available on OSF upon publication.

QUANTIFICATION AND STATISTICAL ANALYSIS

Subject-level Beta and Contrast Maps

Functional data were analyzed with a whole brain voxel-wise general linear model (GLM) using custom MATLAB scripts. The GLM included 4 condition regressors (faces, bodies, objects, and scenes), 6 motion regressors, a linear trend regressor, and 5 PCA noise regressors. Condition regressors were defined as a boxcar function for the duration of the stimulus presentation (18 s blocks). Infant inattention or sleep was accounted for using a single impulse nuisance (“sleep”) regressor. The sleep regressor was defined as a boxcar function with a 1 for each TR the infant was not looking at the stimuli, and the corresponding TR was set to 0 for all condition regressors. Boxcar condition and sleep regressors were convolved with an infant hemodynamic response function (HRF) that is characterized by a longer time to peak and deeper undershoot compared to the standard adult HRF.⁵⁶ PCA noise regressors were computed using a method similar to GLMDenoise,⁵⁷ defined in Deen et al.¹⁷ Data and regressors were demeaned for each subrun. Next, demeaned data and regressors were concatenated across subruns, run regressors were added to account for differences between runs, and beta values were computed for each condition in a whole-brain voxel-wise GLM. Three subject-level contrast maps were computed as the difference between the condition of interest beta (i.e., face beta, body beta, or scene beta) and the object beta for each voxel using in-house MATLAB code. Results from subject-level contrast maps are reported in [Figure 2](#).

Group Random Effects Analyses

To test whether there was systematic overlap between areas of activation, we conducted group random effects analyses for eligible data. For visualization and reporting purposes subject-level voxelwise statistical maps were transformed to coil-specific functional template space (see data registration method above) and group random effects analyses were performed using `Freesurfer mri_concat` and `Freesurfer mri_glmfit`. Each group map was then transformed to coilspecific anatomical space. We did not combine data across coils due to difficulty registering the two different image types. Results from group random effects analyses are reported in [Figures 2](#) and [S1](#).

Functional Region of Interest

To constrain search areas for voxel selection, we used anatomically defined parcels (see parcels/search spaces) transformed to subject native space. We used an iterative leave-one-subrun-out procedure such that data were concatenated across all subruns except one prior to the whole-brain voxel-wise GLM and contrasts were computed (described above). The top 5% of voxels with the greatest difference between the category of interest (i.e., faces, bodies, or scenes) and objects within an anatomical constraint parcel were selected as the fROI for that subject, and the parameter estimates were extracted from a GLM on the left-out subrun. Beta values were averaged across folds within a participant and weighted betas were averaged across participants. The use of top 5% and weighted betas was an analytic decision based on previous research.¹⁷ The strength of selecting the top 5% of voxels is that we equally sample all subjects for a representation of the overall pattern of response across subjects. Thus, it is possible to find a response that is unreliable and even has the reverse preference in left out data (see OPA in [Figure 4B](#) as an example – voxels were selected that had a greater response to scenes compared to objects but in independent data, the response to objects was greater than the response to scenes, indicating that across infants, there was not a reliable scene response in OPA.)

Category-Selectivity

To determine whether a region’s response was category-selective, we fit the beta-values using a linear mixed effects model. In each region, we had an *a priori* hypothesis about which of the four conditions would elicit the largest response. So, in each model, we dummy-coded the other three control conditions, to test the hypotheses that the response to each control condition was lower than to the predicted preferred condition. For example, for regions predicted to be face-selective, we fit a model in MATLAB with the expression:

$$\text{fitlme}(\text{data} \sim f1 + f2 + f3 + \text{age} + \text{sex} + \text{motion} + (1|\text{subject}))$$

where the three dummy-coded condition regressors are $f1$ (bodies), $f2$ (objects) and $f3$ (scenes). Fixed effects parameters of no interest were age, motion, and sex. Motion was the fraction of scrubbed volumes. Subject was coded as a random intercept for all models. In the *Coil 2011* analyses, we also included a fixed effect paradigm parameter (“para”) because two slightly different versions of the experimental paradigm were used during data collection.

The response in a parcel was deemed selective if the fixed effect coefficient for each of the three control conditions was significantly negative, using a t test. Because predictions are unidirectional, reported p values are one-tailed. For example, a face parcel was only deemed face-selective if the parameter estimates for the body, object, and scene regressors were significantly negative. Results for the fROI analyses are reported in [Figures 3, 4, 5](#), and [S3](#) and [Tables S1](#) and [S2](#).

Models of stimulus visual statistics

Each frame was extracted from each video. Frames and still images were converted to grayscale and normalized. To obtain spatial frequency information, we computed a Fourier transform on each image. Low- and high-spatial frequency were extracted from each frame of each stimulus using the methods and cut-offs described by Rajimehr et al.⁵⁸ and averaged across all images within a category. Each category value was then normalized relative to the baseline value. Nasr and colleagues²⁹ found that scene and object stimuli are characterized by the presence of angles and that the presence of straight lines and right angles drive activation in the PPA. Similarly, Yue and colleagues^{30,31} found that face stimuli are characterized by the presence of curves and activate areas along the fusiform gyrus. Thus, we also extracted curvilinear information using code provided by Nasr et al.²⁹ and methods described by Yue et al.^{30,31} Briefly, each normalized grayscale image from a block (e.g., each face block was computed separately) was reduced to 140 × 210 pixels and averaged together. For rectilinear information, angled Gabor filters (90° and 180°) with four different spatial frequencies (1, 2, 4, and 8) were applied to each pixel to assess the amount of angular content. Rectilinear values were averaged across pixels and blocks. Curvilinear values were computed in a similar manner except each angled Gabor filter (30°, 60°, 90°, 120°, 150°, and 180°) had five different curve depths. Curvilinear values for each frame from each block were averaged together. All visual statistics results for this analysis are reported in [Figure S2](#).