

Representational similarity precedes category selectivity in the developing ventral visual pathway



Michael A. Cohen^{a,*}, Daniel D. Dilks^{b,1}, Kami Koldewyn^c, Sarah Weigelt^d, Jenelle Feather^e, Alexander J.E. Kell^e, Boris Keil^f, Bruce Fischl^{g,h}, Lilla Zölleiⁱ, Lawrence Wald^{g,h}, Rebecca Saxe^e, Nancy Kanwisher^j

^a Department of Psychology and Program in Neuroscience, Amherst College, USA

^b Department of Psychology, Emory University, USA

^c School of Psychology, Bangor University, UK

^d Faculty of Rehabilitation Sciences, TU Dortmund University, Germany

^e Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

^f Department of Life Science Engineering, Institute of Medical Physics and Radiation Protection, USA

^g Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School, Massachusetts General Hospital, USA

^h Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA

ⁱ Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, USA

^j McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

ABSTRACT

Many studies have investigated the development of face-, scene-, and body-selective regions in the ventral visual pathway. This work has primarily focused on comparing the size and univariate selectivity of these neural regions in children versus adults. In contrast, very few studies have investigated the developmental trajectory of more distributed activation patterns within and across neural regions. Here, we scanned both children (ages 5–7) and adults to test the hypothesis that distributed representational patterns arise before category selectivity (for faces, bodies, or scenes) in the ventral pathway. Consistent with this hypothesis, we found mature representational patterns in several ventral pathway regions (e.g., FFA, PPA, etc.), even in children who showed no hint of univariate selectivity. These results suggest that representational patterns emerge first in each region, perhaps forming a scaffold upon which univariate category selectivity can subsequently develop. More generally, our findings demonstrate an important dissociation between category selectivity and distributed response patterns, and raise questions about the relative roles of each in development and adult cognition.

1. Introduction

The functional organization of the ventral visual pathway is strikingly similar across people (Kanwisher, 2010), raising the obvious question of how this highly systematic structure arises in development (Grill-Spector et al., 2008). To explore this question, researchers have primarily focused on the developmental trajectory of the size and selectivity of category-selective regions including the fusiform face area (FFA; Passarotti et al., 2003; Golarai et al., 2007, 2010, 2015; Scherf et al., 2007, 2011; Peelen et al., 2009; Pelphrey et al., 2009; Cantlon et al., 2011; Natu et al., 2016; Nordt et al., 2017), parahippocampal place area (PPA; Golarai et al., 2007, 2010; Scherf et al., 2007, 2011; Pelphrey et al., 2009; Chai et al., 2010; Meissner et al., 2019), and extrastriate body area (EBA; Peelen et al., 2009; Pelphrey et al., 2009). While a great deal has been learned about the development of the ventral pathway from these

studies, most prior studies rely exclusively on univariate neural measures and do not take advantage of more distributed, multivariate measures. By contrast, a growing literature in adults has shown that multivariate analyses can provide a finer-grained characterization of neural representations within and across cortical regions (Haxby et al., 2014). In particular, representational similarity analysis offers a window into the representations contained within a cortical region through the use of representational dissimilarity matrices (RSMs), which reveal the similarities of the response patterns between all possible stimulus pairs (Kriegeskorte and Kievit, 2013). To the extent that this method reveals neural representations, it should be an important tool for characterizing cortical development.

At this time, only a few studies have examined the development of distributed activation patterns across the ventral pathway using representational similarity analysis. In one study, Deen et al. (2017) found that

* Corresponding author. Department of Psychology and Program in Neuroscience, Amherst College, USA.

E-mail address: michaelthecohen@gmail.com (M.A. Cohen).

¹ These authors contributed equally to this work.

the relationship for large-scale patterns for different object categories across ventral visual cortex were very different between adults and infants. Meanwhile, in another set of studies, Golarai et al. (2010, 2015) compared the similarity of activation patterns between older children and adults and found those patterns to be strikingly similar between the two groups. This particular set of results stands in stark contrast to the numerous studies that have shown that category-selective regions in this swath of cortex are still developing both in terms of their size and their selectivity in older children (Golarai et al., 2007; Scherf et al., 2007, 2011; Peelen et al., 2009; Meissner et al., 2019). Together, these results raise an interesting question: do distributed representational structures mature more quickly than category-selective regions? In this study, we sought to answer this question.

At first blush, answering this question seems relatively straightforward. The first step would simply entail finding category-selective regions in children that are significantly smaller than those in adults. The next step would then be to perform representational similarity analysis in those neural regions and show a high correlation between the RSMs for children and those for adults. However, such a conclusion would not be so straightforward and interpreting such a result would actually be surprisingly difficult. Because the univariate and multivariate measures are so different from one another, it is unclear what pattern of results would enable us to confidently conclude that representational similarity precedes category similarity. How much smaller do category-selective regions need to be in children? What if the regions are significantly smaller in children, but are not entirely absent? How strong do the correlations between children and adults need to be to conclude that children's activation patterns are adultlike? What if the correlations are statistically significant, but considerable unexplained variance remains?

With these difficult questions in mind, we developed a different approach to examine this issue with the strongest test we could think of. Specifically, we asked if we could find mature representational similarity structures in children that have no category-selective regions at all (i.e., literally 0 category-selective voxels). We scanned adults and children with functional magnetic resonance imaging (fMRI) while they passively viewed a variety of object categories. For each participant, we first measured the size of several category-selective regions: FFA, PPA, EBA, as well as the occipital face area (OFA), the face selective portions of the superior temporal sulcus (STS), the occipital place area (OPA), and retrosplenial cortex (RSC). For each region we then selected every child with 0 category-selective voxels for the defining contrast of that region, and we also identified every adult with at least 100 category-selective voxels in each region. Thus, we have a set of participants that are maximally different in terms of category-selectivity: the children have no selective voxels, while the adults have a substantial amount. Within these participants, we then asked whether the representational similarity patterns (i.e., the matrix of similarities in the pattern of responses across voxels between each pair of stimulus categories) were correlated

between children and adults in each region. This procedure enabled us to ask whether children who lacked any hint of category selectivity would nonetheless have mature representational similarity patterns.

Overall, our results were unambiguous; even when comparing children with no category selective voxels to adults with significant category-selective regions, we still found strong correlations between the two groups when using representational similarity analysis. Thus, distributed representational structures appear to precede the formation of category selective regions in the developing ventral visual pathway. Taken together, these findings raise new questions about how these univariate structures and multivariate patterns develop, their relationship to each other, and their respective causal roles in development and behavior.

2. Materials and methods

2.1. Subjects

We scanned 38 adults (mean 25.1 years old; standard deviation 4.48 years) and 41 children, ages 5–7 years old (mean 6.6 years old; standard deviation 0.91 years). Excessive amounts of motion in 4 of the children resulted in their data being excluded from all further analyses due to an inability to reconstruct the images. All participants had normal or corrected-to-normal vision and no known neurological or psychiatric conditions or structural brain abnormalities. Adult participants and the parents of children participants provided written, informed consent and all children gave verbal assent to participate in the experiment. The Massachusetts Institute of Technology (MIT) Institutional Review Board approved of all experimental protocols.

2.2. Stimuli

In order to hold the interest of children, we used colored movie clips as stimuli, which showed movies of faces, bodies, scenes, objects, and scrambled objects (Pitcher et al., 2011; see Fig. 1). Movies of faces and bodies were filmed on a black background, and framed close-up so that only the faces or bodies of 7 children were visible as they danced, or played with toys or adults that were out of frame. The scene stimuli were mostly pastoral scenes shot from the window of a car that drove through suburbs. There were also some stimuli that were clips of flying over canyons or walking through tunnels. Moving objects were selected that minimized any suggestion of animacy of the object itself or of a hidden actor pushing the object. Examples of these objects included mobiles, windup toys, toy planes and tractors, balls rolling down sloped inclines, etc. The scrambled objects were created by dividing each object clip into a 15×15 box grid and rearranging the location of each of the resulting frames. Finally, rather than using a stationary fixation point as baseline, we used six uniform color fields that were designed to maintain the interest of children, while approximating a fixation baseline condition by

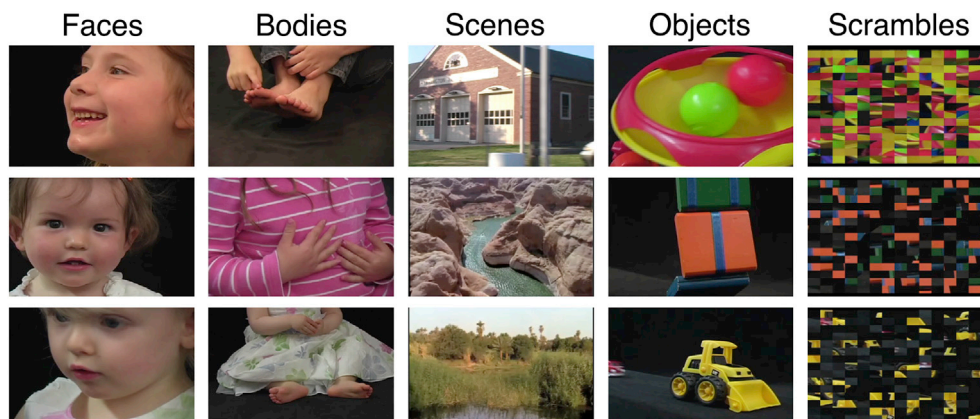


Fig. 1. Sample stimuli. Example frames from movies used as dynamic stimuli.

avoiding any patterned visual input. All stimuli were created using MATLAB and Psychtoolbox (Brainard, 1997; Pelli, 1997) and were presented by a liquid crystal display projector onto a screen in the scanner, which subjects viewed via a mirror attached to the head coil.

2.3. fMRI acquisition

All participants were scanned using functional magnetic resonance imaging (fMRI) at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research, MIT on a Siemens 3T MAGNETOM Trim Trio Scanner (Siemens AG Healthcare). Two weeks before a child's visit, he or she received a CD and illustrated booklet that introduced the experimenters to the child, described the MRI procedure, and included recordings of scanner sounds. Earbuds similar to those the child would wear in the scanner were also included so that he or she could become accustomed to them. Parents were encouraged to review all materials with their child and asked to help him or her practice lying still while listening to the noises of the scanner. Immediately before their scanning session, all children were trained for 15–30 min in a “mock” scanner, designed to simulate the appearance, noise, and confinement of the actual scanner. During these training sessions, children practiced lying still while watching a movie. The movie was turned off by a motion tracking system any time children moved too much in order to teach them how still they had to be to get “good brain pictures.”

For the children, functional images were acquired using a custom made 32-channel phased array head-coil (Keil et al., 2011) optimized to the average head size of 5–7 year olds, and a gradient echo single-shot echo-planar imaging sequence (32 slices, repetition time (TR) = 2 s, echo time (TE) = 30 ms, voxel size = $1 \times 1 \times 1$ mm, and 0.6 mm inter-slice gap). For the adults, functional imaging parameters were identical to the children, with the exception of using a commercially available Siemens 32-channel phased array head-coil, which is ideally suited for an adult head. For all scans, slices encompassed the whole brain aligned to the AC/PC line. Prior to each scan, four “dummy” scans were acquired and discarded to allow longitudinal magnetization to reach equilibrium. High-resolution T1-weighted anatomical images were also acquired for each participant.

2.4. Experimental design

Functional data were acquired over four blocked-design functional runs lasting 234 s each. Each functional run contained three 18-s rest blocks, at the beginning, middle, and end of the run, during which a series of six uniform color fields were presented for 3 s each. Each run contained two sets of five consecutive stimulus blocks (i.e., faces, bodies, scenes, objects, or scrambled objects) sandwiched between the rest blocks, to make two blocks per stimulus category per run. Each block lasted 18 s and contained six 3-s movie clips from each of the five

stimulus categories. The order of stimulus category blocks in each run was palindromic (e.g., fixation, faces, objects, scenes, bodies, scrambled objects, fixation, scrambled objects, bodies, scenes, objects, faces, fixation), and counterbalanced across runs. Participants were asked to passively view the stimuli.

2.5. fMRI data analysis

All fMRI data were analyzed using the Freesurfer software package (Dale et al., 1999; Fischl et al., 1999, 2001) and custom MATLAB scripts. Preprocessing steps included 3-dimensional motion correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), slice scan-time correction, and spatial smoothing (5 mm FWHM kernel). In addition, both children and adult scans were spatially registered using the combined volume and surface-based (CVS) non-linear registration method (Postelnicu et al., 2009) via the T1-weighted MRI images. The surfaces were constructed from a 1 mm isotropic MPRAGE with real-time motion correction using nVas (Van der Kouwe et al., 2008; Tisdall et al., 2012). This allowed for direct comparisons between the two groups since this transformation into a common space normalizes for absolute brain volume. All statistical analyses were based on the general linear model (GLM). GLM analyses all included box-car regressors for each stimulus block, which were convolved with a gamma function to approximate the idealized hemodynamic response. For each experimental protocol, separate GLMs were computed for each participant, yielding regression-weights (i.e., beta maps) for each condition for each subject.

2.6. Defining, characterizing, and measuring category-selective regions of interest

Because category-selective regions of interest (ROIs) are sometimes ambiguous, such as when a subject has two FFAs (Weiner and Grill-Spector, 2012), traditional methods of identifying functionally defined ROIs sometimes require judgment calls about which activation cluster should be taken as the ROI in question, raising the possibility of bias. To minimize the subjectivity inherent in hand picking ROIs, our primary analysis used an algorithmic method for ROI selection. All category-selective ROIs were defined using a Group-Constrained Subject Specific (GSS) Method (Fedorenko et al., 2010). This analysis is based on a previously published parcel atlas that was derived from 42 adult subjects to constrain the definition of numerous ROIs (Julian et al., 2012). These parcels are identified as relatively large swaths of the cortical surface in which most subjects show activation for a particular contrast. Therefore, each of our category-selective ROIs was defined by conjoining contrast maps (e.g., faces vs. objects) with a particular parcel. The particular contrasts used to define our ROIs were faces vs. objects for FFA, OFA, and STS, scenes vs. objects for PPA, OPA, and RSC, and bodies vs. objects for EBA (Fig. 2). In all cases, for our primary analyses, we

Example regions of interest (ROIs)

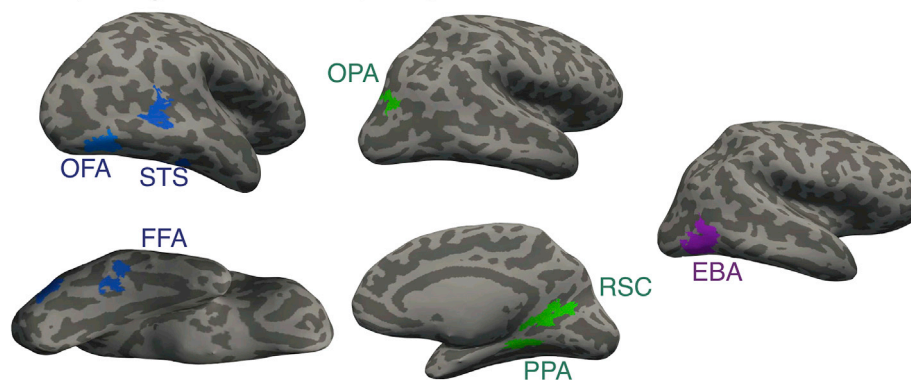


Fig. 2. Example of each of the category-selective regions we identified on representative adult participants.

combined regions across the two hemispheres to form one bilateral region of interest. For all contrasts, we used a statistical threshold of $P < 0.001$ uncorrected.

To define and characterize every ROI, two runs were always used to select the voxels for a particular region, while the remaining two runs were used to obtain an independent measure of the response profile of that region. When quantifying the volume of a region, we would measure the number of voxels that passed our statistical threshold within each subject for the odd runs and the even runs separately and then average them together. For example, in the FFA, if a participant had 500 significant voxels from the two odd runs and 300 significant voxels from the two even runs, we would ultimately say that participant has an average of 400 significant FFA voxels. Meanwhile, to characterize a region, we would use the held out runs (i.e., define on odd runs, measure on even runs; and define on even runs, and measure on odd runs), and the results from those analyses were then averaged together. These steps were taken to ensure that we avoided any issues of statistical non-independence: data used to define a region were never also used to characterize a region (Kriegeskorte et al., 2009; Vul et al., 2009).

2.7. Statistical analysis: representational similarity analysis

To compare the similarity structures between adults and children, we used representational similarity analysis to compute a series of brain/brain correlations focused on a variety of subdivisions within the visual hierarchy. This analysis requires the formation of a representational similarity matrix from neural measures in both adults and children that could then be directly compared to one another (Kriegeskorte and Kievit, 2013). A representational similarity matrix is the set of pairwise similarities (i.e. correlations) between the pattern of response of voxels in a given region to two stimulus classes (e.g., the correlation across voxels between the patterns of response to faces and scenes in FFA). Once these similarity matrices were computed for each child and each adult, they were averaged within each group to create a child group similarity matrix and an adult group similarity matrix. We then measured the correlation between those group-level matrices. The statistical significance of the observed correlations between two correlation matrices, one from adults and the corresponding one from children, was assessed using group-level analyses. That is, the condition labels of the data of each group-level matrix were shuffled, the newly labeled matrices were correlated with one another, and the resulting correlation value was Fisher z-transformed. This procedure was repeated 10,000 times, resulting in a distribution of correlation values. A particular correlation between representational similarity matrices was considered significant if it fell within the top 5% of values in this distribution.

When correlating the representational similarity matrices between adults and children, it is possible that we would see artificially low

correlations simply because of unreliable neural data. To assess this possibility, for every correlation we observed between adults and children, we also computed a reliability-adjusted correlation. The first step of computing these adjusted correlations requires determining the split-half reliability for each particular participant. To compute two similarity matrices, we would first use one set of runs (i.e., runs 1 and 3) to define an ROI and the other runs (i.e., runs 2 and 4) to generate a similarity matrix. We would then switch the runs to generate a second similarity matrix (i.e., define an ROI with runs 2 and 4 and generate the matrix with runs 1 and 3). The result of this process was an odd and even similarity matrix for each participant and ROI. Once these matrices were computed for each participant, they were averaged together to form a group level odd and even similarity matrix for both the adults and the children. Those group level odd and even similarity matrices were then correlated with one another to get an estimate of the reliability of the data for each group of participants. These correlation values were then adjusted using the Spearman-Brown formula to estimate the reliability of the full data set (Spearman, 1910; Brown, 1910). Finally, to adjust the observed correlations as a function of the reliability of the data, we used the correction for attenuation formula: the observed correlation between adults and children from a given neural region divided by the square root of the product of reliability of the data in that region from both adults and children (Nunnally and Bernstein, 1994; Cohen et al., 2017).

2.8. Similarity analysis in category-selective regions and participant selection

Rather than perform our analyses in every participant we scanned, we examined only those children who had zero category-selective voxels for each region and those adults who had at least 100 category-selective voxels for each region. The rationale for these selection criteria is as follows: If we find significant correlations between the representational similarity matrices of children and adults with these two groups, it would provide the strongest evidence that representational similarity precedes category selectivity in the ventral pathway. If, for example, we find strong correlations between children that do not have a single FFA voxel and adults that have at least 100 FFA voxels, then it suggests that distributed representational structures develop before category-selective regions, since there are no category-selective voxels in these particular children.

To select the voxels we want to use for these analyses, we first counted the number of category-selective voxels in each individual adult and only selected those adults with at least 100 voxels (Fig. 3A). Once we identified those adults, we then computed the average number of category-selective voxels within that group (Fig. 3B). In the case of the FFA, we found that our selected adults (N = 32) had on average 886

Creating similarity matrices in adults (FFA as an example)

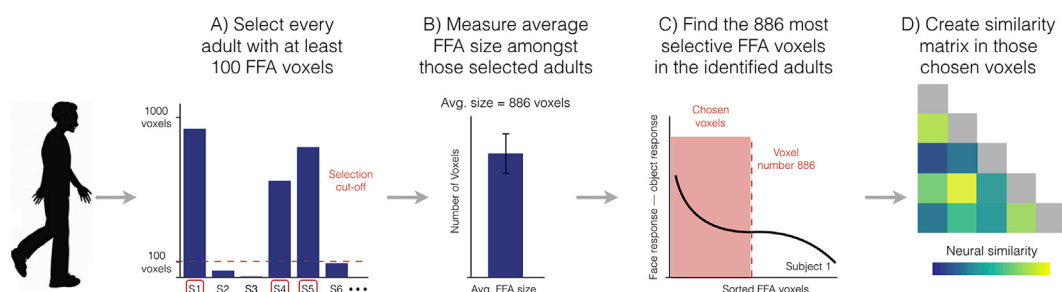


Fig. 3. Visualization of the method used to form representational similarity matrices in adults. In this case, we use the FFA as an example. A) First, we measured the size of the FFA in every adult and only selected adults with at least 100 voxels. B) Then we determined the average size of the FFA across the selected adults (i.e., 886 voxels). C) Next, we selected the top 886 voxels in every adult such that we had the exact same number of voxels in every participant. D) Once those voxels were selected, we created a similarity matrix in each individual participant, which we then averaged together across participants to make one adult group-level matrix.

significant voxels. Then, in order to have the same number of voxels across individuals, we went back to our contrast maps and selected the same number of voxels for each individual. Thus, for the FFA, this meant selecting the 886 most face-selective voxels in each individual adult (Fig. 3C). It should be noted that in some cases this meant selecting some voxels that did not reach statistical significance in some people and excluding some voxels that did reach statistical significance in others. For example, imagine a participant with only 800 significant FFA voxels ($P < 0.001$ uncorrected). To get to 886 voxels, we chose those significant 800 voxels and we also chose the 49 voxels with the lowest P -values even though those values were all $P > 0.001$. Conversely, imagine a participant with 900 significant FFA voxels. To get to 886 voxels with this participant, we would excluded the 51 voxels with the lowest P -values even though those voxels all had P -values that were still $P < 0.001$. It was with these selected voxels that we would create a similarity matrix (Fig. 3D) for each individual participant, which we then averaged together to make a single group-level matrix for the adults.

Once we identified the adults we wanted to examine and determined how many voxels were in each category-selective region for those adults, we then selected our children and the voxels we want to use within those children. Since the goal of our analyses is to examine representational similarity in children with no category-selective voxels, the first step of this process is to identify every child that has zero category-selective voxels in a given region, again with a threshold of $P < 0.0001$ uncorrected (Fig. 4A). We then consulted the number of category-selective voxels we found in a given region amongst the adults that we selected above (Fig. 4B). In the case of the FFA, for example, we found an average of 886 FFA voxels. Similar to the procedure described above, we then selected that exact number of voxels in every child even though none of those voxels were above our statistical threshold. (Fig. 4C). In other words, we selected the 886 most “selective” FFA voxels in children in spite of the fact that none of those voxels are reliably more responsive to faces than objects.

It should be noted that even though we selected voxels in the children that did not significantly respond to faces over objects, it is still possible that once grouped together across participants, we would still find significant activation to the preferred category over objects in a particular region. For example, imagine that every FFA voxel we selected in children had a hypothetical P -value of 0.06 in terms of faces vs. objects. In that case, while no single voxel is significantly responsive to faces, once averaged together within and across participants, it is likely we would find that the group-level response to faces is greater than the group-level response to objects. Such a finding would defeat the purpose of these analyses, as the primary goal is to identify children that have zero category selectivity. Thus, with the selected children, we performed group-level comparisons of the average response to faces/scenes/bodies and the average response to objects. If the overall difference between the

selected category and objects was $P < 0.50$, we removed the participants with the greatest effect size until we reached $P > 0.50$ for the group. This procedure was intended to ensure that we not only had selected individual children with no category-selective voxels, but that the children we did analyze did not even have a trend of a significant preference as a group for the selected category over objects in the different category-selective regions. Once we identified those children and those voxels in every individual child, we then performed our similarity analyses within those voxels for each child and then averaged them together to form a group similarity matrix that we could compare to adults (Fig. 4D).

2.9. Data and code summary statement

All data and code are available upon request and this sharing policy complies with the requirements of the bodies that funded this research.

3. Results

3.1. Size of category-selective region and participant selection

The number of significant voxels we found in each child and adult for each of our seven category-selective-regions is presented below (Fig. 5). For the children, this resulted in the selection of 19 participants for the FFA, 32 for the OFA, 13 for the STS, 14 for the EBA, 11 for the PPA, 29 for the OPA, and 17 for the RSC. For the adults, this resulted in the selection of 32 participants for the FFA, 19 for the OFA, 31 for the STS, 29 for the EBA, 29 for the PPA, 8 for the OPA, and 20 for the RSC.

Overall, this selection process resulted in our identifying a group of adults within each category-selective region with an average total of 886 voxels in the FFA, 805 in the OFA, 1,576 in the STS, 2,387 in the PPA, 561 in the OPA, 2,338 in the RSC, and 3,146 in the EBA (Fig. 6). Naturally, since we selected children with no category-selective voxels, there was an average of zero category-selective voxels in each region.

3.2. Response properties of category-selective regions

Even though we selected children with no category-selective voxels, it is still possible that on average, the response to the selective category (e.g., scenes in PPA or bodies in EBA) will still be significantly higher than to the control category (objects) in the group analysis. Indeed, we found this exact pattern of results in several of our category-selective regions in children: PPA: $t(10) = 2.28$, $P < 0.05$; OPA: $t(31) = 2.62$, $P < 0.05$; RSC: $t(16) = 3.55$, $P < 0.01$. For OFA, STS, and EBA, the selective category (i.e., faces and bodies) were not significantly greater than objects: OFA: $t(30) = 0.36$, $P = 0.72$; STS: $t(12) = 0.87$, $P = 0.40$; EBA: $t(13) = 1.62$, $P = 0.13$. Interestingly, in FFA, we actually found that the response to objects was slightly higher than the response to faces in

Creating similarity matrices in children (FFA as an example)

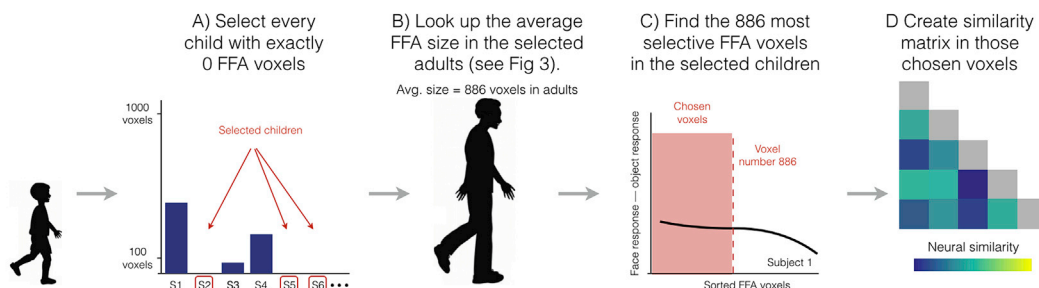


Fig. 4. Visualization of the method used to form representational similarity matrices in children. A) First, we measured the size of the FFA in every child and only selected children with 0 voxels. B) Then we consulted how many FFA voxels we found across our group of selected adults (i.e., 886 voxels). C) Next, we selected the top 886 voxels in every child (even though none of those voxels reached statistical significance) such that we had the exact same number of voxels in every participant. D) Once those voxels were selected, we created a similarity matrix in each individual participant, which we then averaged together across participants to make one child group-level matrix.

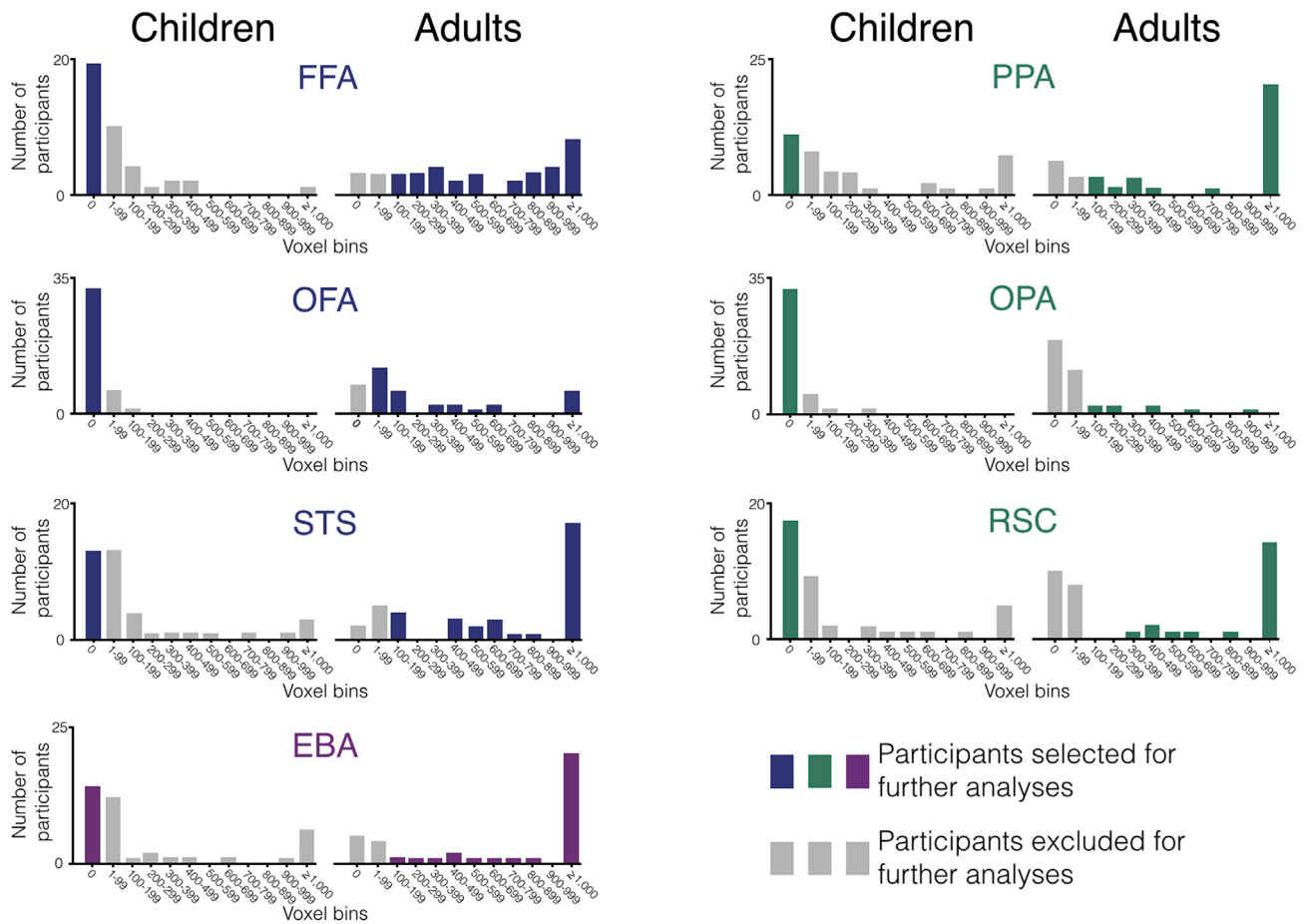


Fig. 5. Summary distribution of the size of the category-selective regions in every participant. On the x-axis of every plot is a series of voxel bins in which we group the number of voxels into discrete bins of 100 voxels. On the y-axis is the number of participants whose category-selective region land within a given bin. Across all plots, the colored bars mark the participants selected for further analyses (i.e., children with no category-selective voxels and adults with at least 100 category-selective voxels), while the grey bars mark participants excluded for further analyses (i.e., children with any category-selective voxels and adults with fewer than 100 category-selective voxels).

the children selected with zero FFA voxels, though this effect was not significant (FFA: $t(18) = 1.26$, $P = 0.22$). For those category-selective regions in which the difference between the preferred category and objects was $P < 0.50$, we identified the children with the largest preferences for the selective category over objects (i.e., faces greater than objects in

OFA, bodies greater than objects in EBA) and removed them one by one until we obtained $P > 0.50$. After removing certain subjects, all category-selective regions showed no preference for the selective category over objects in children: STS: $t(11) = 0.51$, $P = 0.62$; EBA: $t(10) = 0.29$, $P = 0.78$; PPA: $t(6) = 0.49$, $P = 0.64$; OPA: $t(24) = 0.57$, $P = 0.58$; RSC: $t(6) = 0.25$, $P = 0.80$. Meanwhile, in adults, each region showed a strong preference for the selective category relative to objects: FFA: $t(31) = 16.23$, $P < 0.001$; OFA: $t(18) = 12.99$, $P < 0.001$; STS: $t(30) = 15.18$, $P < 0.001$; EBA: $t(28) = 19.04$, $P < 0.001$; PPA: $t(28) = 14.84$, $P < 0.001$; OPA: $t(7) = 7.88$, $P < 0.001$; RSC: $t(19) = 12.86$, $P < 0.001$ (Fig. 7). Ultimately, after this selection process we were left with the following numbers of participants for each region in both groups: FFA: 19 children, 32 adults; OFA: 32 children, 19 adults; STS: 12 children, 31 adults; EBA: 11 children, 29 adults; PPA: 7 children, 29 adults; OPA: 25 children, 8 adults; RSC: 7 children, 20 adults.

Number of category-selective voxels in selected participants
(i.e., adults with at least 100 category-selective voxels and children with 0 category-selective voxels)

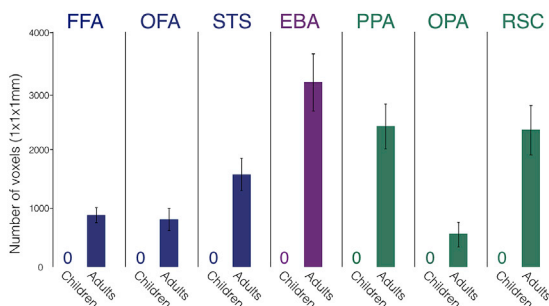


Fig. 6. Size of each category-selective region. The number of voxels in each region was measured for both children and adults and is shown here on the y-axis. Each region was defined using a statistical threshold of $P < 0.001$ uncorrected. There are no bars for the children since we purposefully selected children with no category-selective voxels. The error bars for adults denote the standard error of the mean.

3.3. Representational similarity in category-selective regions

Once we identified a group of children with no discernible category selectivity and a group of adults with strong selectivity (Figs. 6 and 7), we then asked if there is a significant correlation between the representational similarity matrices of these two groups. Again, to perform this analysis in children, we selected the most “selective” voxels for each region even though none of those voxels responded significantly more to the preferred category for that region than to objects. To determine how many voxels to select, we used the size of the adults’ category-selective

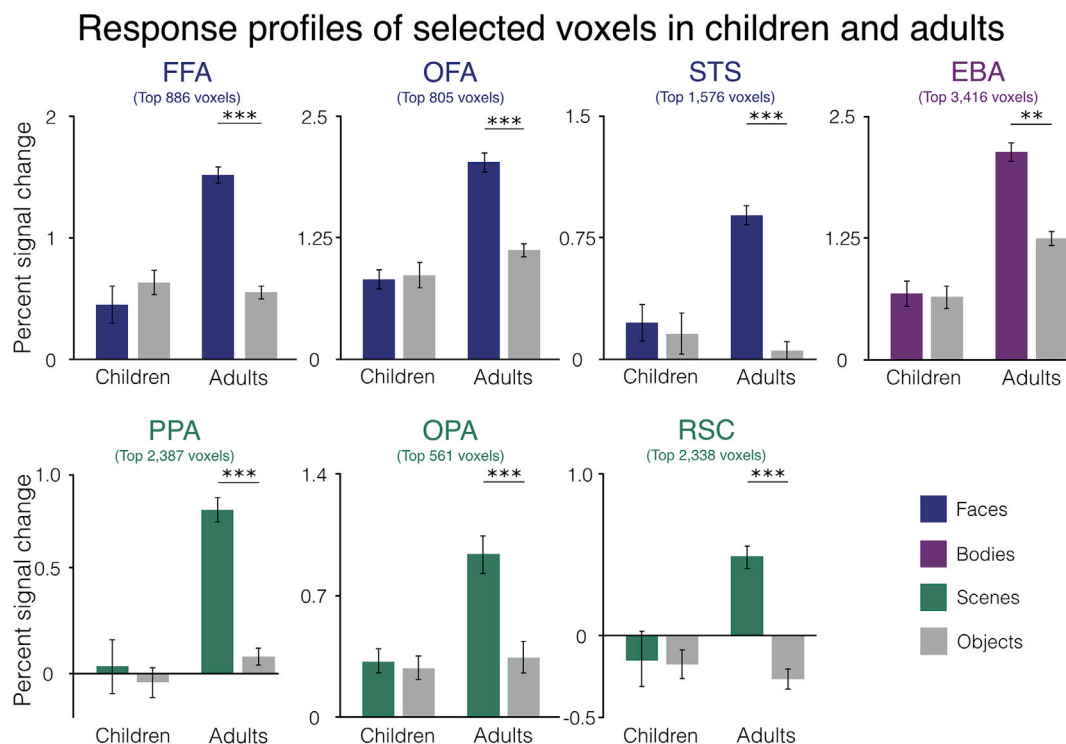


Fig. 7. Univariate responses to selective categories (i.e., faces in FFA, OFA, and STS, bodies in EBA, and scenes in PPA, OPA, and RSC) in the selected voxels of children and adults. These data show the responses in each group after iteratively removing any children until the group analysis across children showed no univariate selectivity for the region-defining contrast. In all cases, the grey bars are the response in those regions to objects, while the colored bars are the response to the preferred category for the region in question, in data independent of those used to define the region. Percent signal change is represented on the y-axis. ** $P < 0.01$, *** $P < 0.001$.

regions as our benchmark (i.e., the top 886 voxels in FFA since on average the adults had 886 FFA voxels). After selecting these voxels in both groups, we created representational similarity matrices by correlating the responses for each pairing of categories across all voxels in a particular neural region (e.g., the correlation across voxels between the patterns of response to faces and scenes in FFA, etc.).

Our results were unambiguous. In every category-selective region, we found strong correlations between children's and adults' matrices: FFA: $r = 0.78$, $P < 0.05$, reliability-adjusted $r = 0.91$; OFA: $r = 0.82$, $P < 0.05$, reliability-adjusted $r = 0.92$; STS: $r = 0.82$, $P < 0.05$, reliability-adjusted $r = 0.95$; EBA: $r = 0.95$, $P < 0.05$, reliability-adjusted $r = 0.99$; PPA: $r = 0.75$, $P < 0.05$, reliability-adjusted $r = 0.97$; OPA: $r = 0.85$, $P < 0.05$, reliability-adjusted $r = 0.89$ (Fig. 8). The one region in which there was a trend of a correlation that did not reach significance was in RSC ($r = 0.54$, $P = 0.08$). However, this relatively low correlation is at least in part due to unreliable data since the reliability-adjusted correlation is higher ($r = 0.65$) and more broadly may stem from RSC being more involved in spatial-memory retrieval than scene perception (Kravitz et al., 2011). Taken together, these data strongly suggest that even in children with no discernible selectivity for the defining contrast for that region, those distributed representational structures are already mature in those regions.

Because we found such strong correlations between adult and child similarity matrices in every category-selective region, a natural question is whether these regions have different representational structures from one another. Are the strong correlations between adults and children across all category-selective regions driven by a common representational structure? Or do the different regions have dissociable structures that vary across the cortex? To address these questions, we calculated the correlations between each of our neural regions within our two groups (i.e., correlate FFA and PPA within children, or EBA and RSC within adults, etc.). In children, nearly two thirds of the neural regions were not significantly correlated with one another: 13 out of 21 possible pairings

(Fig. 9). In adults, meanwhile, almost half of the neural regions were not correlated with one another: 9 out of 21 possible pairings. Thus, while there appears to be some common representational structures between certain neural regions, it does not appear as if our strong correlations between adults and children are driven entirely by our measuring one single structure.

3.4. Does category selectivity ever precede representational similarity?

We have focused so far on the development of representational similarity amongst a group of children who were specifically chosen because they have no discernible selectivity on the defining contrast for the region in question. We have shown that despite their total lack of univariate category selectivity, these children nonetheless show mature representational similarity patterns in the same regions. Is the opposite true? Do children who have no representational similarity with adults have any hint of category selectivity? To answer this question, we first correlated the matrix from each individual child from our original group ($N = 37$) with the group-level adult similarity matrix, for each category-selective region. We then selected every child whose individual correlation with the adults was $r \leq 0.00$ in the region in question (i.e., one group of children were selected for FFA, another for PPA, etc.). Next, we measured the average size of the category-selective regions in the selected group of children using a statistical threshold of $P < 0.001$ uncorrected. Overall, we found that for children that did not show a positively correlated representational similarity pattern in the region in question, the size of the category-selective region was never significantly greater than 0 (FFA: 88 voxels, $t(7) = 1.62$, $P = 0.15$; OFA: 6 voxels, $t(11) = 1.06$, $P = 0.31$; STS: 31 voxels, $t(6) = 1.47$, $P = 0.19$; EBA: 638 voxels, $t(2) = 1.15$, $P = 0.37$; PPA: 42 voxels, $t(2) = 1.0$, $P = 0.42$; OPA: 1 voxel, $t(6) = 1.0$, $P = 0.36$; RSC: 55 voxels, $t(5) = 1.10$, $P = 0.32$). In other words, this analysis revealed that children who have not developed any representational structures that are correlated with adults have also not

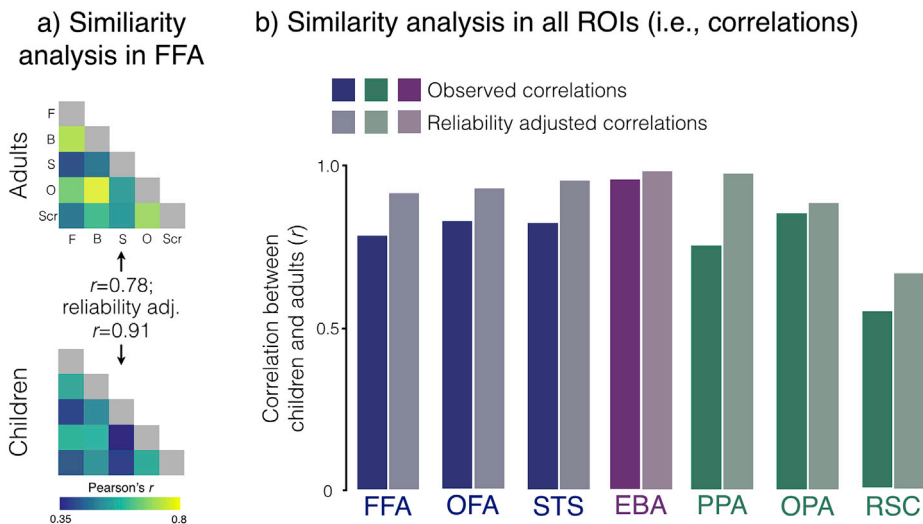


Fig. 8. Representational similarity analysis comparing children and adults. A) Visualization of the similarity matrices based on distributed activation patterns in the FFA voxels in adults (top matrix) and children (bottom matrix). Each cell corresponds to the correlation of the activation patterns within a particular region for two stimulus categories (e.g., the correlation between the pattern of response across voxels to objects and bodies in FFA, etc.). B) Correlations between children and adults in each of our neural regions. The y-axis shows the correlation between the two groups. For each neural region, the saturated bar represents the observed correlation between children and adults while the desaturated bar represents the reliability adjusted correlation (see Methods).

begun to develop category-selective regions (e.g., FFA, PPA, EBA, etc.). Thus, representational similarity patterns are present in children who lack the corresponding univariate selectivity for the region in question, but not vice versa. Of course, these analyses should be interpreted with caution since for several regions, we were only able to identify a small number of children whose similarity matrices were negatively correlated with adults (e.g., PPA). However, the fact that so few children are negatively correlated with adults is itself consistent with our broader claim that distributed activation patterns mature earlier than category-selective regions.

4. Discussion

Here, we report a developmental dissociation between univariate category selectivity and distributed similarity patterns: even in children that have no discernible selectivity for the defining contrast of a given region (e.g., no voxels that respond significantly more to faces than objects in the FFA), the representational similarity patterns in that region are already mature. These findings highlight a key dissociation in the development of the human visual system: rather than developing in synchrony with one another, each region's distributed activation patterns apparently develop before that region's defining univariate selectivity.

These results are consistent with those of Golarai and colleagues (2010; 2015), who showed more mature representational similarity across the ventral visual pathway in 7–11 year old children. However, our findings build on the earlier results by further showing that this more mature representational similarity a) is present at the younger age of 5–7, b) exists within each developing category-selective region, and c) is present even when the defining category selectivity of a region is absent altogether. These results reveal an important dissociation between category selectivity and representational similarity.

Our findings raise numerous questions. First, what is the precise time-course of the development of distributed representational structures in the ventral visual pathway? Deen et al. (2017) showed that representational similarity structures are very different from adults at 6 months of age, so they could mature any time between infancy and the 5–7 year old range studied here. Second, what are the neurobiological mechanisms that construct representational similarity during development? Third, given that representational similarity appears to mature first, does it play a causal role in the development of category selectivity? None of these important questions is answerable with the current data set, but in principle all could be addressed in future work.

A major obstacle in speculating about these questions comes from the fact that we do not understand the relative significance and causal roles of representational similarity and univariate selectivity, or their neurobiological basis, even in adults. Empirically, univariate selectivity in the ventral visual pathway is among the most robust and replicated phenomena in human cognitive neuroscience, and extensive evidence shows that category-selective regions play specific causal roles in behavior (Wada and Yamamoto, 2001; Pitcher et al., 2009; Dilks et al., 2013; Schalk et al., 2017). But computationally, we do not know what goal might be served by category selectivity of specific neural populations, or by the spatial clustering of these neural populations at a sufficient grain to produce category-selective regions detectable with fMRI. Similarly, representational similarity patterns are robust and widely replicated. However, the window they offer into neural/mental representations and their causal role in behavior is unclear. The fact that a given stimulus classification can be performed based on patterns of response across voxels within a particular region of cortex does not guarantee that this information is used (i.e. read out by other brain regions Williams et al., 2007). Although representational similarity is sometimes correlated with behavior (e.g., Cohen et al., 2014, 2015; 2017), evidence from patients and stimulation studies raise questions about its causal role. For example, despite the many studies showing decoding abilities and replicable similarity matrices for non-faces objects in the FFA, intracranial electrical stimulation of the FFA appears to affect only face percepts, not the perception of non-face objects (Parvizi et al., 2012; Schalk et al., 2017),

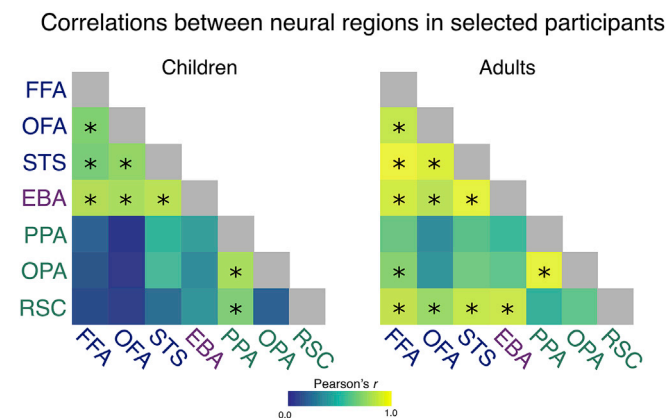


Fig. 9. Representational similarity analysis comparing the correlations between neural regions in both children and adults. In both matrices, each cell corresponds to the correlations between similarity matrices in two given neural regions (e.g., correlation the matrices for the five categories in FFA and OFA, etc.). Asterisks within each cell denote significant correlations between two regions ($P < 0.05$).

suggesting that the pattern information about non-face objects in the FFA may not be causally related to perceptual experience and behavioral performance.

Even if representational similarity is sometimes epiphenomenal in terms of its role in adult perception and cognition, it may nonetheless play an important role in development. But how differences in representational similarity could lead to a later change in the univariate selectivity of a given region is far from clear. One possibility is that early-developing representational similarity in the ventral pathway reflects a “protomap” of cortical organization that serves as a scaffold upon which further development is built (Hasson et al., 2002; Deen et al., 2017; Livingstone et al., 2017). These representational similarities could further reflect retinotopic and featural biases (e.g., for curvature versus rectilinearity) inherited from earlier stages of visual processing, with these biases determining which regions of the ventral pathway will take on which function. Of course, it may be that representational similarity plays a relatively minor role in determining a region's selectivity. Instead, that role may be better accomplished by structural connectivity, which develops very early, varies systematically across functionally different regions (Saygin et al., 2012; Osher et al., 2016), and in at least one case identifies the locus of a functionally distinct region before that region's univariate selectivity arises (Saygin et al., 2012).

The many questions raised by this study about the role of representational similarity in development can be addressed in future work by deriving more detailed similarity matrices from a larger set of stimulus types both within and between categories. Richer matrices might reveal differences between adults and children that were not evident in our study. It will also be important to scan children between the ages of 6 months and five years to learn more about the timeline of the development of representational similarity structures across distributed populations of voxels and neurons. Finally, to understand the role of distributed activation patterns in development, we need to better understand its causal role in adult perception, including the fundamental question of the spatial scale of the neural codes that are read out in behavior.

Acknowledgements

This study was supported by funds from the Ellison Medical Foundation. Thanks to Harris Hoke for assistance with data analysis and Caroline Robertson and Leyla Isik for helpful comments on the project.

References

- Brainard, D.H., 1997. The psychophysics toolbox. *Spatial Vis.* 10, 433–436.
- Brown, W., 1910. Some experimental results in the correlation of mental abilities. *Brit. J. Psychol.* 3, 296–322.
- Cantlon, J.F., Pineda, P., Dehaene, S., Pelphrey, K.A., 2011. Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cerebr. Cortex* 21, 191–199.
- Chai, X.J., Ofen, N., Jacobs, L.F., Gabrieli, J.D.E., 2010. Scene complexity: influence on perception, memory, and development in the medial temporal lobe. *Front. Hum. Neurosci.* 4, 21.
- Cohen, M.A., Konkle, T., Rhee, J., Nakayama, K., Alvarez, G.A., 2014. Processing multiple visual objects is limited by overlap in neural channels. *Proc. Natl. Acad. Sci. Unit. States Am.* 111, 8955–8960.
- Cohen, M.A., Nakayama, K., Konkle, T., Stantic, M., Alvarez, G.A., 2015. Visual awareness is limited, by the representational architecture of the visual system. *J. Cogn. Neurosci.* 27, 2240–2252.
- Cohen, M.A., Alvarez, G.A., Nakayama, K., Konkle, T., 2017. Visual search for object categories can be predicted across all of higher-level visual cortex. *J. Neurophysiol.* 117, 388–402.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Deen, B., Richardson, H., Dilks, D.D., Takahashi, A., Keil, B., Wald, L.L., Kanwisher, N., Saxe, R., 2017. Organization of high-level visual cortex in human infants. *Nat. Commun.* 8, 13995.
- Dilks, D.D., Julian, J.B., Paunov, A.B., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. *J. Neurosci.* 33, 1331–1336.
- Fedorenko, E., Hsieh, P.-J., Nieto-Castanon, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. A new method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104, 1177–1194.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.
- Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery, constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* 20, 70–80.
- Golarai, G., Ghahremani, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D., Grill-Spector, K., 2007. Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.* 10, 512–522.
- Golarai, G., Liberman, A., Yoon, J.M.D., Grill-Spector, K., 2010. Differential development of the ventral visual cortex extends through adolescence. *Front. Hum. Neurosci.* 3, 80.
- Golarai, G., Liberman, A., Grill-Spector, K., 2015. Experience shapes the development of neural substrates of face processing in human ventral temporal cortex. *Cerebr. Cortex* 27, 1229–1244.
- Grill-Spector, K., Golarai, G., Gabrieli, J., 2008. Developmental neuroimaging of the human ventral visual cortex. *Trends Cognit. Sci.* 12, 152–162.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34, 479–490.
- Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* 4, 435–456.
- Julian, J.B., Fedorenko, E., Webster, J., Kanwisher, N., 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60, 2357–2364.
- Kanwisher, N., 2010. Functional specificity in the human brain: a window in- to the functional architecture of the mind. *Proc. Natl. Acad. Sci. Unit. States Am.* 107, 11163–11170.
- Keil, B., Alagappan, V., Mareyam, A., McNab, J.A., Fujimoto, K., Tountcheva, V., Triantafyllou, C., Dilks, D.D., Kanwisher, N., Lin, W., Grant, P.E., Wald, L.L., 2011. Size-optimized 32-channel brain arrays for 3 T pediatric imaging. *Magn. Reson. Med.* 66, 1777–1787.
- Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. *Nat. Rev. Neuro.*
- Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience, the dangers of double dipping. *Nat. Neurosci.* 12, 535–540.
- Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry, integrating cognition, computation, and the brain. *Trends Cognit. Sci.* 17, 401–412.
- Livingstone, M.S., Vincent, J.L., Arcaro, M.J., Srihasam, K., Schade, P.F., Savage, T., 2017. Development of the macaque face-patch system. *Nat. Commun.* 8, 14897.
- Meissner, T.W., Nordt, M., Weigelt, S., 2019. Prolonged functional development of the parahippocampal place area and occipital place area. *Neuroimage* 191, 104–115.
- Natu, V.S., Barnett, M.A., Hartley, J., Gomez, J., Stigliani, A., Grill-Spector, K., 2016. Development of neural sensitivity to face identity correlates with perceptual discriminability. *J. Neurosignals* 36, 10893–10907.
- Nordt, M., Semmelmann, K., Genc, E., Weigelt, S., 2017. Age-related increase of image-invariance in the fusiform face area. *Dev. Cogn. Neurosci.* 31, 46–57.
- Nunnally, J.C., Bernstein, J.H., 1994. *Psychometric Theory*, (3rd ed.). McGraw-Hill, New York.
- Osher, D.E., Saxe, R., Koldewyn, K., Gabrieli, J.D.E., Kanwisher, N., Saygin, Z.M., 2016. Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cerebr. Cortex* 26, 1668–1683.
- Parvizi, J., Jacques, C., Foster, B.L., Withoft, N., Rangarajan, V., Weiner, K.S., Grill-Spector, K., 2012. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* 32, 14915–14920.
- Passarotti, A.M., Paul, B.M., Buxton, J.R., Buxton, R.B., Wong, E.C., Stiles, J., 2003. The development of face and location processing, an fMRI study. *Dev. Sci.* 6, 100–117.
- Peelen, M.V., Glaser, B., Vuilleumier, P., Eliez, S., 2009. Differential development of selectivity for faces and bodies in the fusiform gyrus. *Dev. Sci.* F16–F25.
- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vis.* 10, 437–442.
- Pelphrey, K.A., Lopez, J., Morris, J.P., 2009. Developmental continuity and change in responses to social and nonsocial categories in human extrastriate visual cortex. *Front. Hum. Neurosci.* 3, 25.
- Pitcher, D., Charles, L., Devlin, J.T., Walsh, V., Duchaine, B., 2009. Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curr. Biol.* 19, 319–324.
- Pitcher, D., Dilks, D.D., Saxe, R., Triantafyllou, C., Kanwisher, K., 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* 56, 2356–2363.
- Postelnicu, G., Zolle, L., Fischl, B., 2009. Combined volumetric and surface registration. *IEEE Trans. Med. Imaging* 28, 508–522.
- Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D.E., Saxe, R.R., 2012. Anatomical connectivity patterns predict face-selectivity in the fusiform gyrus. *Nat. Neurosci.* 15, 321–327.
- Schalk, G., Kapeller, C., Guger, C., Ogama, H., Hiroshima, S., Lafer-Sousa, R., Saygin, Z., Kamada, K., Kanwisher, N., 2017. Facephenes and rainbows: causal evidence for functional and anatomical specificity of face and color processing in the human brain. *Proc. Natl. Acad. Sci. Unit. States Am.* 114, 12285–12290.
- Scherf, K.S., Behrmann, M., Humphreys, K., Luna, B., 2007. Visual category-selectivity for faces, places, and objects emerges along different developmental trajectories. *Dev. Sci.* 10, F15–F30.
- Scherf, K.S., Luna, B., Avidan, G., Behrmann, M., 2011. “What” precedes “which”, Developmental neural tuning in face- and place-related cortex. *Cerebr. Cortex* 21, 1963–1980.

- Spearman, C.C., 1910. Correlation calculated from faulty data. *Brit. J. Psychol.* 3, 271–295.
- Tisdall, M.D., Hess, A.T., Reuter, M., Meintjes, E.M., Fischl, B., van der Kouwe, A.J., 2012. Volumetric navigators for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn. Reson. Med.* 68, 389–399.
- Van der Kouwe, A.J.W., Benner, T., Salat, D.H., Fischl, B., 2008. Brain morphometry with multiecho MPRAGE. *Neuroimage* 40, 559–569.
- Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzilingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* 4, 274–290.
- Wada, Y., Yamamoto, T., 2001. Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *J. Neurol. Neurosurg. Psychiatry* 71, 254–257.
- Wiener, K., Grill-Spector, K., 2012. The improbable simplicity of the fusiform face area. *Trends Cognit. Sci.* 16, 251–254.
- Williams, M., Dang, S., Kanwisher, N., 2007. Only some spatial patterns of fMRI response are read out in task performance. *Nat. Neurosci.* 10, 685–686.