Title: Objects seen as scenes: neural circuitry for attending whole or parts

Short title: Objects seen as scenes

Authors: Mitchell Valdés-Sosa[1,2♯*], Marlis Ontivero-Ortega[3,4♯], Jorge Iglesias-Fuster[1], Agustin Lage-Castellanos[3,5], Jinnan Gong[2], Cheng Luo[2], Ana Maria Castro-Laguardia[1], Maria Antonieta Bobes[1,2], Daniele Marinazzo[4] and Dezhong Yao[2]

Affiliations:

[1] Department of Cognitive Neuroscience, Cuban Center for Neuroscience, Havana, Cuba.

[2] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China.

[3] Department of Neuroinformatics, Cuban Center for Neuroscience, Havana, Cuba.

[4] Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium.

[5] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands.

[♯]These authors contributed equally.

*Correspondence to: mitchell@cneuro.edu.cu

Abstract (225 words)

Depending on our goals, we pay attention to the global shape of an object or to the local shape of its parts, since it's difficult to do both at once. This typically effortless process can be impaired in disease. However, it is not clear which cortical regions carry the information needed to constrain shape processing to a chosen global/local level. Here, novel stimuli were used to dissociate functional MRI responses to global and local shapes. This allowed identification of cortical regions containing information about level (independent from shape). Crucially, these regions overlapped part of the cortical network implicated in scene processing. As expected, shape information (independent of level) was mainly located in category-selective areas

specialized for object- and face-processing. Regions with the same informational profile were strongly linked (as measured by functional connectivity), but were weak when the profiles diverged. Specifically, in the ventral-temporal-cortex (VTC) regions favoring level and shape were consistently separated by the mid-fusiform sulcus (MFS). These regions also had limited crosstalk despite their spatial proximity, thus defining two functional pathways within VTC. We hypothesize that object hierarchical level is processed by neural circuitry that also analyses spatial layout in scenes, contributing to the control of the spatial-scale used for shape recognition. Use of level information tolerant to shape changes could guide whole/part attentional selection but facilitate illusory shape/level conjunctions under impoverished vision.

Significance statement

One daily engages hierarchically organized objects (e.g. face-eyes-eyelashes). Their perception is commonly studied with global shapes composed by of local shapes. Seeing shape at one level is easy, but difficult for both at once. How can the brain guide attention to one level? Here using novel stimuli that dissociate different levels over time and examining local patterns of brain-activity, we found that the level and shape of visual objects were represented in segregated sets of cortical regions, each connected into their own pathway. Level information was found in part of the cortical network known to process scenes. Coding of object-level independently from shape could participate in guiding sustained attention within objects, eliminating interference from irrelevant levels. It could also help produce “illusory conjunctions” (perceptual migration of a shape to the wrong level) when attention is limited.

Highlights

- Modified Navon figures allow dissociation in time of fMRI responses for the global/local levels.
- Shape-invariant hierarchical level information was found in scenes selective areas, whereas level-invariant shape information was found in object- and faces- selective areas.
- Level and shape regions were divided by the mid-fusiform sulcus (MFS) in VTC cortex, and each type of region connected into its own pathway.
- Having separate level/shape pathways could facilitate selective-attention, but foster illusory conjunctions.

Author contributions:

Designed research: MVS MOO JIF
Performed research: MVS JIF JG CL ACL MAB DY
Contributed analytic tools: MVS MOO DM ALC
Analyzed data: MVS MOO JIF
Wrote the paper: MVS MOO JIF

Keywords: Hierarchical figures, global/local Navon figures, fMRI, MVPA scene-selective, functional connectivity, VTC

Introduction

Our visual world is full of hierarchically organized objects, and at different times it is more important to attend one echelon than the others (Kimchi, 2015). Consider identifying either the shape of a tree (a whole) in contrast to identifying the form of a leaf (a part). Visual shape is recognized along a series of strongly connected cortical patches within the lateral VTC (Grill-Spector and Weiner, 2014; Moeller et al., 2008), where increased tolerance to features not essential for object recognition (e.g. size, position, and viewpoint), and larger receptive fields (RFs), emerge by stages. It is improbable that each level of an object has its own neural mechanism for shape recognition (Tacchetti et al., 2018). Convergence of inputs from different levels on a common processor is more plausible, and this means potential competition for the same neural resource (Chelazzi, Miller, and Duncan, 1993).

Between-level competition has been traditionally studied with Navon figures (Fig. 1A): global shapes made out of local shapes (Navon, 2003). In real-world examples, shapes, cognitive tasks, and affordances usually differ across the global and local levels, but these can be equated between levels in Navon figures. Hence, differences in behavioral or neural responses to global/local shapes would largely depend on perceptual/attentional factors. However, one limitation of Navon figures is that neural responses for each level are difficult to separate for analysis, since global and local shapes onset and offset at the same time. In a previous work, we developed modified-Navon figures (Fig. 1B) that make it possible to present global/local

elements dissociated over time (López et al., 2002). This allows more precise psychophysics and the separation of neural responses associated to each level.

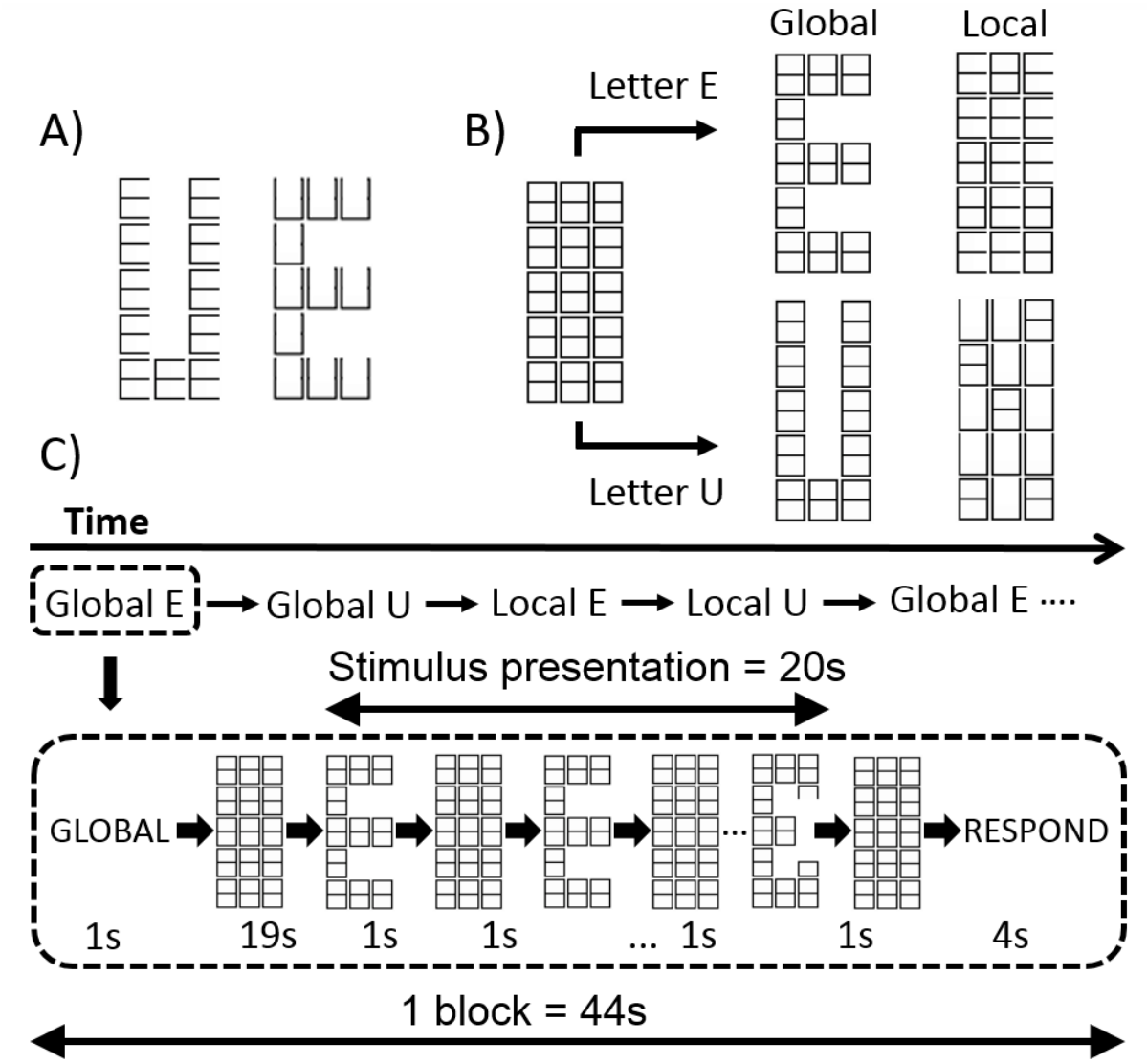

Figure 1. Stimuli and experimental design. A) Two traditional Navon figures: a global 'U' made of local 'E's, and a global 'E' made of local 'U's. Note that letters are present simultaneously at both levels. B) Modified Navon figures which emerged out of a mask. Two letter shapes ('E' and 'U') and two levels (global and local) were used. Note that levels were revealed one at a time, thus allowing separation of their neural responses. This is not possible with traditional Navon figures. C) On the top the order in which the stimulus- blocks were shown. One example stimulus-block is expanded at the bottom, in which the global 'E's alternated with the mask. Participants detected occasional deviant shapes (in this example the last letter).

Using this new paradigm, it is easy to recognize consecutive stimuli with arbitrary shapes within one level (even at fast rates), but it is difficult to do so for both levels at once (Iglesias-Fuster et al., 2015). Also, shifting attention between levels takes time, especially from global to local shapes in typical subjects (Valdés-Sosa et al., 2014). Therefore, fixing attention at the same level -when possible- is an efficient strategy, which requires guidance by cortical representations of level that need to be agnostic about shape. This type of attentional control can fail, producing illusory shape/level perceptual conjunctions (Hübner and Volberg 2005) and is atypical in conditions such as autism (White et al., 2009) and Alzheimer's disease (Jacobson et al., 2005; Slavin et al., 2002).

Oppositely to shape information, which maps reliably to category-specific cortical areas (Kanwisher, 2010), the mapping of level information (i.e.: tolerant to variations in other properties) is unclear. Level and shape are coded conjointly in early visual areas, where rudimentary attention to wholes and parts could arise by respectively selecting larger or smaller regions of visuotopic cortex, that is by varying the attentional-zoom (Sasaki et al., 2001). For higher-order visual areas, alternative hypothesis about shape-invariant level information can be posited. Firstly (H1), features associated with level (i.e. object-size) could simply be discarded during extraction of level-invariant representations of shape in VTC (Rust and DiCarlo, 2010). Another hypothesis (H2), suggested by research in monkeys VTC (Hong et al., 2016), is that codes for shape/level (invariant to each other) emerge together in the same pathway. Finally (H3), invariant codes for level (ignoring shape) could be extracted within a yet unidentified pathway, parallel to the route extracting shape within or outside of VTC. This would be analogous to a circuit in monkey VTC (Chang et al., 2017), in which information about color hue

is refined across cortical patches while of information about shape is reduced. Since the input and output connections of any cortical region determine its function (Saygin et al., 2016), these hypothesis imply also different functional wiring schemes (Osher et al., 2016).

Here we used functional MRI (fMRI) and our modified Navon figures in order to adjudicate among the three hypothesis. Invariant information associated to sensory attributes can be located in the cortex using multivariate pattern analysis (MVPA) based on cross-classifications (Kaplan et al., 2015), whereas network organization can be studied with functional connectivity (FC) methods (Friston, 2011). The three hypotheses described above predict distinct outcomes on these fMRI measures. H1 entails a failure of MVPA in revealing shape-invariant level information in higher-order visual areas, whereas H2 implies that both level/shape information present to the same degree in most of them. H3 -separate pathways- predicts that level and shape information are distributed unequally over the cortex. If the latter is observed, then stronger FC between areas with the same (compared to different) level/shape specialization would support the idea of independent neural pathways extracting each attribute.

To anticipate our results, H3 explained the data best with a cortical separation of level and shape information -invariant to each other- and tighter connections between areas with similar specialization. The location of level/shape information revealed by MVPA were also compared to well-known category-selective visual regions. Although the stimuli used here are clearly objects/letters, unexpectedly more invariant information about level, compared to shape, was found in part of the scene-selective cortex. Unsurprisingly, shape information was stronger in object-selective cortex. These results serve to better typify the networks involved in visual recognition, and suggest two parallel pathways within VTC.

Materials and Methods

Main experiment

Participants

Twenty-six students from the University for Electronic Science and Technology of China (UESTC) participated in the main experiment (ages ranged from 23 to 28 years, 9 females). All had normal (or corrected-to-normal) vision and were also right handed except for two cases. None had a history of neurological or psychiatric disease. The experimental procedures were previously approved by the UESTC ethics committee, carried out in accordance with the declaration of Helsinki, with participants giving written informed consent.

Stimuli and task

Stimuli were projected onto a screen at the subject's feet, viewed through an angled mirror fixed to the MRI head-coil, and generated using the Cogent Matlab toolbox (http://www.vislab.ucl.ac.uk/cogent.php). Modified Navon figures were built as follows. A matrix of white lines (about 2.0° wide and 5.3° high), on a black background, was used as a baseline stimulus. This matrix was built out of 15 small placeholder elements shaped like '8's (spanning visual angles about 40' wide and 1° 3' high). Local letters were uncovered by eliminating selected lines within 10 out of 15 possible placeholders most placeholders, whereas global letters were uncovered by completely eliminating several placeholders (see Fig. 1B). At both levels the letters 'E' and 'U' were presented. Small deviations in letter-shapes were used as oddball stimuli.

Each stimulus-block was initiated by a 1 sec cue ('Global' or 'Local'), that announced the upcoming letter level, followed by a 19 sec baseline (see Fig. 1C and SI movie for examples). The same letter (at a fixed level) was presented 10 times during 1 sec alternating with 1 sec baseline period. To encourage attention to the stimuli, participants were asked to count the number of oddball letters within a block (which occurred either 0, 1, or 2 times per block in equal proportions). Blocks finished with a 4 sec 'respond' signal, with participants reporting the number of oddballs via a MRI-compatible button pad (detection accuracy in all participants were above 85%). Thus each block of recurring letters lasted 20 sec, and was separated from the next one by another 24 sec.

The experiment was divided into runs separated by 1 to 2 minute breaks. Each run contained 12 stimulus-blocks, that were 3 repetitions of the sequence: global 'U', global 'E', local 'U, and local 'E'. Most participants completed 5 runs, except two who only completed four. Consequently each type of stimulus-block was repeated a total of either 15 or 12 times in the experiment.

Data acquisition

Recordings for the experiment were obtained with a GE Discovery MR750 3T scanner (General Electric Medical Systems, Milwaukee, WI, USA) using an 8 channel receiver head coil. High-spatial resolution (1.875 x 1.875 x 2.9 mm) functional images (fMRI), with 35 slices covering all the head except the vertex, were acquired. A T2*- weighted echo planar imaging sequence was used with the parameters: TR=2.5s; TE=40 ms; flip angle=90° and acquisition matrix=128 x 128. There were 135 images per run. The initial 5 volumes were discarded for all runs, to stabilize T1 magnetization. A 262 slice anatomical T1-weighted image was also obtained

with the following parameters: voxel size=1 x 1 x 0.5 mm; TR=8.10 ms; TE=3.16 ms; acquisition matrix=256 x 256; and flip angle=12.

Image preprocessing

Preprocessing was carried out with SPM8 (https://www.fil.ion.ucl.ac.uk/spm/). The functional scans were first submitted to artifact correction using the ArtRepair toolbox (http://cibsr.stanford.edu/tools/ArtRepair/ArtRepair.htm), thus repairing motion/signal intensity outliers and other artifacts (including interpolation using nearest neighbors for bad scans). Then slice-timing, head motion correction (including extraction of motion parameters), and unwarping procedures were applied.

White matter and pial surfaces were reconstructed from each T1-weighted image using Freesurfer software (http://surfer.nmr.mgh.harvard.edu), then registered to the FsAverage template, and subsequently subsampled to 81924 vertices (https://surfer.nmr.mgh.harvard.edu/fswiki/FsAverage). The mid-gray surface was calculated as the mean of white and pial surfaces vertex coordinates. Each T1 image was co-registered with the mean functional image, and the transformation matrix was used to project the mid-gray surface into each subject´s functional native space. Volume BOLD signals were interpolated at the coordinates of the mid-gray vertices, producing surface time-series (without spatial smoothing). Time series were high-pass filtered with cutoff of 0.008 Hz, their means and linear trends removed, and were also normalized (z-scored), all of which was performed for each run separately. A general linear model (GLM) was fitted to the time series of each surface vertex using a regressor for each stimulation block (i.e. square-waves convolved with the canonical hemodynamic function), in addition to the head movement parameters and the global mean of

the white matter signal as nuisance covariates. The beta parameters estimated for each block (trial) were used as features for MVPA, whereas the residual time series or background activity (BA) were used to perform the FC analysis (Al-aidroos et al., 2012).

Searchlight MVPA for decoding invariant information for level and shape

Searchlight analysis

The presence of invariant level/shape information in the fMRI was gauged in a searchlight analysis (Kriegeskorte et al., 2006) over the cortical surface (also known as informational mapping). Discs with 10 mm radii of geodesic distance were defined around all FsAverage surface vertices (mean number of nodes: 186; range: 53-471) by means of the fast_marching toolbox (https://github.com/gpeyre/matlab-toolboxes). MVPA was performed in each subject on the beta parameters for the set of surface vertices within each searchlight using a fast Gaussian Naïve Bayes classifier developed for searchlight computation (Ontivero-Ortega et al., 2017) in a leave-one run out cross-validation scheme (for 5 runs, 12 trials for training and 3 trials for testing per condition).

Cross–classification was employed to ascertain invariant level and shape information in each searchlight (Kaplan et al., 2015). Shape-invariant level information was considered to be present if letter level was predicted accurately (i.e. above chance) for one letter by a classifier trained on exemplars of the other letter (Fig. 2A). This would imply patterns for level indifferent to letter shape. Conversely, level-invariant shape information was considered to be present if letter identity at one level could be predicted accurately by a classifier trained on exemplars from

the other level. This would imply multivariate patterns for shape tolerant to large changes in physical format between different levels (Fig. 2B).

The individual accuracies in each searchlight were averaged across folds and both directions of the cross-classifications, logit transformed (logit-acc), and assigned to the central node to create searchlight maps. These logit-acc were entered into a group t-test vs chance level (accuracy=50%, logit-acc=0). A false discovery rate (FDR) threshold of q=0.05 was used to control the effects of multiple comparisons in all maps, based on estimation of empirical Gaussian null distributions (Schwartzman et al., 2009). Note that the use of cross-classification mitigates some limitations of group-level t tests for MVPA (Allefeld et al., 2016).

Searchlights may be informative due to differences across conditions in the local patterns of activity, in the response amplitudes, or in both (Coutanche, 2013). To test if local patterns contribute, for each trial the mean of the betas in a searchlight was removed (i.e. the data was centered) before MVPA. To test if the amplitudes contribute, the beta values in each searchlight were replaced by their average across nodes before MVPA (i.e. the data was smoothed). This corresponds to spatial smoothing with a cylindrical (instead of a Gaussian) window. These transformed maps were analyzed as described above. The correspondence between different types of maps was assessed with the Dice coefficient (Dice, 1945).

Index of Specialization (IOS)

The degree of specialization for level/shape invariant information was characterized in each searchlight by an index of specialization (IOS, Fig. 2C), which reflected the relative accuracy of the two types of cross-decoding. IOS was expected to reveal spatially segregated

cortical sectors according to H3 but not H1 or H2. After expressing logit-acc for the two attributes in polar coordinates (negative values were substituted by zero), IOS was defined as the polar angle minus 45 degrees. Therefore -45 degrees corresponded to maximum specialization for shape information, 45 degrees to maximum specialization for level information, and zero to no preference. IOS was only calculated in informative searchlights (i.e. in which at least one of the t-tests vs chance was significant). This measure is similar to the activity profiles of different stimuli (expressed on spherical coordinates) that have been used in clustering voxels (Lashkari et al., 2010), although here informational- instead of activity- profiles were employed. Henceforth, two non-overlapping functional regions (ROI) were defined in each hemisphere: a shape-invariant level ROI considering all nodes with IOS>10, and a level-invariant shape ROI with IOS<-10.

To help describe the searchlight maps, nodes corresponding to the mid-fusiform sulcus (MFS) (Weiner, 2018) were identified manually on the FsAverage surface. Also, the VTC region was defined as all the vertices included in fusiform gyrus, the lingual gyrus, and the lateral occipito-temporal, collateral, and transverse collateral sulci, according to the Freesurfer atlas (Destrieux et al., 2010), but only posterior to the rostral tip of MFS.

Comparison of MVPA with face- and scene-selective areas

Localizer experiment

Face- and scene- selective areas were localized using a standard paradigm (Epstein and Kanwisher, 1998) in an independent group of participants composed of sixteen Cuban participants, with ages from 25 to 36 years (7 females). All participants had normal (or

corrected-to-normal) vision and were right handed except for one case. None had a history of neurological or psychiatric disease. The experimental procedures were previously approved by the ethics committees of the Cuban Center for Neuroscience, and participants gave written informed consent. One scanning run was used, each consisting of 16 blocks of stimuli. The blocks lasted 30 s and each containing 60 stimuli, which in turn were presented during 500 ms. In half of these blocks pictures of faces were presented, and on the other half pictures of houses. Both types of blocks were alternated, separated by 30 s of fixation. All stimuli had the same dimensions (6.2° x 8.25°) and were presented in grayscale on black background. Stimuli were projected on a screen at the subject's head, viewed through an angled mirror fixed to the MRI head-coil, and generated using the Cogent Matlab toolbox. The data were acquired with a Siemens MAGNETOM Allegra 3T scanner (Siemens, Erlangen, Germany), using a single channel receiver head coil.

High-spatial resolution functional images (1.8 x 1.8 x 1.8 mm) were collected using a T2*-weighted echo planar imaging sequence, with 47 axial slices, interleaved, effectively covering the entire brain. The parameters for image acquisition were as follows: TR = 3s, TE = 30 ms, flip angle = 90 °, acquisition matrix = 128 x 120. For each subject, 340 volumes were collected. A 176 slice anatomical T1-weighted image was obtained with the following parameters: voxel size = 1 x 1 x 1 mm; TR = 1940 ms; TE = 3.93 ms; acquisition matrix = 256 x 256; and flip angle = 9.

Analysis of the localizer data

The localizer data was analyzed as the main experiment up to the massive GLM to estimate the beta values, except that the data was smoothed on the surface with an 8 mm

Gaussian kernel. Then, the contrasts of the face>house, face>fixation-point, and house>fixation-point were calculated. The latter two contrasts were entered into a group t-test vs zero and thresholded with an FDR at q=0.05 (based on an empirical null distribution), and were also expressed as an IOS values (see Fig. 4B). This face/house-IOS was calculated as the polar angle between the mean face>fixation-point, and house>fixation-point contras for all nodes in which at least one of the t-tests was significant.

To further test the relationship between level/shape information with the category-specific regions, we performed a more detailed analysis limited to VTC region using the face>house contrast. These contrast values were averaged in each individual within the level/shape masks created with the IOS, separately for the two hemispheres. The averaged contrast values were then submitted to a repeated measures-ANOVA using Category (face/scene) and Hemisphere as main effects. Subsequently, the Pearson correlation coefficient was calculated between the average IOS values and the results of t-tests for the face/house contrast. Note that in this analysis the dependent and independent measures were defined in by experiments in separate groups of participants.

Background activity connectivity analysis

Comparison of the FC between regions with similar and dissimilar IOS.

We first tested if the FC was stronger between areas with the same specialization, rather than dissimilar specialization. The regions for invariant level and shape (excluding V1 and V2) were each divided into spatially separated clusters. Only clusters containing more than 100 nodes were considered. In each subject, BA time series were averaged across nodes in each cluster and

the partial correlations (used as a FC measure) were calculated between all pairs of these IOS-defined clusters (controlling for the correlation with the other clusters). Cells of the resulting partial correlation matrices were averaged within each individual after grouping according to two main effects: Pair-similarity (shape-shape, level-level or shape-level) and Hemisphere (clusters in the same vs. different hemisphere). A repeated measure ANOVA was then performed on the reduced matrix values after a Fisher transformation, using the main effects described before.

Prediction of IOS from FC

To verify the relationship between FC and cortical specialization we carried out a comparison between a functional connectivity parcellation and IOS maps but only in the VTC region for each hemisphere separately. First, a vertex-wise FC matrix was estimated in each individual by calculating the Fisher transformed Pearson correlation between all cortical nodes in VTC. The group-level matrix was thresholded with a t-test against zero that was corrected for multiple comparisons ($p<0.05$, Bonferroni corrected). This mean FC matrix was then partitioned into two new FC-defined clusters using Spectral Clustering (SC) method (Von Luxburg, 2007). The hypothesis was that two regions specialized (for shape and level) would emerge from SC method. Two ensure stability, the SC was repeated 100 times and a consensus matrix C (Lancichinetti and Fortunato, 2012) was built based on the different parcellations across the iterations. The final two FC-defined clusters were generated over C using again SC, and the mean IOS across their nodes was calculated.

Tests of models of FC structure in VTC

Finally, we studied the topology of FC inside VTC, testing several theoretical models that could explain partial correlations within this region, with an approach analogous to that used in Representational Similarity Analysis (RSA) (Kriegeskorte et al., 2008). In each hemisphere, the VTC ROI was restricted by excluding non-informative-searchlights, as well as the V1 and the V2 regions. The BA time-series of each node was spatially smoothed on the surface by replacing it with the average of all nodes in the (10 mm radius) searchlight surrounding it. Each restricted VTC was divided in 10 patches along the caudal-rostral direction by a k-means partition of the Y coordinates of the surface nodes. The two FC-defined clusters in each hemisphere described above were then subdivided with these 10 patches. This yielded in both hemispheres 8 and 9 patches for each cluster respectively (Fig. 5C). A patch-wise observed FC matrix was estimated in each participant, by calculating partial correlations (to remove spurious or indirect associations) with the averages of BA within patches.

Alternative theoretical models (expressed as patch-wise matrices) were built as explanations of the pattern of connectivity in VTC. These models were: a) connectivity by simple spatial proximity (indexed by the matrix of average between-node geodesic distances over the cortex for all pairs of patches); b) original FC-defined cluster membership; c) patch contiguity in the caudal-rostral direction (CR adjacency); and d) lateral patch contiguity (lateral adjacency). The models and the patch-wise FC matrices were vectorized. Multiple regression was carried out separately for the two hemispheres in each subject using the theoretical correlation values (corresponding to each model) as the independent variables and the observed connectivity values as the dependent variable ($obsCorr = Intercept + B_1 * geodesic\ distance +$

B_2*cluster + B_3*CR-adjacency + B_4*Lateral-adjacency*). The intercept was considered a nuisance variable. The resulting beta values (B_i) across participants were submitted to a random effect t-test against zero. The effect size of each predictor was estimated by bootstrapping the test (*n=1000*). This allowed selection of the theoretical matrix most similar to the observed patch-wise FC matrices.

Results

Information about level and shape are carried by different cortical regions

Invariant level and shape searchlight maps (Fig. 2D, E and Supplementary Fig. S2A, B) overlapped only moderately (Dice coefficient = 0.5). Referenced to anatomical landmarks (Destrieux et al., 2010), level information was most accurately decoded in the occipital pole, but also in the medial portion of the fusiform gyri and the lingual gyrus, the collateral and traverse collateral gyri, as well as medial occipital areas. Conversely, shape information was concentrated in the lateral occipital and posterior lateral fusiform gyri, as well as the lateral-occipital sulci.

The centered searchlight maps (Fig. 3A, B and Fig. S3A, B).were very similar to the original (untransformed) maps (Dice coefficients: for level=0.82; for shape=0.74) Thus local patterns contribute to information for both level and shape at most sites. Although the smoothed (Fig. 3D, E; and Fig. S4A, B) and original maps for shape were also similar (Dice coefficient=0.69), they were different for level (Dice coefficients for level=0.17). This discrepancy is due to the fact that the smoothed map for level was only informative in caudal, but not rostral (higher-order) visual areas. Thus, information about level was carried by amplitude exclusively in caudal early visual cortex.

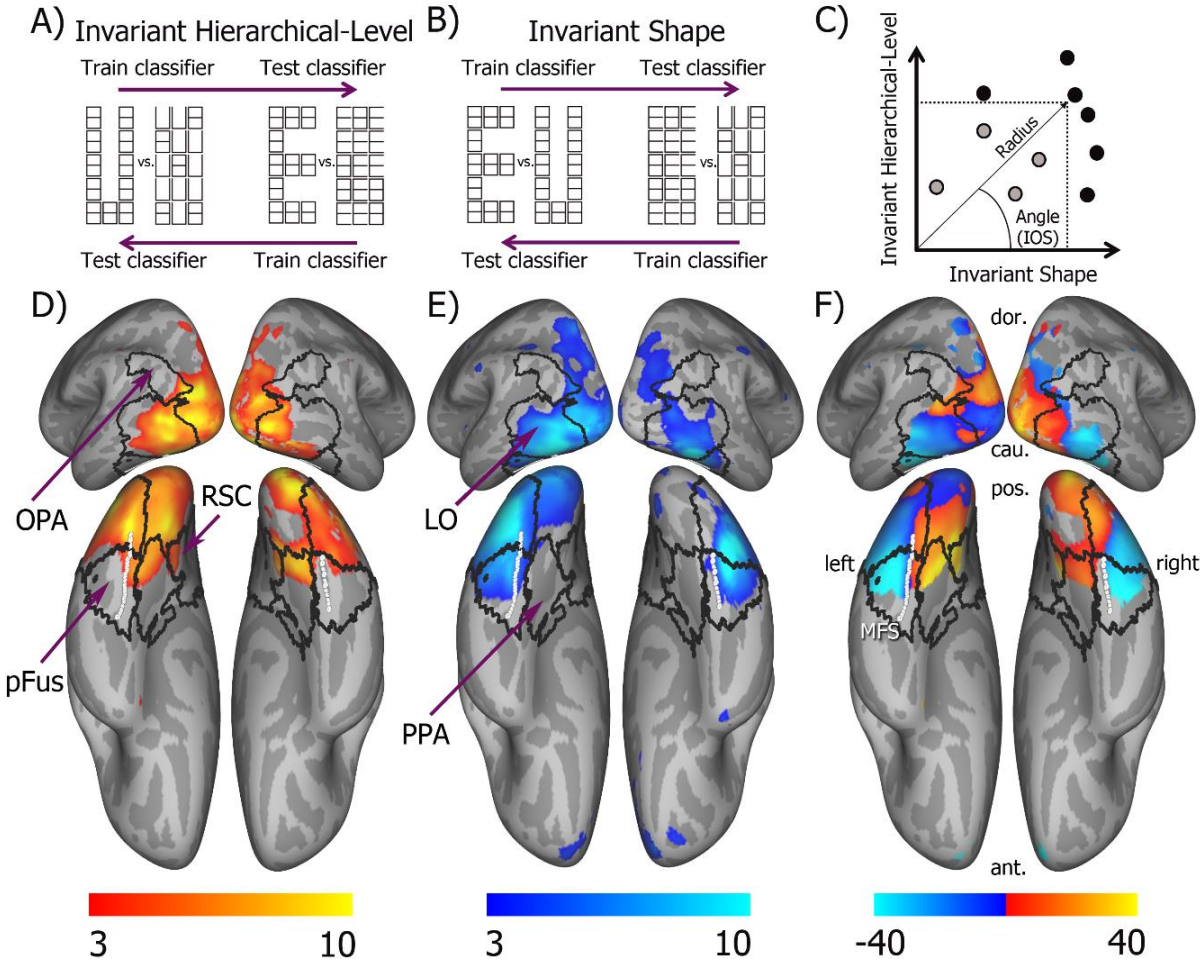

Figure 2. Searchlight maps. A) Design of cross-classification procedure for shape-invariant level. B) Design of cross-classification procedure for level-invariant shape. For A and B accuracies of classification in the two directions were averaged. C) IOS defined as the polar angle between the accuracies of level and shape classification at each searchlight (minus 45 degrees). Dotted lines are FDR thresholds for each cross-classification. D) Searchlight map for shape-invariant hierarchical-level decoding, showing group Z-scores of above-chance classification (thresholded FDR q=0.05). E) Corresponding searchlight map for level-invariant shape decoding. F) Group IOS map. Only searchlights informative for at least one cross-classification are shown (drawn in C as black circles). Also, borders of scene- and object-selective areas are shown in black lines. Object-selective areas: LO lateral occipital; pFus posterior fusiform. Scene-selective areas: PPA Para-hippocampal place area; RSC retrosplenial complex; OPA occipital place area. The MFS is depicted in white dotted lines. Acronyms: dor.: dorsal; cau.: caudal; pos.: posterior; ant.: anterior

Description of IOS regions respect to ROIs from probabilistic atlases

The IOS map (Fig. 2F and Fig. S2C) confirmed the spatial segregation level or shape information preference. Singularly, a sharp boundary at the mid-fusiform sulcus (MFS), divided VTC into areas with different informational profiles: a lateral area more specialized for shape and a medial one more specialized for level (Fig. S6). The fact that level and shape were carried by different (albeit overlapping) cortical areas rebutted H1-2 and vindicated H3. The centered IOS (Fig. 3C and Fig. S3C) and original IOS maps were highly similar (Dice coefficient = 0.83), however the smoothed IOS (Fig. 3F and Fig. S4C) and original IOS maps overlapped less (Dice coefficient =0.48).

Regions where the IOS map showed that level and shape were invariant to each other were compared to well-characterized functional regions of interest (fROIs) using population templates of visual category-specific cortex. Scene-selective areas (Zhen et al., 2017), defined by the contrast of scenes vs visual objects, included the para-hippocampal place area (PPA), the retrosplenial complex (RSC) and the occipital place area (OPS). Object selective areas, defined by the contrast of objects vs scrambled objects, (http://www.brainactivityatlas.org/atlas/atlas-download/), included two portions of the lateral occipital complex (LOC): the lateral occipital (LO) and the posterior fusiform areas (pFus). An atlas of 25 retinotopic maps (Wang et al., 2015) was also used. Searchlights preferring shape-invariant level information occupied (see Supplementary Table S1) the posterior portion of the PPA in both hemispheres, with a slight

overlap with the ventral-posterior portion of the RSC and the OPA. Additionally, overlap was found with the superior posterior portions of LO division in both hemispheres.

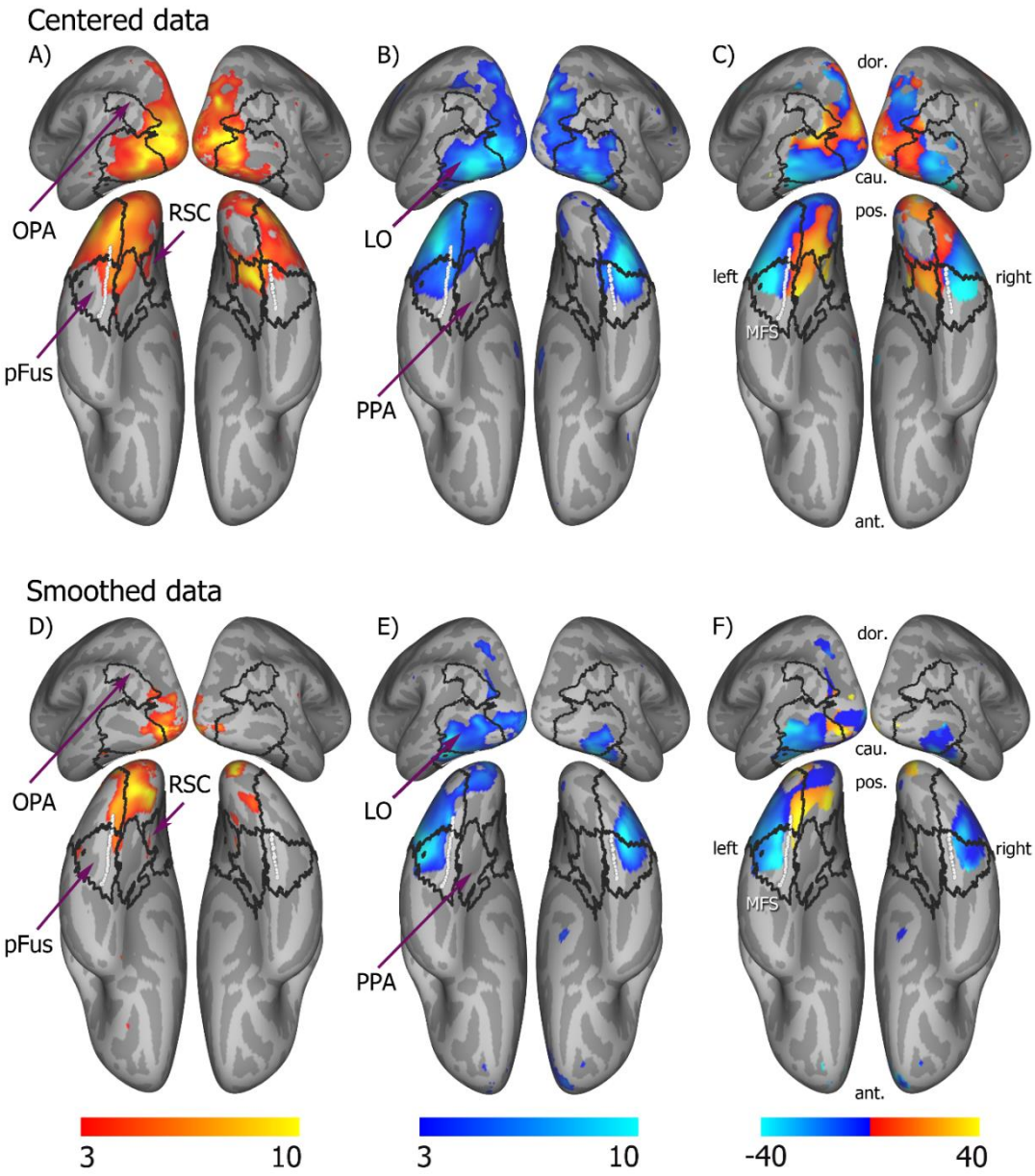

Figure 3. Searchlight maps for centered and smoothed data. Centered maps: A) Shape-invariant level map; B) Level invariant shape map; C) IOS map. Smoothed maps: D) Shape-invariant level map; E) Level invariant shape map; F) IOS map. ROI names and orientation conventions as in Fig. 2.

Referred to visuotopic maps (see Supplementary Figure S5 and Table S1), shape-invariant level information was found in PHC1, VO2, V3 and ventral V3 as well as V3A/B and superior posterior parts of LO1-2 in both hemispheres. PHC2, ventral V2, and V1 on the right, as well as VO1 on the left side were also involved. Note that PHC1-2 corresponds to the retintopically-organized posterior part of PPA. Similar results have been obtained in previous mapping of scene-selective ROIS onto retinotopic areas (Epstein and Baker, 2019; Malcolm et al., 2016; Edward H Silson et al., 2016). Importantly, whereas mean-corrected and smoothed decoding of shape-invariant level were both accurate in all early visual (V1-V3, hV4), only mean-centered decoding was found in VO1-2 and PHC1-2. This indicated that the visual features underlying level information shifted in the caudal to rostral direction.

Conversely to level, and unsurprisingly, searchlights preferring level-invariant shape information in both hemispheres overlapped the LOC, specifically antero-ventral LO and most of pFus, as well retinotopic areas LO1-2 and TO1-2 (see Supplementary Figure S5 and Table S1). These areas are well known to be selective for visual objects (Kourtzi and Kanwisher, 2000). Note that the border between level and shape dominance corresponded better to the MFS than to the borders of the PPA, LO, and pFus fROIs.

Characterization of IOS regions using a face/house localizer

The unexpected location of shape-invariant level information in scene-selective cortex motivated a further test taking advantage of the fact that houses reliably activate scene-selective cortex. The face/house-IOS map showed remarkable analogies with the shape/level IOS map. The level/shape divide in VTC across MFS was mirrored by a house/face separation. Again the face/house border corresponded better to MFS than the borders of PPA, LO and pFus. In LO the

transition from shape at the antero-ventral towards level in the posterior-dorso portion was mirrored by a face/house transition (Fig. 4A).

The ROIs specialized for level had larger responses for houses than for faces, in both the left ($t(15)=-7.08$, $p<0.001$) and right hemisphere ($t(15)=-4.1$, $p<0.001$). The ROIs for shape had larger responses for faces than for houses in the right hemisphere only ($t(15)=4.1$, $p<0.001$) (Fig. 4C). Correspondingly, a strong correlation was found ($r=-0.64$, $p<10^{-5}$) across cortical nodes between the IOS and t-test scores for the face/house contrast (Fig. 4D) in the VTC.

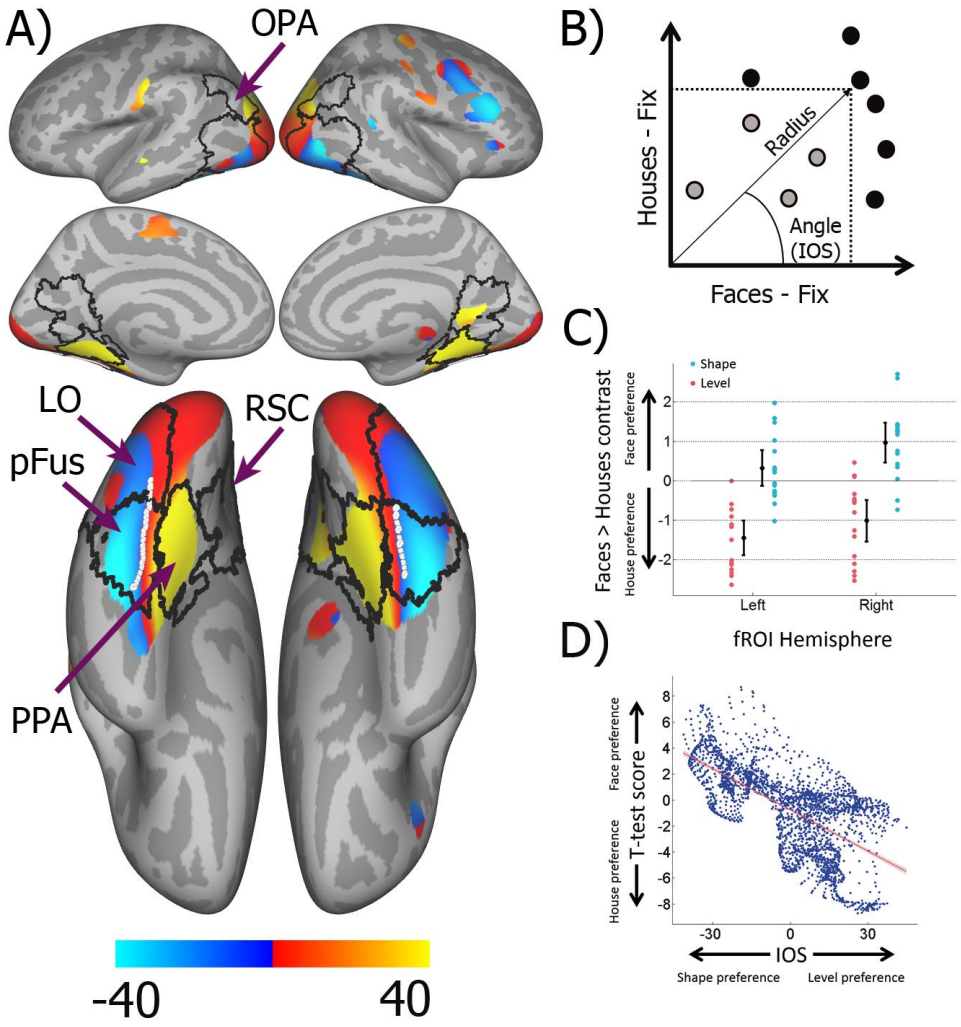

Figure 4. Relationship of regions where level and shape information was invariant with scene- and face-selective areas. A) Group face/house-IOS map. B) Face/house-IOS defined as the polar angle between the contrasts at each node (minus 45 degrees). Dotted lines are FDR thresholds for each t-test. Only nodes significant in at least one t-test were considered (drawn as black circles). ROI name conventions as in Fig. 2. C) Scatterplots of mean value of face-house contrast within masks of cortex specialized for level (red dots) and for shape (blue dots) in each individual, separated by hemisphere. Black dots and whiskers represent means and standard errors of means respectively for each ROI. D) t-test scores of the face-house contrast as a function of IOS plotted for cortical nodes in VTC (excluding V1 and V2).

Areas specialized for level and for shape belong to independent pathways

Pairs of IOS-defined clusters with similar level/shape specialization (IOS) had stronger connectivity than pairs with different specialization. The repeated measure ANOVA showed a highly significant effect of Pair-similarity ($F(2,50)=169$, $p<10^{-5}$) on the partial correlation (Fig. 5A), which was driven by lower values for shape-level than for shape-shape and level-level ($F(1,25)=285$, $p<10^{-5}$) in both hemispheres. No effect of Hemisphere was found, although it interacted significant with Pair-similarity ($F(2,50)=7$, $p<0.002$). This interaction corresponds to significantly larger partial correlations ($F(1,25)=23$, $p<10^{-3}$) for level-level than for shape-shape pairs of different hemisphere.

IOS similarity predicted FC, so we asked if FC could predict the IOS? FC-based parcellation defined in each hemisphere two compact areas that differed in mean IOS values: a lateral region in which shape was predicted better than level (mean IOS = -11.1), and a medial region in which level was predicted more accurately than shape (mean IOS = 16.1) (Fig. 5B). The clusters derived from FC analysis corresponded very well to the segmentation produced by the IOS map. The medial cluster overlapped level-informative areas, whereas lateral cluster overlapped shape- informative areas. In other words, FC predicted cortical specialization. Fig. 5D shows the plots of the group-mean partial correlations between adjacent patches in the caudo-

rostral direction, and between adjacent patches in the lateral direction. It is clear that the patches in the caudo-rostral direction were more connected (i.e. had higher partial correlations) than those in the lateral direction (which crossed the MFS).

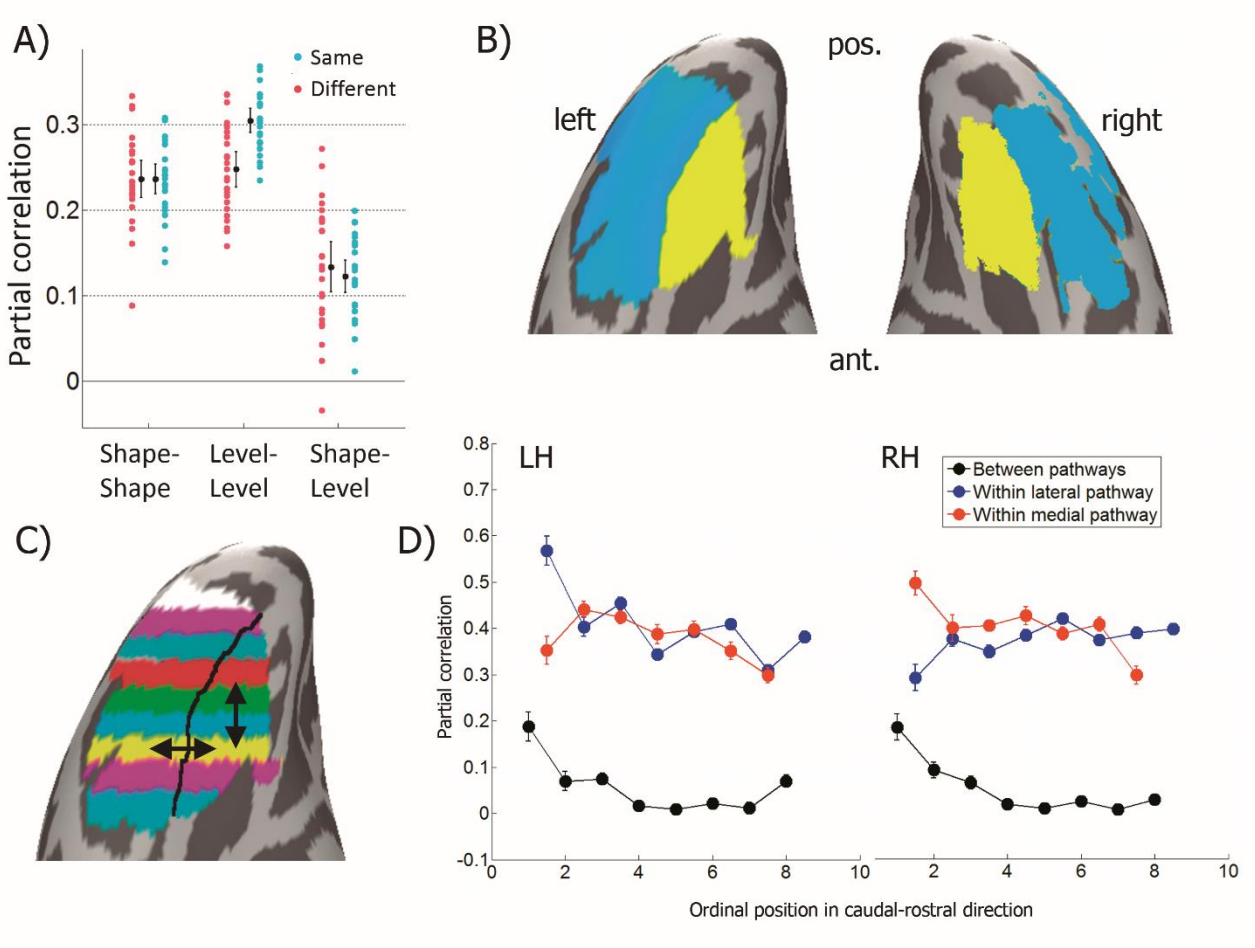

Figure 5. Functional connectivity analysis. A) Scatterplots of mean partial correlations between initial clusters as a function of their similarity in level/shape specialization and if they are in the same or different hemispheres. B) The two clusters obtained by Spectral Clustering method for each hemisphere. C) Subdivision of the latter clusters into parallel bands (here and in D-E shown only in the left hemisphere). Arrows show examples of bands that are adjacent in the caudo-rostral direction and the lateral direction. D) Partial correlations between patches in VTC for each hemisphere; red (lateral cluster) and blue (medial cluster) plots correspond to values between consecutive bands in the anterior-posterior direction, whereas black plot corresponds to values between adjacent patches in the two clusters. In A and D, dots and whiskers respectively represent means and standard errors of means.

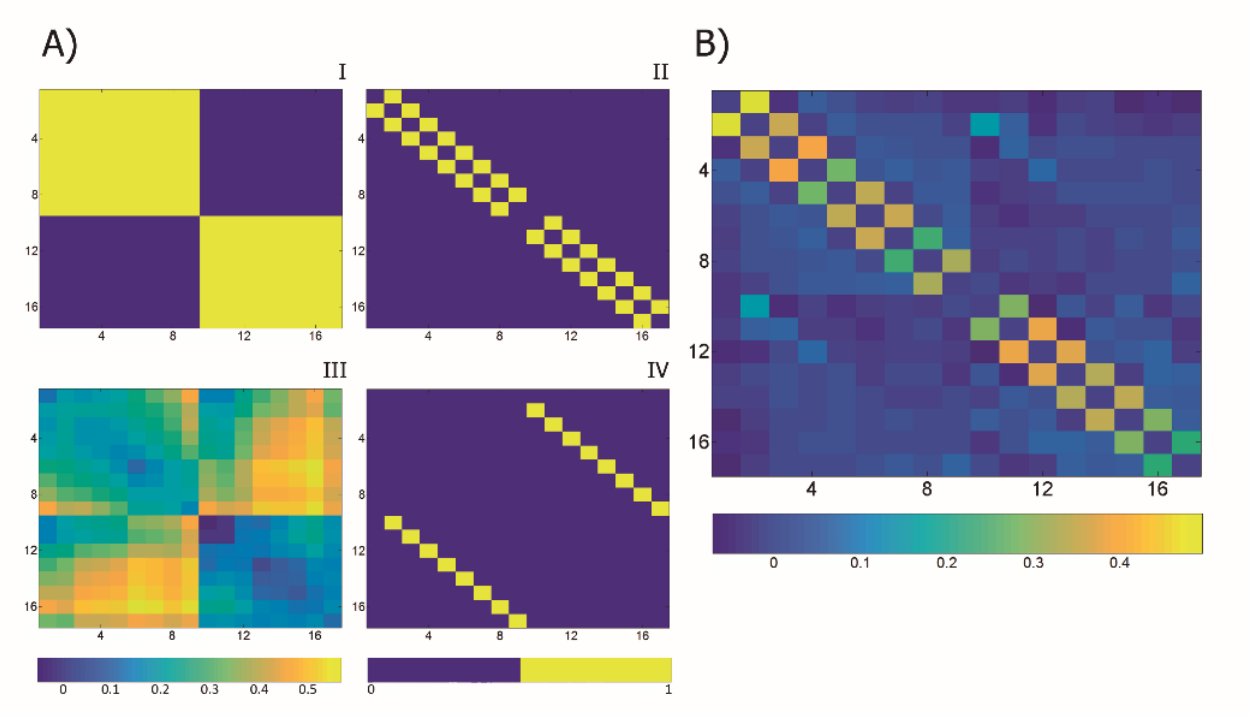

Figure 6. Topology of FC in VTC. A) Competing models: I - Cluster membership; II - Caudo-rostral within-cluster adjacency between bands; III - Geodesic distance between all pairs of bands; and IV - Lateral adjacency between bands from different clusters. B) Experimentally observed group-mean partial correlation matrix.

The topology of FC in VTC was tested more formally by comparing the amount of variance of the observed partial correlation matrix associated with each of the four theoretical matrices outlined above (Fig. 6A). The observed correlation matrix (Fig. 6B) was best explained by the model specifying adjacency in the caudo-rostral direction in both hemispheres, consistent with the effects described above. The effect size of this organizational model was much larger than for the other three models (Table I). This suggests two parallel functional pathways within VTC, each coursing in a caudal-rostral orientation.

In summary, similarity in the cortical specialization of cortical areas for level vs. shape (indexed by the IOS) predicted their FC. On the other hand, FC parcellated the VTC into areas that were specialized for either level or for shape. These results buttress H3, which posits that

hierarchical level and shape of visual objects are processed in separate neural circuitry, and negates H1 and H2.

Table 1. Assessment of competing models as predictors of between-band FC in VTC (multiple regression analysis).

Hemisphere		Intercept	Geodesic distance	Cluster membership	CR direction	Lateral direction
Left	t value	19.225	-0.130	9.192	83.225	7.479
	p value	0.0001	0.898	0.0001	0.0001	0.0001
	effect size	3.770	-0.025	1.803	**16.322**	1.467
Right	t value	19.517	-1.300	7.765	75.631	6.468
	p value	0.0001	0.206	0.0001	0.0001	0.0001
	effect size	3.828	-0.255	1.523	**14.833**	1.268

Discussion

There were two main findings, both of which validate H3 among the competing hypothesis from the introduction. First, cross-decoding of level and of shape (evincing invariance to each other) was unevenly distributed over the cortical surface. Shape-invariant level decoding was better in areas previously characterized as scene-selective. Oppositely, level-invariant shape decoding was better in regions reported to be object- and face-selective. The boundary between these regions with different informational profiles was marked sharply by the MFS, thus extending the list of functional and anatomical parcellations delimited by this sulcus (Grill-Spector and Weiner, 2014). Second, cortical patches having a similar informational specialization were more strongly interconnected than those that had divergent specialization, with a pattern within VTC suggesting two parallel pathways, each placed on a different side of the MFS.

Finding level information in scene-selective cortex was a novel finding based on a methodological innovation. To our knowledge, MVPA has not been used in previous fMRI studies of traditional Navon figures (Han et al., 2002; Weissman and Woldorff, 2005). But even if it were used, shape-invariant level and level-invariant shape information would probably go undetected with these figures. With traditional stimuli, global and local shapes are presented at once, therefore their activity patterns cannot be separated. This is unfortunate because the cross-decoding approach used to diagnose invariance to secondary attributes depends on this separation. Since in our modified Navon figures the two levels are presented separately over time, it is possible to estimate the activity patterns separately for global and local target shapes, thus overcoming this obstacle.

Decoding of hierarchical-level from scene-selective cortex seems counter-intuitive. Navon figures are visual objects not places. However, features implicated in scene processing (Groen et al., 2017) could also be used to code level. The low-level visual feature of spatial frequency serves to distinguish scenes (Andrews et al., 2015; Berman et al., 2017), and is also important in selective attention to hierarchical levels (Flevaris et al., 2014; Flevaris and Robertson, 2011; Valdés-Sosa et al., 2014). In mid-level vision, clutter (Park et al., 2015), amount of rectilinear contours (Nasr et al., 2014), and statistics of contour junctions help distinguish between scenes (Choo and Walther, 2016). These features also differ markedly between the global/local levels of Navon figures (e.g. see Figure 1 in (Flevaris et al., 2014)). Finally high-level features such as spatial layout (in conjunction with object content) can be used to guide the navigation of gaze and attention within a scene (Malcolm et al., 2016). Perhaps the same features would allow Attention to be navigated within Navon figures.

On the other hand, finding shape information in object-selective cortex was unsurprising. Previous MVPA studies have shown that the shape of visual objects can be reliably decoded in LOC with tolerance to variations in their position (Cichy et al., 2011; Stokes et al., 2011; Williams et al., 2008), size and viewing angle (Eger et al., 2008a, 2008b), and rotation (Christophel et al., 2017). Also if the overall shape is maintained, fMRI-adaptation for the same object survives in LOC despite interruptions of its contours (Kourtzi and Kanwisher, 2001). Our results suggest that hierarchical level should be added to this list. It is not clear how this invariance of shape to level is related to the invariance to the other attributes, although possible links are apparent.

As mentioned in the introduction, shape-invariant level decoding could simply reflect the use of different attentional-zooms in the visual field for global and local shapes. Global/local shapes differ in size and spatial frequency content, properties for which peripheral and foveal sectors of visuotopic maps are differentially tuned (Henriksson et al., 2008). Two studies have shown that foveal sectors of early visual areas (V2, V3, V3A, and hV4) are more activated by attention to local -compared to global- stimuli (Rijpkema et al., 2007; Sasaki et al., 2001). Here shape-invariant level was found only at the foveal confluence of early visual areas (V1-3) in the smoothed IOS searchlight maps. Since smoothed maps index amplitude changes, this is consistent with previous reports. Although this effect was absent in higher-order visual areas (VO1-2, PHC1-2), level information was found with centered searchlight maps (which are pattern- but not amplitude- dependent). This suggests a change in code between early (i.e. retinotopic) and higher-order areas (i.e. non-retinotopic). In the latter, the link between visuotopic mapping and anatomy is less direct and strongly modified by attention (Kay et al.,

2015), which implies that a more rigorous test of the attentional-zoom explanation requires measurement of cross-decoding and pRF mapping within the same experiment.

Since the medial and lateral VTC have different retinal eccentricity biases, respectively linked preferentially to peripheral/foveal stimuli and sectors of V1 (Baldassano et al., 2016b; Hasson et al., 2002), there is another twist to the attentional-zoom explanation. Switching attention between the global/local levels could potentially shift activity between medial and lateral VTC. Nevertheless, this scheme cannot explain why we found invariant-level decoding within searchlights restricted to only one sub-region (nor why decoding is more accurate in medial VTC).

Non-retinotopic coding of level is implied by two experimental findings from the literature. Attention to the one level of a Navon figure selectively primes subsequent report of one of two superimposed sinusoidal gratings (Flevaris and Robertson, 2011). Attention to the global/local level respectively favors the grating with lower/higher spatial frequency, in apparent support of retinotopic coding. However, the relative (not the absolute) frequency of the two gratings determines this outcome (Flevaris and Robertson, 2016), which precludes coding by mapping onto retinal eccentricity. Moreover, although larger activations for big –compared to small- real world size objects have been found in medial-VTC, this effect is tolerant to changes in retinal size. Both these phenomena suggest that a mechanism for extracting relative scale exists in the medial VTC, which could be used to represent shape-invariant level.

Our results suggest that the internal topology of VTC is dominated by rostro-caudal connections, with a division that maps onto level/shape specialization. A recent analysis (Haak and Beckmann, 2016) using functional connectivity patterns of twenty-two visuotopic areas

(based on the same atlas used here, Wang et al., 2015) found a tripartite organization that is consistent to some extent with our results. One pathway lead from V1 into VO1-2 and PH1-2, largely overlapping our medial route. Another pathway lead into LO1-2 and TO1-2, partially overlapping our lateral route. Unfortunately in this study pFus (which looms large in our results) connectivity was not examined. Other studies have shown (Hutchison et al., 2014; Stevens et al., 2015) stronger FC between areas with the same, as compared to different, category-selectivity. Given the mapping of level/shape onto different category-specific regions, this can be seen as congruent with our findings.

Two studies with diffusion-tensor-imaging (DTI) tractography suggest an anatomical basis for the two functional VTC pathways posited here. One study (Gomez et al., 2015) found fiber tracts coursing parallel to each other in a caudal-rostral orientation within VTC. One of the tracts was found within face-selective, and the other within scene-selective areas (corresponding to our level/shape areas). In the other study, probabilistic tractography was used to cluster fusiform gyrus voxels based on their anatomical connectivity with the rest of the brain. The fusiform gyrus was divided into medial, lateral, and anterior regions (Zhang et al., 2016), the first two separated by MFS (also corresponding to our level/shape division).

Only the posterior portions of PPA, OPA and RSC (roughly corresponding to PHC1-2, VO1-2 and V3AB) contain shape-invariant level information, contrariwise to their anterior sections. This limited overlap thus involved a posterior scene-selective sub-network found with FC (Baldassano et al., 2016a; Çukur et al., 2016; Silson et al., 2016). This sub-network, centered on the posterior PPA and OPA and not coupled to the hippocampus, would processes visual features needed to represent spatial layout. An anterior subnetwork containing the anterior

portions of RSC and PPA (and strongly connected to the hippocampus), would play a role nonvisual tasks like memory for scenes, which would be poorly mobilized by our task thus explaining why level information is not detectable there. Furthermore, the posterior subnetwork would be biased toward static stimuli (our case), whereas the anterior network would be biased toward dynamic stimuli (Çukur et al., 2016).

Further work should take into consideration the following limitations of our study. We only included two shapes in the design, as in previous studies of shape decoding (Stokes et al., 2011), to improve the signal-to noise-ratio for estimation of multivariate patterns. The generality of our findings must be tested with a larger range of shapes and stimuli of different retinal sizes (to help answer determine if absolute or relative size contributes to the level decoding). With more diverse stimuli, representational similarity analysis (Kriegeskorte, 2008) could be used to better characterize level/shape representation in higher-order visual areas. Visual field mapping of pRFs (Dumoulin and Wandell, 2008) and level/shape cross-decoding must be performed in the same experiment to adequately test the attention-zoom explanation. Other limitations were that eye-movements were not controlled, and the face-house/contrast localizer was obtained from different subjects than the main experiment. Additionally, other patterns classifiers should be tested, since comparing their accuracy could give clues about differences in coding between areas (as found here by comparing smoothed and centered searchlight maps). Finally, although coding of level plausibly contributes to the representation of visual whole and parts, other sensory attributes probably involved should be examined.

Separate pathways for invariant level and shape have many functional implications. Shapes from different levels of the same object could compete for the same neuronal receptive

fields (Chelazzi, Miller, Duncan, 1993). This is illustrated by a study presenting two fast streams of our modified Navon figures, one global and the other local, at the same visual location (Iglesias-Fuster et al., 2015). Shape identification was more accurate when attention was focused at only one level than when it was split between two. To guide filtering of the irrelevant level, invariant information about level is required in higher-order visual areas. We posit that scene-selective cortex can play this role. Moreover, brief presentation of Navon figure leads to illusory shape/level conjunctions (Hübner and Kruse 2011; Hübner and Volberg 2005; Flevaris, Bentin, and Robertson 2010). Letters from the non-target level are fallaciously seen at the target level This implies that shape/level codes must separate somewhere in the brain (Hübner and Volberg, 2005), perhaps in the two VTC pathways described here. Thus independent coding of level and shape may facilitate sustained attention to a single level, but at the price of risking illusory conjunctions with impoverished attention.

Conclusions

While rejecting H1 and H2, our results support a nuanced version of H3. Consistent with previous work (Cichy et al., 2011; Golomb and Kanwisher, 2012; Hong et al., 2016; Rauschecker et al., 2012; Zopf et al., 2018), information about the non-preferred attribute does not disappear in the shape- and level-preferring pathways. However, a relative reduction of information about the non-preferred attribute was found in each pathway. This suggests computational circuitry specialized for achieving invariance for a primary attribute, geared to specific real-world demands (Peelen and Downing, 2017), that however does not completely discard information about secondary attributes. The fact that shape-invariant level was decodable from scene-selective cortex was an unexpected result. Although Navon figures are clearly visual

objects, they are not only that. If we accept the definition (Epstein and Baker, 2019) that we act ***upon*** objects, but act ***within*** scenes, it is possible to understand the dual nature of Navon figures. Shapes are acted ***on*** to recognize them, but are selected ***within*** the framework of their organization into levels. We postulate that coordinated use of invariant level and shape codes not only helps navigate attention between levels in Navon figures, but possibly within all real-life visual objects possessing wholes and parts.

Acknowledgments

We thank the University of Electronic Science and Technology of China (UESTC) for funding and providing special equipment and facilities for this project. We thank the VLIR-UOS scheme for funds related to the project "A Cuban National School of Neurotechnology for Cognitive Aging". M.V.S., M.O.O., J.I.F., A.C.L., M.B.L. and A.L.C. were also supported by the National Fund for Science and Innovation of Cuba.

Competing interests:

The authors declare that no competing interests exist.

References

Al-aidroos, N., Said, C.P., Turk-browne, N.B., 2012. Top-down attention switches coupling between low- level and high-level areas of human visual cortex 109, 14675–14680. https://doi.org/10.1073/pnas.1202095109

Allefeld, C., Görgen, K., Haynes, J.D., 2016. Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. Neuroimage 141, 378–392. https://doi.org/10.1016/j.neuroimage.2016.07.040

Andrews, T.J., Watson, D.M., Rice, G.E., Hartley, T., 2015. Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway visual pathway. J. Vis. 15, 1–12. https://doi.org/10.1167/15.7.3.doi

Baldassano, C., Esteva, A., Fei-fei, L., Beck, D.M., 2016a. Two Distinct Scene-Processing Networks Connecting Vision and Memory 3.

Baldassano, C., Fei-Fei, L., Beck, D.M., 2016b. Pinpointing the peripheral bias in neural scene processing networks during natural viewing. J. Vis. 16, 9. https://doi.org/10.1167/16.2.9.doi

Berman, D., Golomb, J.D., Walther, D.B., 2017. Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. PLoS One 12, 1–16. https://doi.org/10.1371/journal.pone.0189828

Chang, L., Bao, P., Tsao, D.Y., 2017. The representation of colored objects in macaque color patches. Nat. Commun. 8, 1–14. https://doi.org/10.1038/s41467-017-01912-7

Chelazzi, Leonardo, Earl K. Miller, John Duncan, R.D., 1993. A neural basis for visual search in inferior temporal cortex. Nat. 363, 345–347.

Choo, H., Walther, D.B., 2016. Contour junctions underlie neural representations of scene categories in high-level human visual cortex: Contour junctions underlie neural representations of scenes. Neuroimage 135, 32–44. https://doi.org/10.1016/j.neuroimage.2016.04.021

Christophel, T.B., Allefeld, C., Endisch, C., Haynes, J.-D., 2017. View-Independent Working Memory Representations of Artificial Shapes in Prefrontal and Posterior Regions of the Human Brain. Cereb. Cortex 32, 1–16. https://doi.org/10.1093/cercor/bhx119

Cichy, R.M., Chen, Y., Haynes, J.D., 2011. Encoding the identity and location of objects in human LOC. Neuroimage 54, 2297–2307. https://doi.org/10.1016/j.neuroimage.2010.09.044

Coutanche, M.N., 2013. Distinguishing multi-voxel patterns and mean activation: Why, how, and what does it tell us? Cogn. Affect. Behav. Neurosci. 13, 667–73. https://doi.org/10.3758/s13415-013-0186-2

Çukur, T., Huth, A.G., Nishimoto, S., Gallant, J.L., 2016. Functional subdomains within scene-selective cortex: Parahippocampal place area, retrosplenial complex, and occipital place area. J. Neurosci. 36, 10257–10273. https://doi.org/10.1523/JNEUROSCI.4033-14.2016

Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn. Sci. 15, 254–62. https://doi.org/10.1016/j.tics.2011.04.003

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Dice, L.R.., 1945. Measures of the Amount of Ecologic Association between Species. Ecology 26, 297–302.

Dumoulin, S.O., Wandell, B.A., 2008. Population receptive field estimates in human visual cortex. Neuroimage 39, 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Eger, E., Ashburner, J., Haynes, J.D., Dolan, R.J., Rees, G., 2008a. fMRI activity patterns in human LOC carry information about object exemplars within category. J. Cogn. Neurosci. 20, 356–370. https://doi.org/10.1162/jocn.2008.20019

Eger, E., Kell, C.A., Kleinschmidt, A., 2008b. Graded size sensitivity of object-exemplar-evoked activity patterns within human LOC subregions. J. Neurophysiol. 100, 2038–2047. https://doi.org/10.1152/jn.90305.2008

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392, 598–601.

Epstein, R.A., Baker, C.I., 2019. Scene Perception in the Human Brain. Annu. Rev. Vis. Sci. 5, 1–25. https://doi.org/10.1146/annurev-vision-091718-014809

Flevaris, A. V., Robertson, L.C., 2016. Spatial frequency selection and integration of global and local information in visual processing: A selective review and tribute to Shlomo Bentin. Neuropsychologia 83, 192–200. https://doi.org/10.1016/j.neuropsychologia.2015.10.024

Flevaris, A. V, Bentin, S., Robertson, L.C., 2010. Local or global? Attentional selection of spatial frequencies binds shapes to hierarchical levels. Psychol. Sci. 21, 424–31. https://doi.org/10.1177/0956797609359909

Flevaris, A. V, Martínez, A., Hillyard, S.A., 2014. Attending to global versus local stimulus features modulates neural processing of low versus high spatial frequencies: an analysis with event-related brain potentials. Front. Psychol. 5, 1–11. https://doi.org/10.3389/fpsyg.2014.00277

Flevaris, A. V, Robertson, L.C., 2011. Attentional selection of relative SF mediates global versus local processing : Evidence from EEG 11, 1–12. https://doi.org/10.1167/11.7.11.Introduction

Friston, K.J., 2011. Functional and Effective Connectivity: A Review. Brain Connect. 1, 13–36. https://doi.org/10.1089/brain.2011.0008

Golomb, J.D., Kanwisher, N., 2012. Higher level visual cortex represents retinotopic, not spatiotopic, object location. Cereb. Cortex 22, 2794–810. https://doi.org/10.1093/cercor/bhr357

Gomez, J., Pestilli, F., Witthoft, N., Golarai, G., Liberman, A., Poltoratski, S., Yoon, J., Grill-Spector, K., 2015. Functionally Defined White Matter Reveals Segregated Pathways in Human Ventral Temporal Cortex Associated with Category-Specific Processing. Neuron 85, 216–228. https://doi.org/10.1016/j.neuron.2014.12.027

Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex

and its role in categorization. Nat. Rev. Neurosci. 15, 536–548. https://doi.org/10.1038/nrn3747

Grill-Spector, K., Weiner, K.S., Kay, K., Gomez, J., 2017. The Functional Neuroanatomy of Human Face Perception. Annu. Rev. Vis. Sci. 3, annurev-vision-102016-061214. https://doi.org/10.1146/annurev-vision-102016-061214

Groen, I.I.A., Silson, E.H., Baker, C.I., 2017. Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. Philos. Trans. R. Soc. B Biol. Sci. 372, 20160102. https://doi.org/10.1098/rstb.2016.0102

Haak, K. V., Beckmann, C.F., 2016. Objective analysis of the topological organization of the human cortical visual connectome suggests three visual pathways. Cortex 1–11. https://doi.org/10.1016/j.cortex.2017.03.020

Han, S., Weaver, J.A., Murray, S.O., Kang, X., Yund, E.W., Woods, D.L., 2002. Hemispheric asymmetry in global/local processing: Effects of stimulus position and spatial frequency. Neuroimage 17, 1290–1299. https://doi.org/10.1006/nimg.2002.1255

Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas 34, 479–490.

Henriksson, L., Karvonen, J., Salminen-Vaparanta, N., Railo, H., Vanni, S., 2012. Retinotopic maps, spatial tuning, and locations of human visual areas in surface coordinates characterized with multifocal and blocked FMRI designs. PLoS One 7, e36859. https://doi.org/10.1371/journal.pone.0036859

Henriksson, L., Nurminen, L., Vanni, S., 2008. Spatial frequency tuning in human retinotopic visual areas 8, 1–13. https://doi.org/10.1167/8.10.5.Introduction

Hong, H., Yamins, D.L.K., Majaj, N.J., DiCarlo, J.J., 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. Nat. Neurosci. 19, 613–622. https://doi.org/10.1038/nn.4247

Hübner, R., Kruse, R., 2011. Effects of stimulus type and level repetition on content-level binding in global/local processing. Front. Psychol. 2, 134. https://doi.org/10.3389/fpsyg.2011.00134

Hübner, R., Volberg, G., 2005a. The integration of object levels and their content: a theory of global/local processing and related hemispheric differences. J. Exp. Psychol. Hum. Percept. Perform. 31, 520–41. https://doi.org/10.1037/0096-1523.31.3.520

Hübner, R., Volberg, G., 2005b. The integration of object levels and their content: a theory of global/local processing and related hemispheric differences. J. Exp. Psychol. Hum. Percept. Perform. 31, 520–541. https://doi.org/10.1037/0096-1523.31.3.520

Hutchison, R.M., Culham, J.C., Everling, S., Flanagan, J.R., Gallivan, J.P., 2014. Distinct and distributed functional connectivity patterns across cortex reflect the domain-specific constraints of object, face, scene, body, and tool category-selective modules in the ventral visual pathway. Neuroimage 96, 216–236. https://doi.org/10.1016/j.neuroimage.2014.03.068

Iglesias-Fuster, J., Santos-Rodríguez, Y., Trujillo-Barreto, N., Valdés-Sosa, M.J., 2015.

Asynchronous presentation of global and local information reveals effects of attention on brain electrical activity specific to each level. Front. Psychol. 6, 1–14. https://doi.org/10.3389/fpsyg.2015.00570

Jacobson, M.W., Delis, D.C., Lansing, A., Houston, W., Olsen, R., Wetter, S., Bondi, M.W., Salmon, D.P., 2005. Asymmetries in global-local processing ability in elderly people with the apolipoprotein e-epsilon4 allele. Neuropsychology 19, 822–829. https://doi.org/10.1037/0894-4105.19.6.822

Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. Proc. Natl. Acad. Sci. U. S. A. 107, 11163–70. https://doi.org/10.1073/pnas.1005062107

Kaplan, J.T., Man, K., Greening, S.G., 2015. Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. Front. Hum. Neurosci. 9, 1–12. https://doi.org/10.3389/fnhum.2015.00151

Kay, K.N., Weiner, K.S., Grill-Spector, K., 2015. Attention reduces spatial uncertainty in human ventral temporal cortex. Curr. Biol. 25, 595–600. https://doi.org/10.1016/j.cub.2014.12.050

Kimchi, R., 2015. The perception of hierarchical structure. Oxford Handb. Percept. Organ. 129–149. https://doi.org/10.1093/oxfordhb/9780199686858.013.025

Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital complex. Science (80-.). 293, 1506–1509. https://doi.org/10.1126/science.1061133

Kourtzi, Z., Kanwisher, N., 2000. Cortical Regions Involved in Perceiving Object Shape 20, 3310–3318.

Kriegeskorte, N., 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–8. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4. https://doi.org/10.3389/neuro.06.004.2008

Lancichinetti, A., Fortunato, S., 2012. Consensus clustering in complex networks. Sci. Rep. 2. https://doi.org/10.1038/srep00336

Lashkari, D., Vul, E., Kanwisher, N., Golland, P., 2010. NeuroImage Discovering structure in the space of fMRI selectivity profiles. Neuroimage 50, 1085–1098. https://doi.org/10.1016/j.neuroimage.2009.12.106

Leibo, J.Z., Liao, Q., Anselmi, F., Poggio, T., 2015. The Invariance Hypothesis Implies Domain-Specific Regions in Visual Cortex. PLoS Comput. Biol. 11, 1–29. https://doi.org/10.1371/journal.pcbi.1004390

López, K., Torres, R., Valdés, M., 2002. Medición directa del tiempo de tránsito atencional entre distintos niveles de figuras jerárquicas : observadores normales y autistas . Rev. CENIC

Ciencias Biológicas 33, 111–117.

Malcolm, G.L., Groen, I.I.A., Baker, C.I., 2016. Making Sense of Real-World Scenes. Trends Cogn. Sci. 20, 843–856. https://doi.org/10.1016/j.tics.2016.09.003

Moeller, S., Freiwald, W. a, Tsao, D.Y., 2008. Patches with links: a unified system for processing faces in the macaque temporal lobe. Science 320, 1355–9. https://doi.org/10.1126/science.1157436

Nasr, S., Echavarria, C.E., Tootell, R.B.H., 2014. Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. J. Neurosci. 34, 6721–6735. https://doi.org/10.1523/JNEUROSCI.4802-13.2014

Navon, D., 2003. What does a compound letter tell the psychologist's mind? Acta Psychol. (Amst). 114, 273–309. https://doi.org/10.1016/j.actpsy.2003.06.002

Ontivero-Ortega, M., Lage-Castellanos, A., Valente, G., Goebel, R., Valdes-Sosa, M., 2017. Fast Gaussian Naïve Bayes for searchlight classification analysis. Neuroimage 1–9. https://doi.org/10.1016/j.neuroimage.2017.09.001

Osher, D.E., Saxe, R.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. Cereb. Cortex 26, 1668–1683. https://doi.org/10.1093/cercor/bhu303

Park, S., Konkle, T., Oliva, A., 2015. Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain. Cereb. Cortex 25, 1792–1805.

https://doi.org/10.1093/cercor/bht418

Peelen, M. V., Downing, P.E., 2017. Category selectivity in human visual cortex: Beyond visual object recognition. Neuropsychologia 105, 177–183. https://doi.org/10.1016/j.neuropsychologia.2017.03.033

Rauschecker, A.M., Bowen, R.F., Parvizi, J., Wandell, B. a, 2012. Position sensitivity in the visual word form area. Proc. Natl. Acad. Sci. U. S. A. 109, E1568-77. https://doi.org/10.1073/pnas.1121304109

Rijpkema, M., Aalderen, S.I. Van, Schwarzbach, J. V, Verstraten, F.A.J., 2007. Activation patterns in visual cortex reveal receptive field size-dependent attentional modulation 89. https://doi.org/10.1016/j.brainres.2007.10.100

Rust, N.C., DiCarlo, J.J., 2010. Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT. J. Neurosci. 30, 12978–12995. https://doi.org/10.1523/JNEUROSCI.0179-10.2010

Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, A.K., Marret, S., Dale, A.M., Tootell, R.B.H., 2001. Local and global attention are mapped retinotopically in human occipital cortex. Proc. Natl. Acad. Sci. 98, 2077–2082. https://doi.org/10.1073/pnas.98.4.2077

Saygin, Z.M., Osher, D.E., Norton, E.S., Youssoufian, D.A., Beach, S.D., Feather, J., Gaab, N., Gabrieli, J.D.E., Kanwisher, N., 2016. Connectivity precedes function in the development of the visual word form area. Nat. Neurosci. 19, 1250–1255. https://doi.org/10.1038/nn.4354

Scholte, H.S., Losch, M.M., Ramakrishnan, K., de Haan, E.H.F., Bohte, S.M., 2018. Visual pathways from the perspective of cost functions and multi-task deep neural networks. Cortex 98, 249–261. https://doi.org/10.1016/j.cortex.2017.09.019

Schwartzman, A., Dougherty, R.F., Lee, J., Ghahremani, D., Taylor, J.E., 2009. Empirical null and false discovery rate analysis in neuroimaging. Neuroimage 44, 71–82. https://doi.org/10.1016/j.neuroimage.2008.04.182

Silson, Edward H, Groen, I.I.A., Kravitz, D.J., Baker, C.I., 2016. Evaluating the correspondence between face- , scene- , and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. J. Vis. 16, 1–21. https://doi.org/10.1167/16.6.14

Silson, Edward H., Steel, A.D., Baker, C.I., 2016. Scene-Selectivity and Retinotopy in Medial Parietal Cortex. Front. Hum. Neurosci. 10, 1–17. https://doi.org/10.3389/fnhum.2016.00412

Slavin, M.J., Mattingley, J.B., Bradshaw, J.L., Storey, E., 2002. Local – global processing in Alzheimer ' s disease : an examination of interference , inhibition and priming. Neuropsychologia 40, 1173–1186.

Snodgrass, J.G., Vanderwart, M., 1980. A Standardized Set of 260 Pictures : Norms for Name Agreement , Image Agreement , Familiarity , and Visual Complexity. J. Exp. Psychol. Hum. Learn. Mem. 6, 174–215.

Stevens, D.W., Tessler, M.H., Peng, C.S., Martin, A., 2015. Functional connectivity constrains the category-related organization of human ventral occipitotemporal cortex. Hum. Brain Mapp. 36, 2187–2206. https://doi.org/10.1002/hbm.22764

Stokes, M., Saraiva, A., Rohenkohl, G., Nobre, A.C., 2011. NeuroImage Imagery for shapes activates position-invariant representations in human visual cortex. Neuroimage. https://doi.org/10.1016/j.neuroimage.2011.02.071

Tacchetti, A., Isik, L., Poggio, T.A., 2018. Invariant Recognition Shapes Neural Representations of Visual Input. Annu. Rev. Vis. Sci. 4, annurev-vision-091517-034103. https://doi.org/10.1146/annurev-vision-091517-034103

Valdés-Sosa, M.J., Iglesias-Fuster, J., Torres, R., 2014. Attentional selection of levels within hierarchically organized figures is mediated by object-files. Front. Integr. Neurosci. 8, 91. https://doi.org/10.3389/fnint.2014.00091

Von Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput. 17, 395–416. https://doi.org/10.1007/s11222-007-9033-z

Wang, L., Mruczek, R.E.B., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25, 3911–3931. https://doi.org/10.1093/cercor/bhu277

Weiner, K.S., 2018. The mid-fusiform sulcus (sulcus sagittalis gyri fusiformis). Anat. Rec. 1–36. https://doi.org/10.1002/ar.24041

Weissman, D.H., Woldorff, M.G., 2005. Hemispheric asymmetries for different components of global/local attention occur in distinct temporo-parietal loci. Cereb. Cortex 15, 870–6. https://doi.org/10.1093/cercor/bhh187

White, S., O'Reilly, H., Frith, U., 2009. Big heads, small details and autism. Neuropsychologia

47, 1274–81. https://doi.org/10.1016/j.neuropsychologia.2009.01.012

Williams, M.A., Baker, C.I., Op de Beeck, H.P., Mok Shim, W., Dang, S., Triantafyllou, C., Kanwisher, N., 2008. Feedback of visual object information to foveal retinotopic cortex. Nat. Neurosci. 11, 1439–1445. https://doi.org/10.1038/nn.2218

Zhang, W., Wang, J., Fan, L., Zhang, Y., Fox, P.T., Eickhoff, S.B., Yu, C., Jiang, T., 2016. Functional organization of the fusiform gyrus revealed with connectivity profiles. Hum. Brain Mapp. 37, 3003–3016. https://doi.org/10.1002/hbm.23222

Zhen, Z., Kong, X.Z., Huang, L., Yang, Z., Wang, X., Hao, X., Huang, T., Song, Y., Liu, J., 2017. Quantifying the variability of scene-selective regions: Interindividual, interhemispheric, and sex differences. Hum. Brain Mapp. 38, 2260–2275. https://doi.org/10.1002/hbm.23519

Zopf, R., Butko, M., Woolgar, A., Williams, M.A., Rich, A.N., 2018. Representing the location of manipulable objects in shape-selective occipitotemporal cortex: Beyond retinotopic reference frames? Cortex 106, 132–150. https://doi.org/10.1016/j.cortex.2018.05.009

Supplementary material

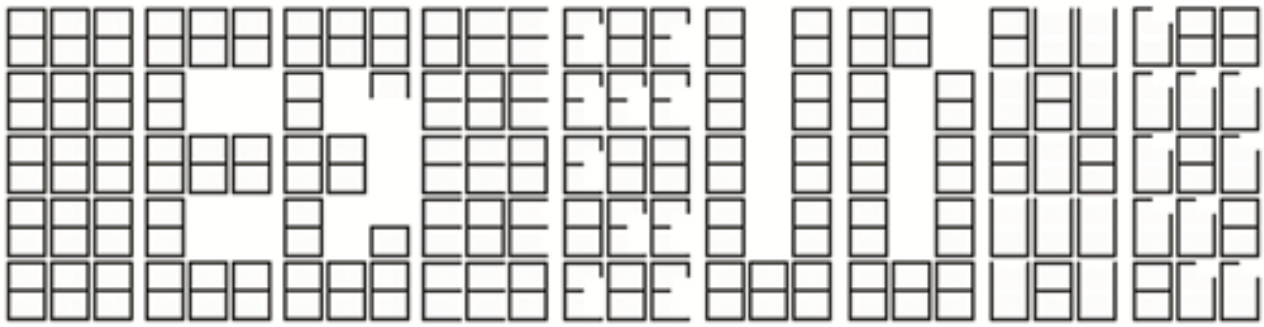

Figure S1. All stimuli used in the main fMRI experiment.

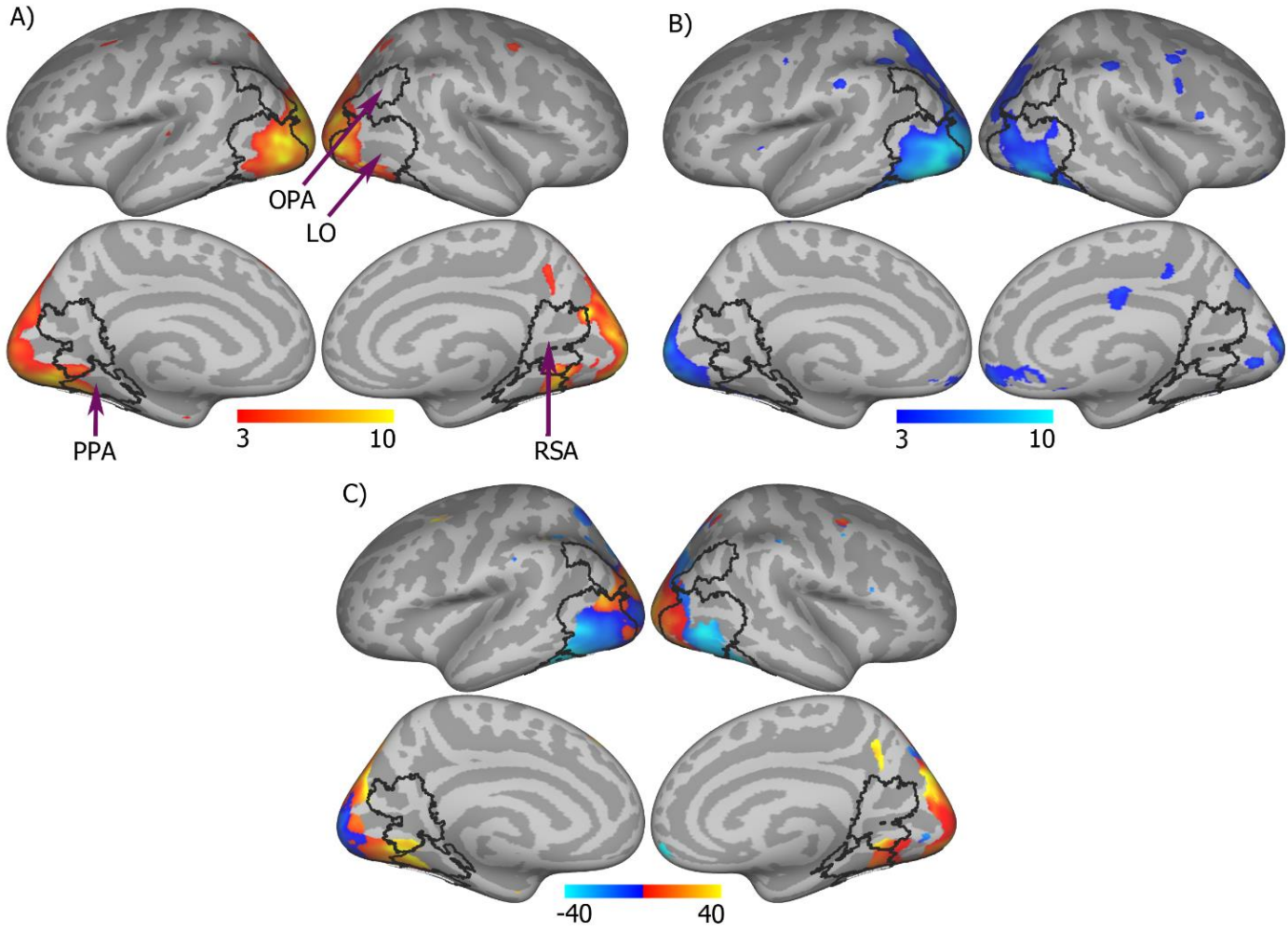

Figure S2. Additional views of searchlight maps for original data: A) Shape-invariant level map. B) Level-invariant shape map. C) IOS map.

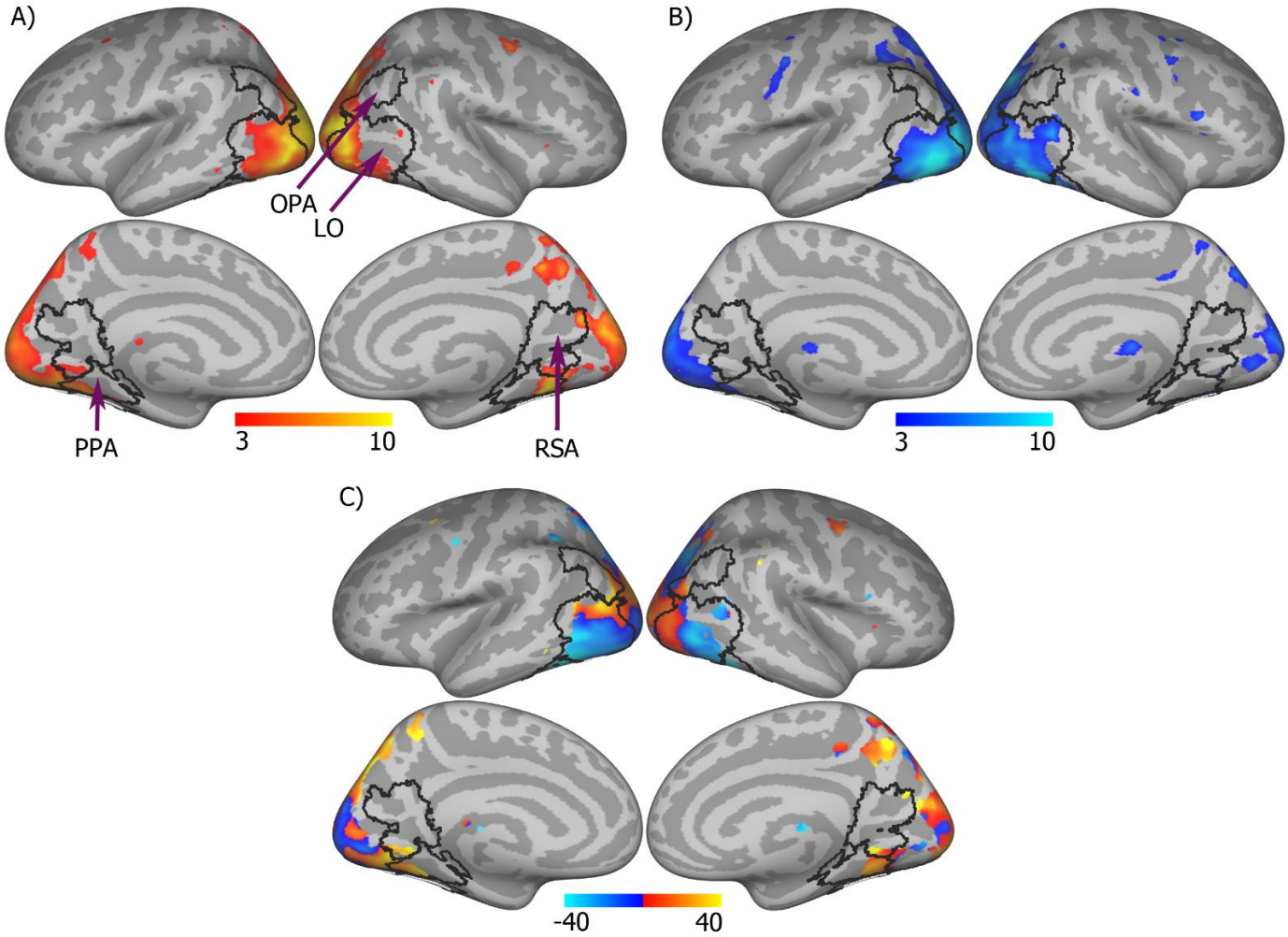

Figure S3. Additional views of searchlight maps for centered data: A) Shape-invariant level map. B) Level-invariant shape map. C) IOS map.

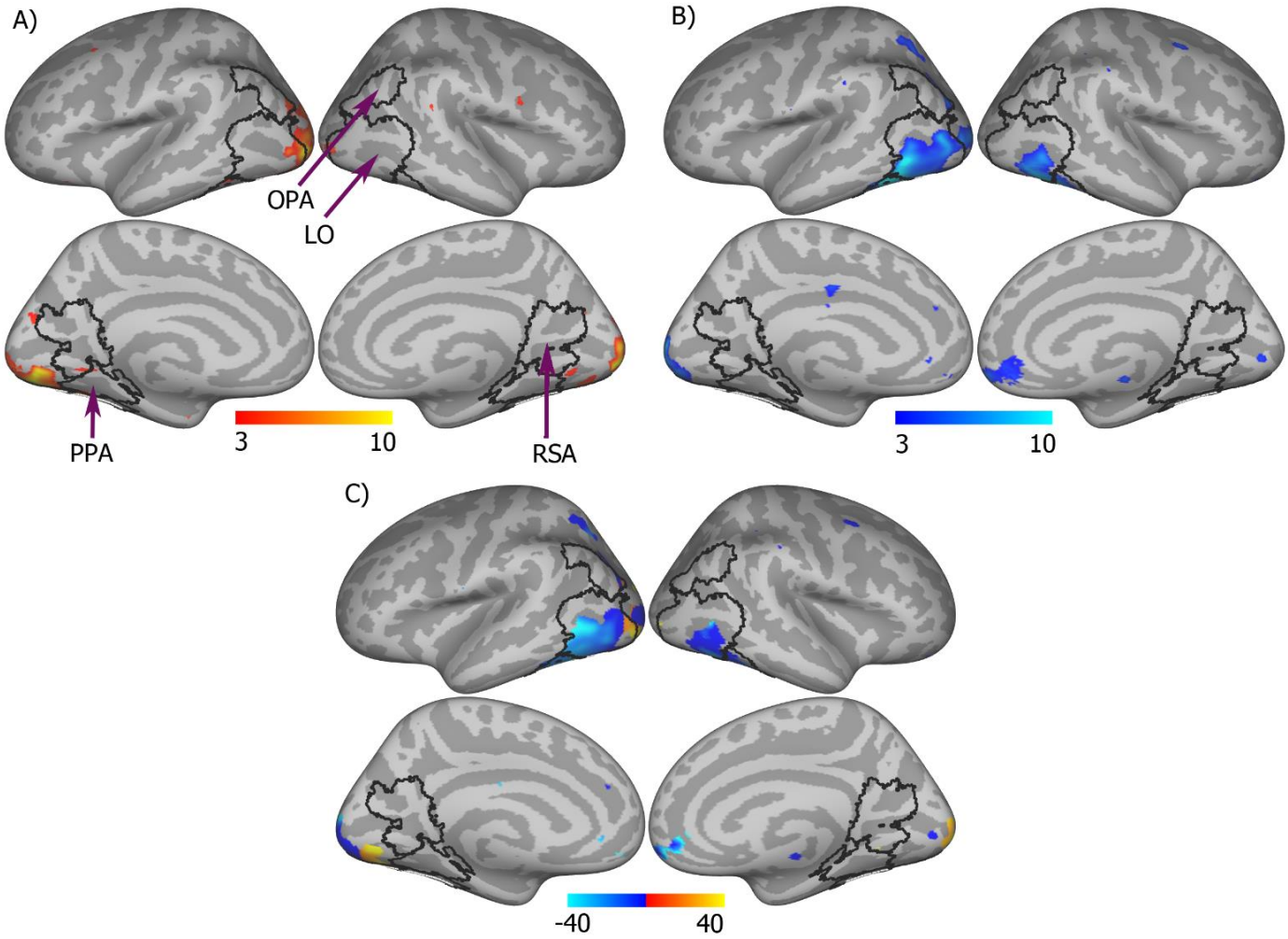

Figure S4. Additional views of searchlight maps for smoothed data: A) Shape-invariant level map. B) Level-invariant shape map. C) IOS map.

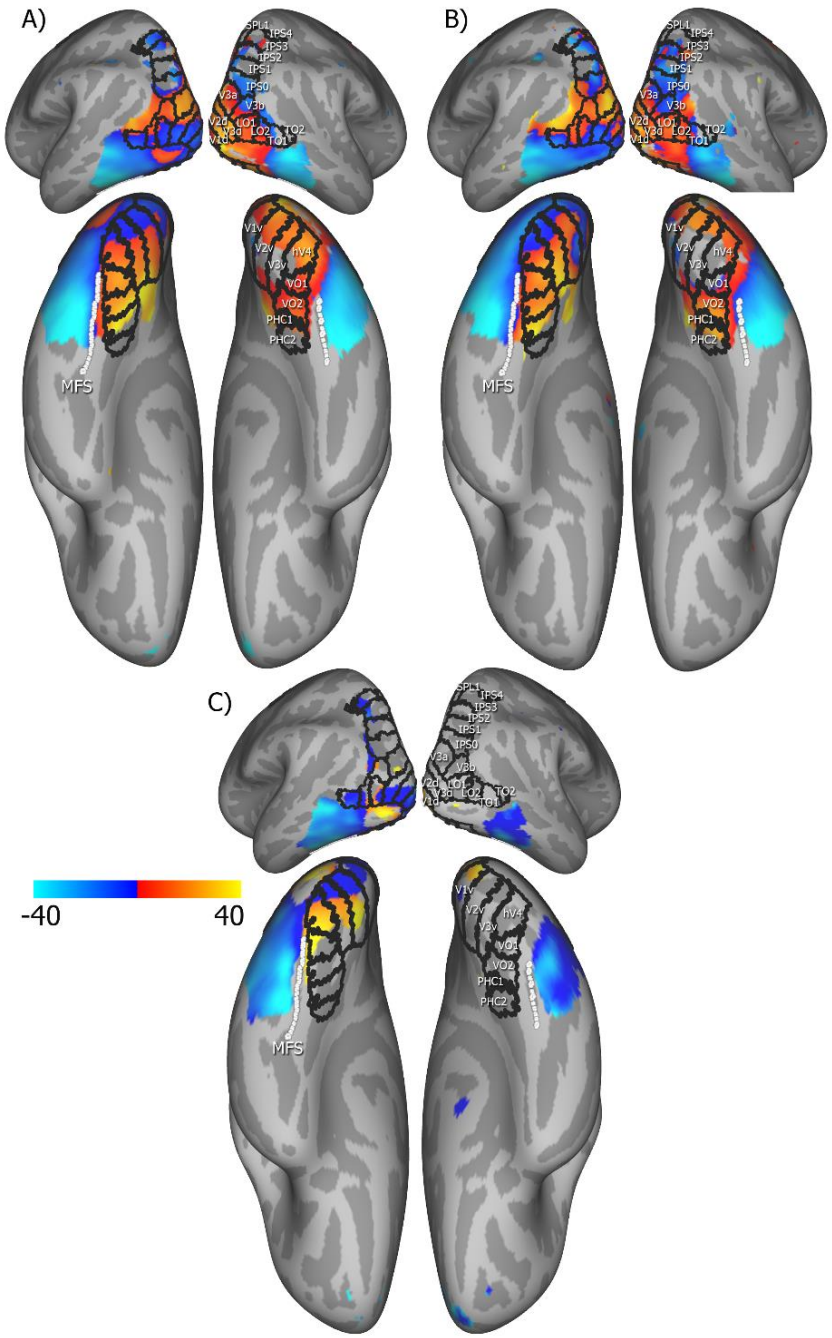

Figure S5. IOS map with superimposed borders of retinotopic areas (Wang atlas). A) Original map. B) Centered map. C) Smoothed map.

Table S1. Percentage of informative searchlight centers in each fROI. Percentages larger than 20% are shown in bold.

fROIs	**Percentage (LH)**				**Percentage (RH)**			
	Houses	**Level**	**Shape**	**Faces**	**Houses**	**Level**	**Shape**	**Faces**
PPA	**58**	19	0	0	**76**	19	0	0
RSC	0	9	0	0	15	6	0	0
OPA	19	4	0	0	19	0	1	0
LO	8	5	**36**	0	9	19	**23**	8
pFus	**21**	2	**41**	**32**	10	0	**33**	**47**
V1v	1	1	2	0	0	19	4	0
V1d	0	6	30	0	0	**36**	0	0
V2v	1	**27**	0	0	0	**43**	0	0
V2d	2	0	14	0	0	**92**	0	0
V3v	5	**40**	0	0	7	**42**	0	0
V3d	13	**23**	9	0	9	**93**	0	0
hV4	19	**24**	0	0	9	**78**	0	0
VO1	**73**	**93**	0	0	**55**	**46**	0	0
VO2	**91**	**98**	0	0	**75**	**74**	0	0
PHC1	**97**	**54**	0	0	**100**	**87**	0	0
PHC2	**100**	4	0	0	**100**	17	0	0
MST	0	0	0	0	0	0	0	0
hMT	3	1	16	0	9	0	2	14
LO2	**89**	0	0	0	**81**	**23**	0	0
LO1	**78**	0	0	0	**94**	**71**	0	0
V3b	**100**	17	0	0	**100**	12	8	0
V3a	13	**96**	0	0	15	**52**	0	0
IPS0	**28**	1	10	0	**52**	0	**23**	0
IPS1	0	1	11	0	4	0	11	0
IPS2	0	7	12	0	0	0	0	0
IPS3	0	0	**43**	0	0	1	0	0
IPS4	0	0	0	0	0	0	0	0
IPS5	0	0	0	0	0	0	0	0
SPL1	0	0	2	0	0	0	0	0
FEF	0	0	0	0	0	0	0	0

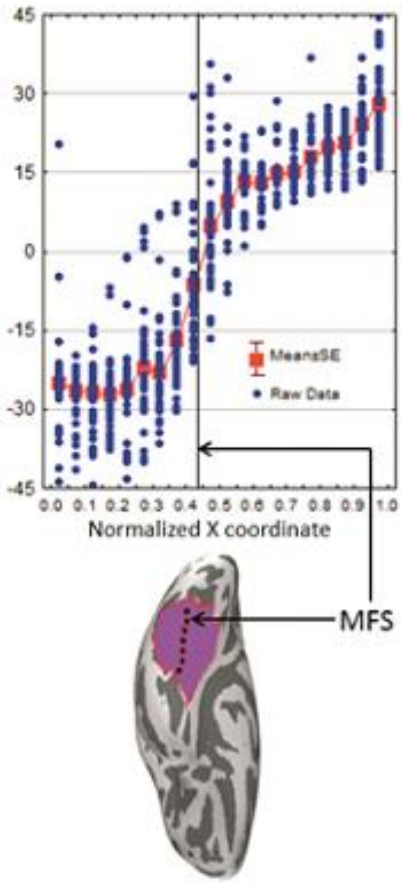

Figure S6. IOS values for each surface node in VTC as a function of its normalized laterality along the MFS. The distance between the lateral and medial borders of VTC shown in the inset were measured and normalized to 1. The IOS values of each node was then plotted as a function of this normalized value. The approximate position of MFS is indicated by arrows