Title: Shape coding in occipito-temporal cortex relies on object silhouette, curvature and medial-axis

Authors and affiliations: Paolo Papale[1,2,¶], Andrea Leo[1,3,¶], Giacomo Handjaras[1], Luca Cecchetti[1], Pietro Pietrini[1] and Emiliano Ricciardi[1, *]

1. *Molecular Mind Lab, IMT School for Advanced Studies Lucca, 55100 Lucca, Italy*
2. *Department of Vision & Cognition, Netherlands Institute for Neuroscience (KNAW), 1105 BA Amsterdam, Netherlands*
3. *Department of translational research and advanced technologies in medicine and surgery, University of Pisa, 52100 Pisa, Italy*

¶. *These authors contributed equally*

Running head: Shape coding in natural vision

***Corresponding author**: Emiliano Ricciardi: *emiliano.ricciardi@imtlucca.it*, Piazza San Francesco 19, 55100 Lucca, Italy

Abstract

Object recognition relies on different transformations of the retinal input, ranging from local contrast to object shape and category. While some of those transformations are thought to occur at specific stages of the visual hierarchy, the features they represent are correlated (e.g., object shape and identity) and the selectivity for the same feature can be found across many brain regions. This overlap may be explained either by collinearity across representations, or may instead reflect the coding of multiple dimensions by the same cortical population. Moreover, orthogonal and shared components may differently impact on distinctive stages of the visual hierarchy. We recorded functional MRI (fMRI) activity while participants passively attended to object images and employed a statistical approach that partitioned orthogonal and shared object representations to reveal their relative impact on brain processing. Orthogonal shape representations (silhouette, curvature and medial-axis) independently explained distinct and overlapping clusters of selectivity in occitotemporal (OTC) and parietal cortex. Moreover, we show that the relevance of shared representations linearly increases moving from posterior to anterior regions. These results indicate that the visual cortex encodes shared relations between different features in a topographic fashion and that object shape is encoded along different dimensions, each representing orthogonal features.

New & Noteworthy

There are several possible ways of characterizing the shape of an object. Which shape description is encoded in our brain and thus really mirrors the way we perceive shapes? Here, we employed three competing shape models to explain brain representations when viewing real objects. We found that object shape is encoded in a multi-dimensional fashion and thus defined by the interaction of multiple features.

Introduction

Objects are typically defined by their shapes. However, from a theoretical perspective, the concept of shape may appear quite elusive, since the shape of an object could be defined in several possible ways. Consequently, many computational models of object shape could be constructed (Figure 1A): closed shapes can be easily and reliably generated by combining simple elements (e.g., *geons* or medial axes), or by connecting few salient points with acute curvature, or by modulating its radial frequency. Hence, which of these different descriptions is encoded in our brain and truly reflects the way we perceive and represent shapes? Here, we show that shape coding in the brain relies on multiple dimensions.

Indeed, all these different representations are sensitive to specific aspects of a shape and produce different ways of encoding the same object. Consider Figure 1B. A silhouette descriptor (Figure 1A, red), consisting of a simple stimulus vectorization, would vary mostly depending on the global layout of a shape, but would be quite insensitive to small perturbations of the outline. Conversely, the curvature descriptor from Figure 1A (blue) would be unaffected by rotations or large transformation of a shape, but would be highly sensitive to the number of acute points on the outline. Of note, these different representations can either diverge independently to each other, as in the horizontal and vertical shapes, or can covary together, as for the shapes placed on the diagonal. Evidence from previous studies suggests that visual dimensions in natural vision are indeed highly correlated (Figure 1C-E; Bracci and Op de Beeck 2016; Kay 2011; Papale et al. 2019). Thus, addressing the extent to which brain regions represent different dimensions has so far proven challenging: how can we disentangle the role of different shape properties if they likely covary together?

To answer these questions, we recorded functional MRI (fMRI) activity while participants passively attended to object pictures. We employed a statistical approach that partitions

63 orthogonal and shared shape representations and reveals their relative impact on brain
64 processing, while controlling, at the same time, for low- and high-level features (Figure 1;
65 Lescroart et al. 2015). Since a recent behavioral work demonstrates that multiple dimensions are
66 necessary to model behavioral shape similarity (Morgenstern et al. 2020), we expected multiple
67 shape properties to be independently represented in the human visual cortex. At the same time,
68 we explored the way in which the brain exploits shared information between different visual
69 dimensions. In line with recent neurophysiological work (Hong et al. 2016), we expected to
70 observe a linear decrease of the role of orthogonal representations along the visual hierarchy.

71 In our approach, we tested three competing shape models. A first description was
72 computed by extracting the silhouette. The link between shape silhouette and representations in
73 the occipito-temporal cortex (OTC) has been extensively investigated in neuroimaging studies
74 (Bracci and Op de Beeck 2016; Kaiser et al. 2016; Khaligh-Razavi and Kriegeskorte 2014; Proklova
75 et al. 2016). Second, a skeletal representation of each stimulus was extracted by performing the
76 medial axis transform (Blum 1973). The spike rate of inferotemporal (IT) neurons in monkey are
77 sensitive to the medial-axis of objects (Hung et al. 2012), which also captures behavioral ratings of
78 shape similarity (Ayzenberg et al. 2019a; Ayzenberg and Lourenco 2019; Lowet et al. 2018) and
79 whose spatiotemporal association with brain activity in humans has been described in several
80 neuroimaging studies (Ayzenberg et al. 2019b; Handjaras et al. 2017; Leeds et al. 2013; Lescroart
81 and Biederman 2013; Papale et al. 2019). A third description was obtained by computing the
82 curvature distribution for each object contour. It has been showed that V4 neurons in monkey are
83 selective to a specific degree of curvature (Cadieu et al. 2007; Carlson et al. 2011; Connor et al.
84 2007). Moreover, the pivotal role of contour curvature in object perception has been extensively
85 demonstrated both by behavioral (Elder and Velisavljevic 2009; Lawrence et al. 2016; Long et al.
86 2017; Wolfe et al. 1992) and neuroimaging studies (Caldara et al. 2006; Long et al. 2018; Vernon et

al. 2016; Yue et al. 2014). In addition, we introduced two additional models: the inked area (in pixels) of each stimulus as a low-level representation, and object identity, to account high-level information (Khaligh-Razavi and Kriegeskorte 2014; Kriegeskorte et al. 2008). Of note, all these models differ in their complexity: the silhouette and inked area return a low-level description of which parts and what extent of the visual field (and thus of the retinotopic cortex) are covered by the picture of an object, respectively. Instead, medial-axis, curvature and identity provide high-level, abstract representations, insensitive to view-point and local position.

As expected, we found both distinct and overlapping clusters of selectivity in OTC and in parietal regions independently explained by different shape representations (i.e., silhouette, curvature and medial-axis: Figures 2-4). Moreover, we showed that, while the prominence of retinotopic processing on abstract information shifts abruptly moving from the occipital to the temporal cortex, shared representations linearly increase from posterior to anterior regions along the visual hierarchy (Figure 5).

** Figure 1 near here **

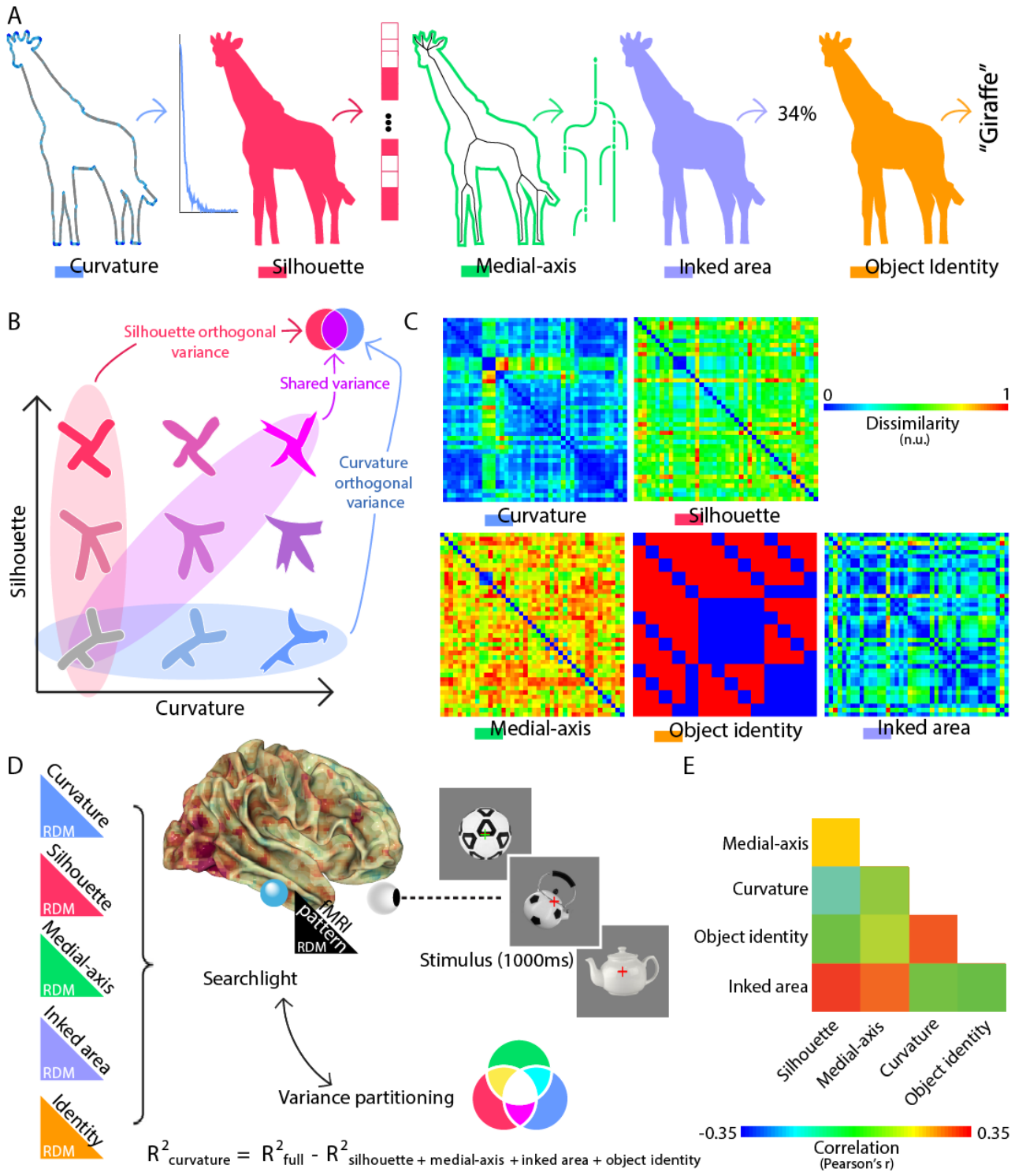


Figure 1. Schematic of the shape models and experiment.
A) Five different object representations are employed: three shape models and two further descriptions. From left: silhouette, medial axis, curvature, inked area (low-level) and object identity (high-level).
B) Different features capture specific aspects of object shape. For instance, silhouette and curvature descriptions of the same shapes may be orthogonal to each other (red- and blue-shaded areas) or vary in a linear fashion (purple-shaded area). Thus, our brain may represent object shape by extracting one specific and more reliable feature, by focusing on shared representations across multiple features, or even encoding the orthogonal components of different features.

C) Representational dissimilarity matrices (RDMs) of each model: they represent all the possible pairwise distances between the stimuli.
D) Methodological pipeline. Brain responses were recorded while subjects maintained fixation on a colored fixation cross, paying attention to color switching between red and green. Orthogonal to the task, we presented 42 grayscale pictures of real objects, for 1s each. Activity patterns were used to test the association between the five model RDMs and each brain activity RDM, computed combining a searchlight procedure with a variance partitioning analysis.
E) Similarity between the five model RDMs. As expected, the five representations are correlated. However, the variance partitioning approach control for the effect of model collinearity.

Methods

Subjects

Seventeen subjects were enrolled for the study. Two subjects participated as pilot subjects with a different version of the experimental protocol and their data were not used for the subsequent analyses; data from one subject who abruptly terminated the experiment were discarded. Fourteen subjects were further considered. The final sample comprised six females, age was 24 ± 3 years, all subjects were right-handed with normal or corrected-to-normal vision and were recruited among the students at the University of Pisa, Italy. Signed informed consent was acquired from all subjects and all the experimental procedures were performed according to the Declaration of Helsinki, under a protocol (1616/2003) approved by the Ethical Committee at the University of Pisa, Italy.

Task

For this study, an event-related design was adopted. Stimuli consisted of 42 static images of grayscale unfamiliar and common objects, presented against a fixed gray background, with a superimposed fixation cross (size: 2x2°), followed by a baseline condition characterized by a gray screen with a red fixation cross (Figure 1D).

A set of stimuli was selected, consisting of 24 common (animate and inanimate) and 18 unfamiliar objects (500x500 pixels – 20°x20°). The latter group represented existing objects that

combine the function and the shape of two of the common objects (e.g., a fish-shaped teapot). Of note, a similar criterion has been employed for stimuli selection also in a recent study (Bracci et al. 2019). To build the final set of stimuli, pictures of existing objects were found on Internet, resized, normalized for luminance and root-mean-square contrast.

Stimuli were presented with the Presentation software (Neurobehavioral Systems, Albany, CA, USA) on MR-compatible goggles (VisuaStim, Resonance Technology Inc., CA, USA), with a LCD at the resolution of 800x600 pixels (32°x24°). The study was organized in six runs, comprising 56 trials (in total: 8 repetitions for each stimulus) which consisted of 1000ms of stimulus presentation and 7000ms of inter-stimulus interval; each run started and ended with 15 seconds of rest, to estimate baseline levels of BOLD signal, and lasted 7:30 minutes. The total duration of the experiment, including anatomical scans, was about 55 minutes.

During the functional runs, subjects were asked to fixate the cross at the center of the screen. On selected trials, the cross changed its color from red to green, and subjects were asked to detect such changes by pressing a key on a MR-compatible keyboard with the index finger of their dominant hand. Order of trials was randomized across runs, and a different randomization schema was used for each participant.

Functional MRI data acquisition

Data were acquired with a 3-Tesla GE Signa scanner (General Electric Inc., Milwaukee, WI, USA) equipped with an 8-channel phased-array coil. For functional images, a gradient-echo echo-planar imaging sequence (GE-EPI) was used, with TE = 40ms, TR = 2500ms, FA = 90°, 184 volumes, acquisition time 7’40”, including four additional dummy scans; image geometry parameters were: Field-Of-View 258x258mm, 128x128 in-plane matrix, voxel size 2.03x2.03x4mm, 37 axial slices for total brain coverage (z-axis extent = 148mm). To acquire detailed information of subject anatomy,

a 3D Fast Spoiled Gradient Echo T1-weighted sequence was also acquired (TE = 3.18ms, TR = 8.16ms, FA = 12°, Field-Of-View 256x256mm, 256x256 matrix size, $1mm^3$ isotropic voxels, 256 axial slices, z-axis extent 256mm).

Functional MRI data processing

Data preprocessing was carried out with AFNI (Cox 1996) and FSL 5.0 (Jenkinson et al. 2012). Preprocessing of functional data comprised slice timing correction with Fourier method (*3dtshift*), rigid-body motion correction using the first volume of the third run as reference (*3dvolreg*), spike removal (elimination of outliers in the functional time series, *3dDespike*), smoothing with a Gaussian filter (fixed FWHM 4 mm, *3dmerge*), and scaling of BOLD time series to percentage of the mean of each run (*3dTstat*, *3dcalc*). Processing of anatomical images consisted of brain extraction (*bet*), segmentation for bias-field estimation and removal (*FAST*, *fslmaths*), linear (*FLIRT*) and nonlinear registration (*FNIRT*) to MNI152 standard space.

For each subject, data from the six concatenated runs (960 time points) were used for a GLM analysis (*3dDeconvolve*) with the responses for each stimulus – modeled with 1 seconds-long block functions convolved with a canonical HRF – as predictors of interest, and the six motion parameters plus polynomial trends up to 4th order as predictors of no-interest.

Responses for individual stimuli were converted to MNI152 space by applying the transformation matrices estimated as explained above, and resampled to a resolution of 2x2x2mm.

Computational models

Five different representations of the 42 stimuli were developed: three shape-based descriptions of interest and two further models. For each model, we obtained a stimulus-specific

feature space, and pairwise dissimilarities between stimuli were computed to obtain a representational dissimilarity matrix (RDM). Before computing shape-related information, stimuli were binarized.

For the silhouette model, pairwise dissimilarity was computed using correlation distance (1 – Pearson's rho). For the medial-axis model, pairwise distance between skeletal representations was computed using the ShapeMatcher algorithm (http://www.cs.toronto.edu/~dmac/ShapeMatcher/; (Van Eede et al. 2006)). In sum, the ShapeMatcher algorithm builds the shock-graphs of each shape and then estimates their dissimilarity as the deformation required to match different objects (Sebastian et al. 2004). Curvature was computed as the chord-to-point distance (Monroy et al. 2011) in a 40-pixels window. Pairwise dissimilarity was computed using correlation distance between the histograms of curvature from each pair of stimuli. Finally, two further RDMs were built. For the inked-area, pairwise dissimilarity was computed as the absolute difference between the number of pixels covered by different objects. For identity, a binary representation was employed (Khaligh-Razavi and Kriegeskorte 2014; Kriegeskorte et al. 2008). Unfamiliar stimuli were considered as belonging to categories according to both their function and shape.

Shape selectivity

A variance partitioning analysis (Lescroart et al. 2015) was performed to determine whether the three shape models in this study significantly explain unique components of the variance of brain representations (computed using Pearson's correlation distance), as computed in 6 mm-radius spherical searchlights (Kriegeskorte et al. 2006). To this aim, explained variance coefficient (R^2) was computed for each model RDM in independent linear regressions, and then all the different combinations of models were tested in further multiple linear regressions. The final

statistic reporting the partial goodness of fit for unique and shared components was computed following the work by Nimon and colleagues (2008). For example, the unique variance explained by the curvature model in a specific searchlight was determined as the difference between the full-model R^2 and the variance explained by the combination of all other models (i.e., $R^2_{curvature} = R^2_{full} - R^2_{silhouette + medial\text{-}axis + inked\ area + identity}$). In the context of multiple linear regression, this approach is better known as 'commonality analysis' (Nimon and Oswald 2013), and its popularity is growing in neuroimaging (de Heer et al. 2017; Groen et al. 2018; Lescroart et al. 2015).

1 – r correlation distance was used to compute the RDM of fMRI activity patterns in each searchlight. Only voxels pertaining to the cerebral cortex with a probability higher than 50% were included in the procedure (i.e., by applying a threshold over the Harvard-Oxford cortical probabilistic atlas). The z-scored partial correlation coefficient (de Heer et al. 2017) for each component of unique and shared variance were then assigned to the center of the searchlight, so obtaining a map for each subject and component. For each model, threshold free cluster enhancement (TFCE: Smith and Nichols 2009) was used to detect group-level clusters significantly explained by the corresponding unique variance component (5000 randomizations with 6mm variance smoothing, as implemented in FSL's *randomise*: www.fmrib.ox.ac.uk/fsl/randomise). Statistical maps were then thresholded at one-tailed $p < 0.05$, corrected for multiple comparisons across gray matter voxels (minimum cluster size = 10 voxels; Figure 2).

Orthogonality and complexity testing

Orthogonality was computed by dividing the group-averaged (between subjects) sum of variance explained uniquely by the five models with the group-averaged (between subjects) sum of variance explained by their shared components for each searchlight (Orthogonality = $\Sigma R^2_{unique\ components} / \Sigma R2_{shared\ components}$); a higher value indicates, therefore, that a higher fraction of variance

is explained by individual models, rather than being shared across them. In principle, this value can take extreme values as 0, when only shared information is coded, or infinite, when only orthogonal information is represented. We tested whether a linear trend between the Y coordinate and average orthogonality within XZ-slices was present by searching for abrupt changes in the slope, as high as 50% of the maximum value. As we found no significant changes, the strength of the linear dependency between orthogonality and the posterior-to-anterior direction was calculated using the Spearman's correlation (Figure 5A) and significance was then computed with a parametric test.

Following (Vernon et al., 2016), two different groups of features were identified: low-level representations, sensitive to retinotopic information, and abstract representations, that are independent of the extent of retinotopic cortex stimulated. Inked-area and silhouette were labeled as low-level models, and medial-axis, curvature and identity as abstract ones. Then, complexity was measured by the ratio between variance uniquely explained either by low-level or abstract models: Complexity = ($R^2_{\text{silhouette}} + R^2_{\text{inked area}}$) / ($R^2_{\text{medial-axis}} + R^2_{\text{curvature}} + R^2_{\text{identity}}$). Thus, within each searchlight, group-averaged (between subjects) sum of variance explained uniquely by the low-level models was divided by the group-averaged sum of variance explained by the abstract ones. Linearity was tested as for the orthogonality index.

Surface plots in Figures 3 and 4 were produced with the *Pycortex* toolbox for Python (Gao et al. 2015). Second-level analyses were performed using custom-made code written in MATLAB (MathWorks Inc.).

Results

As expected from both theoretical and experimental investigations on this topic (Kay, 2011; Bracci and Op de Beeck, 2016; Papale et al., 2019), the five models reveal moderate-to-high

261 degrees of collinearity (Figure 1E). Consequently, to account for multicollinearity before
262 considering the significance of the association of each model with brain representations, the
263 variance partitioning analysis (Lescroart et al., 2015) and a searchlight procedure (whole brain,
264 6mm radius: Kriegeskorte et al. 2006) were combined to identify group-level clusters significantly
265 explained by three shape models independently from competing representations (Figure 1D).
266
267 *The human visual cortex encodes multiple orthogonal shape representations*
268 Group-level results show both distinct and overlapping clusters of shape selectivity in OTC,
269 mildly extending also to posterior dorsal regions ($p < 0.05$ one-tailed, TFCE corrected). The
270 silhouette model (Figure 2, in red) shows a significant association with brain representations along
271 the Calcarine sulcus (CalcS), the occipitotemporal sulcus (OTS), the right collateral sulcus (CollS),
272 the right inferior temporal sulcus (ITS), the right fusiform gyrus (FusG), the cuneus (Cun) and in
273 posterior portions of the middle temporal gyrus (pMTG) and intraparietal sulcus (pIPS). The
274 medial-axis (Figure 2, in green) explains a significant portion of unique variance in the right lateral
275 occipital area (LO) only. Finally, curvature (Figure 2, in blue) significantly explains fMRI
276 representational geometries in the left lingual gyrus (LinG), in the bilateral FusG, along bilateral
277 OTS and ITS, along the right CollS, in the right MTG, bilaterally in the Cun and along the right IPS.
278 The significant clusters for the low- and high-level models are also represented in Figure 2.
279 As all orthogonal components of our tested models indicate the presence of at least a
280 significant cluster of selectivity, shape representation does not rely on a single feature, but on a
281 multi-dimensional coding scheme.
282
283 ** Figure 2 near here **

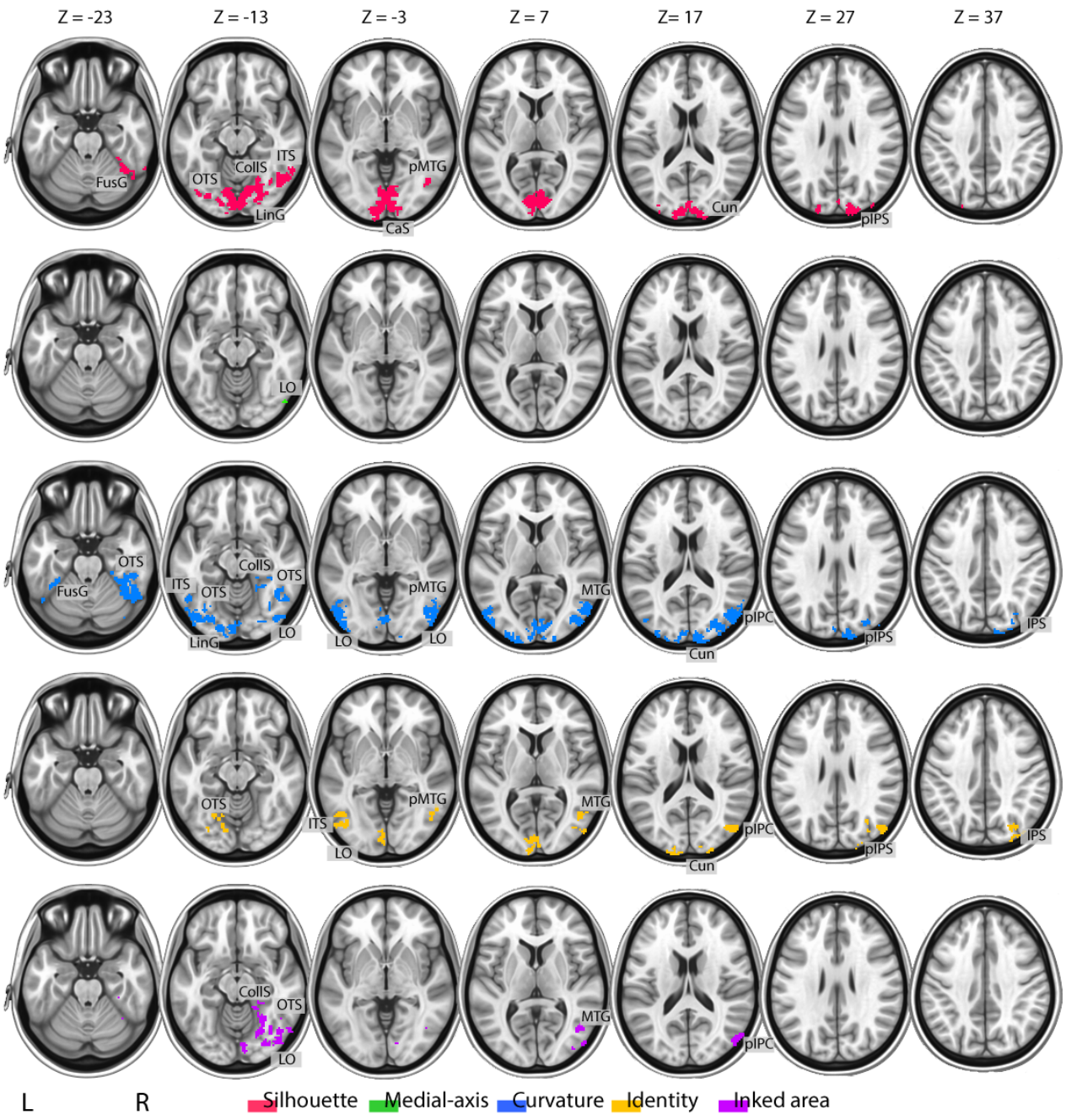
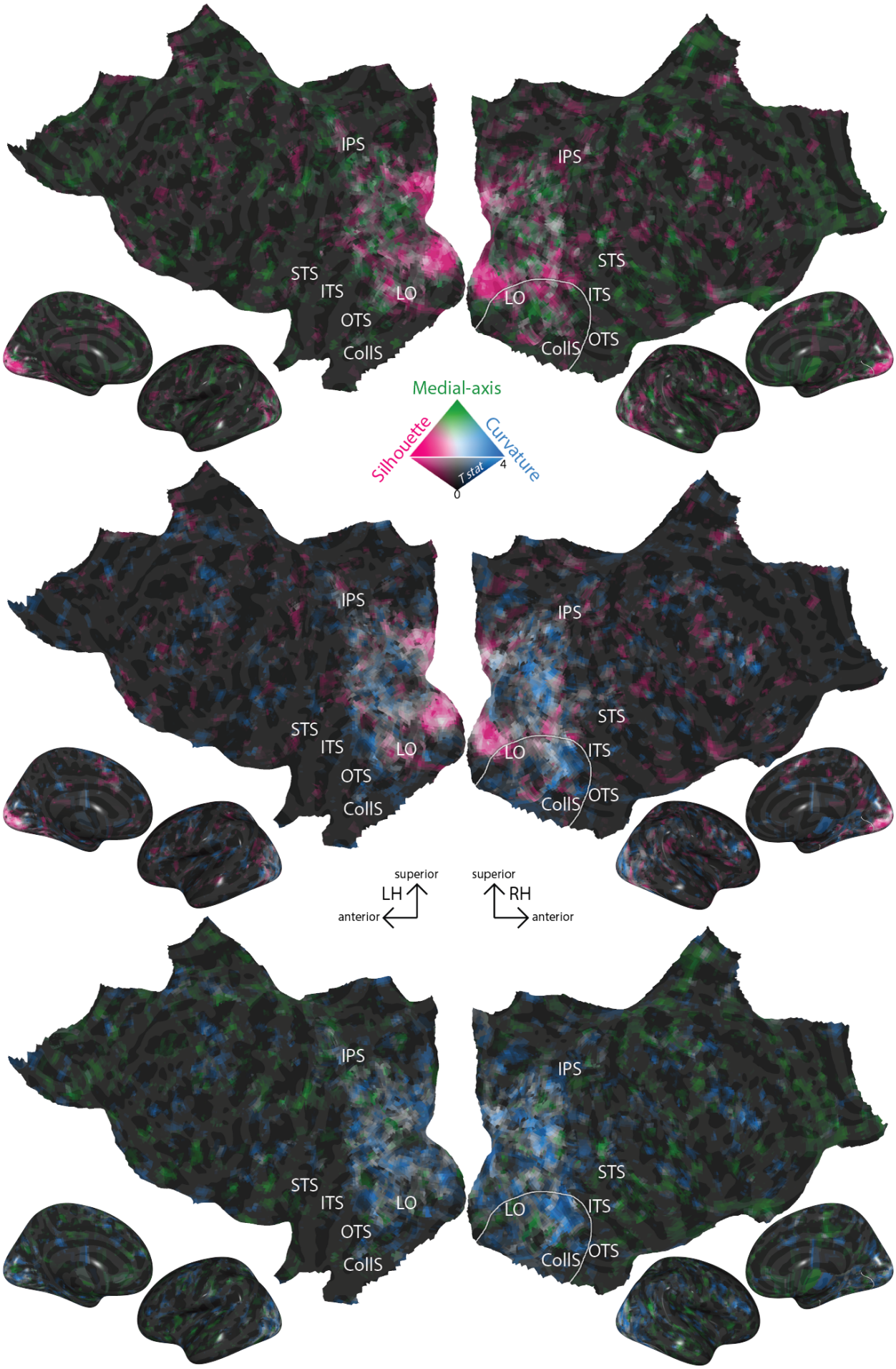


Figure 2. The human visual cortex encodes orthogonal shape representations. Group-level maps showing significant clusters of shape selectivity in OTC and in posterior dorsal regions (one-tailed $p < 0.05$, TFCE corrected). Selectivity to orthogonal components of silhouette (red), medial-axis (green) curvature (blue), object identity (orange) and inked area (purple).

Selectivity to orthogonal shape representations coexist in the same cortical regions

We further explored the overlap between the selectivity to orthogonal shape representations. Figure 3 depicts the pairwise comparisons between the three shape models in our study. Qualitatively, a stronger overlap is observed in LO for medial-axis and curvature, and in

294 IT, right FusG, Cun, right pMTG and right pIPS for silhouette and curvature. Thus, those brain
295 regions encode multiple shape features, independently from the shared variance between them.
296
297 ** Figure 3 near here **
298


IPS
STS
ITS
LO
OTS
CollS
Medial-axis
Silhouette
Curvature
T stat
0
4
superior
LH
anterior
superior
RH
anterior

299

Figure 3. Coding of orthogonal shape components overlap in the human visual cortex
Pairwise comparisons between group-level unthresholded T-maps of orthogonal shape components show that several regions encode more than a single orthogonal description. Colored voxels have high T-value in a single model. Silhouette is represented in red, medial-axis in green and curvature in blue. The overlap between two orthogonal representations is indicated by white voxels, while brightness represents the value of the T-statistic in each voxel (i.e. gray and black voxels have low T-value in both models). White lines enclose right OTC, where all three shape models are significant.

Topographic organization of object shape in right OTC

Of note, all three models are significant only within right OTC (enclosed by a white line in Figure 3). Figure 4 depicts right OTC in isolation with a greater detail: when combining the three models (Figure 4B), a topographic organization emerges. Silhouette coding is medial with respect to the CollS, encompassing the LinG and parahippocampal gyrus (PHG, red voxels in Figure 4B). Proceeding laterally, the silhouette and medial-axis coexist in the fundus of the CollS (orange voxels in Figure 4B), while the medial-axis extends also to the FusG (green voxels in Figure 4B). Finally silhouette and curvature are both encoded medial to the OTS, with curvature being encoded also in the fundus of the OTS.

These qualitative observations return a complex picture on shape coding in the human brain. However, some general considerations can be made by looking at the interactions between features.

** Figure 4 near here **

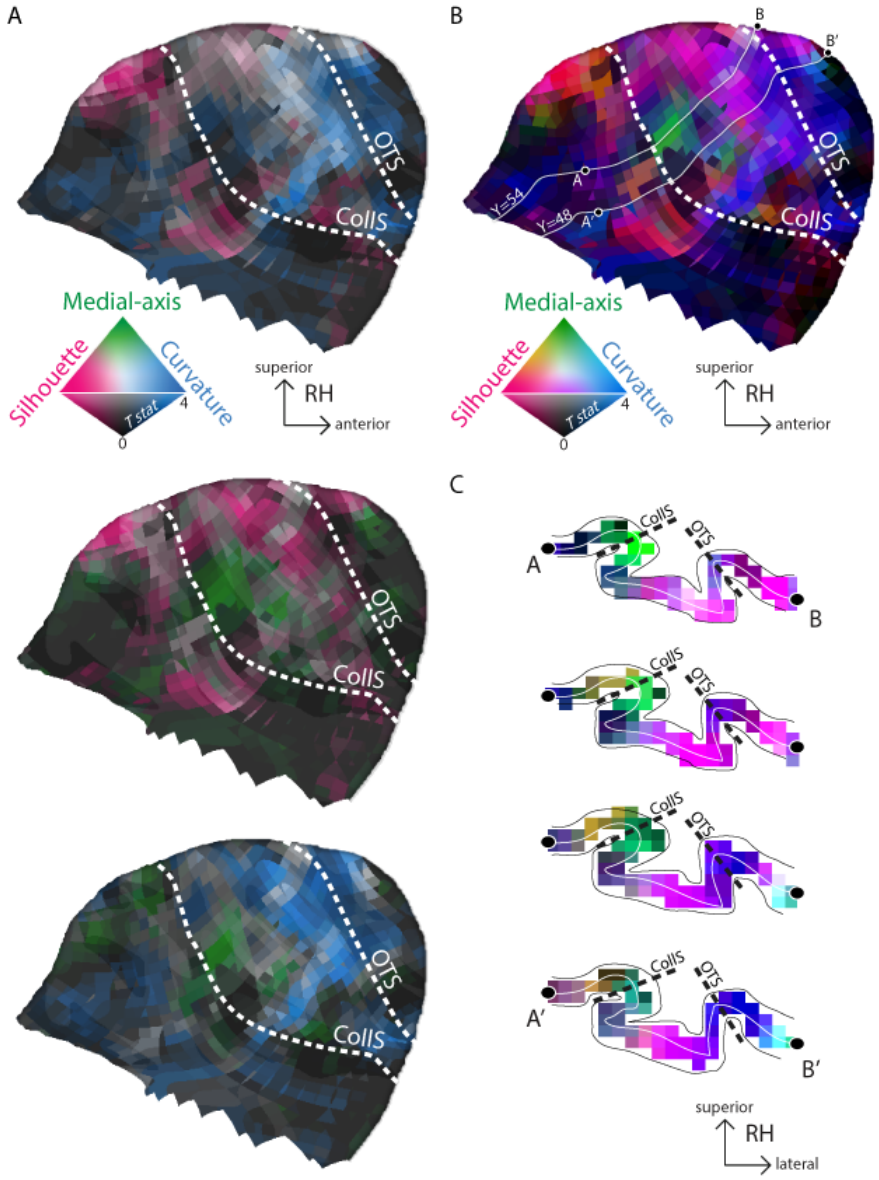


Figure 4. Topographic organization of object shape in right OTC

A) Pairwise comparisons between group-level T-maps of orthogonal shape components in right OTC. Colored voxels have high T-value in a single model. Silhouette is represented in red, medial-axis in green and curvature in blue. The overlap between two orthogonal representations is indicated by white voxels, while brightness represents the value of the T-statistic in each voxel (i.e. gray and black voxels have low T-value in both models).

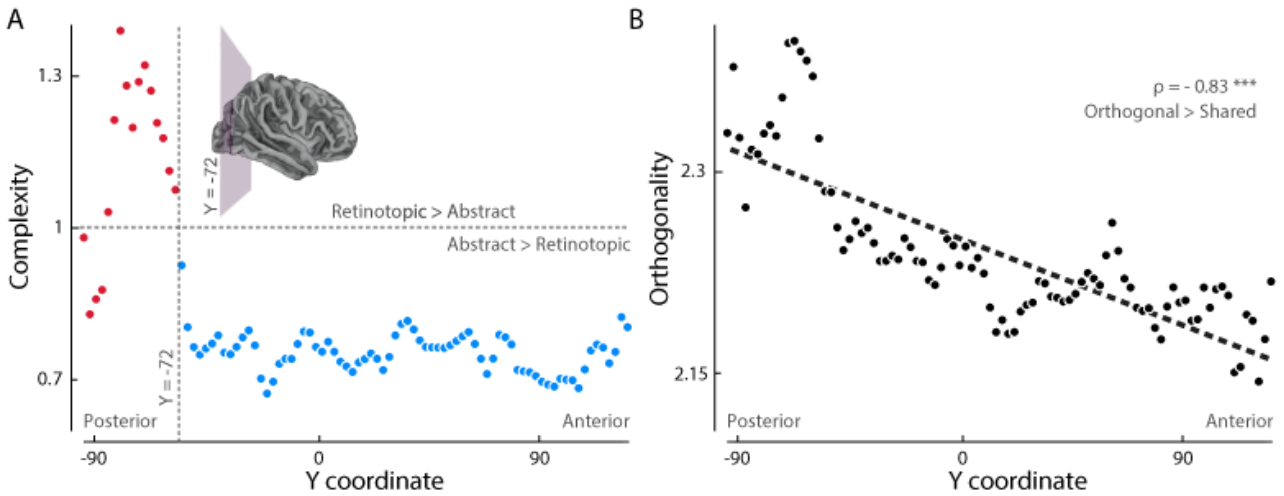
B) Overlap between the three group-level T-maps of orthogonal shape components in right OTC. Silhouette is represented in red, medial-axis in green and curvature in blue. The overlap between two orthogonal representations is indicated by intermediate colors: pink for silhouette and curvature, orange for silhouette and medial-axis, cyan for medial-axis and curvature. Brightness represents the value of the T-statistic in each voxel (i.e. gray and black voxels have low T-value in all models). The solid lines show the slices depicted in (C).

Coding of orthogonal object representations decreases from posterior to anterior regions

338 In a previous study, Vernon *et al.* (2016) explored the relationship between retinotopic and
339 more abstract object representations, including contour curvature. They defined two orthogonal
340 components enclosing low-level and higher-level, complex features, and described a shift between
341 retinotopic and more abstract features in LO. Given the pattern of results from the previous
342 analysis, we further explored what has been observed by Vernon et al. (2016) and also tested the
343 relative weight of orthogonal and shared components. Indeed, the tuning to increasingly complex
344 features is considered a cornerstone of hierarchical object processing (Riesenhuber and Poggio
345 2000). However, it has been proposed that interaction between features may play a pivotal role in
346 evolving reliable selectivity in the brain (Benjamin et al. 2019). Thus, we hypothesized that shared
347 information should become more relevant along the visual hierarchy, moving from posterior to
348 anterior brain regions.

349 We defined two independent components, one for the low-level features and one for the
350 abstract ones, as in Vernon et al (2016). The first comprised the orthogonal variance of silhouette
351 and inked-area, since both are linked to the local retinotopic arrangement and to the extent of
352 retinotopic cortex stimulated. The second includes the orthogonal variance of medial-axis,
353 curvature (both insensitive to differences in object orientation and size) and object identity
354 models. The ratio between the explained variance of low-level and abstract features (i.e.,
355 complexity) in the posterior-to-anterior axis was computed by averaging the values in the XZ
356 plane: values higher than one indicate that brain representations are better accounted for by
357 retinotopic information, while values smaller than one indicate that abstract representations are
358 more relevant. When looking at the slope of complexity along the posterior-to-anterior axis, we
359 observed an abrupt shift (higher than 50% of the maximum) from retinotopic to abstract features
360 around Y_{MNI} = -72 (Figure 5A). Of note, the shift occurs at the limit between occipital and temporal
361 or parietal cortex. As a matter of fact, previous studies (e.g., Haxby et al. 2001; Rice et al. 2014) on

362 ventral temporal cortex selectivity constrained their analysis between Y_{MNI} = -70 and Y_{MNI} = -20
363 (but see: Grill-Spector and Weiner 2014 for a different definition based on anatomical landmarks).
364 Then, we looked at the ratio between orthogonal and shared variance components (i.e.,
365 orthogonality) in the posterior-to-anterior axis. The variance explained by the orthogonal
366 components of the five models was first summed, and then divided by the sum of the shared
367 components between the five models. Here, values higher than one indicate that brain
368 representations are better explained by orthogonal components of variance. Orthogonality
369 linearly decreases along the posterior-to-anterior axis, without shifts (ρ = -0.83, $p < 0.001$,
370 parametric test; Figure 5B). Thus, while orthogonal information is always more represented than
371 shared variance (min = 2.15), it becomes less relevant proceeding along the visual hierarchy.
372 Of note, both these ratios abstract away from magnitude of response and goodness of fit
373 *per se*. Instead, they reveal the relative contribution of different features to different brain areas
374 beyond their overall performance.
375
376 ** Figure 5 near here **



377

378 **Figure 5. The link between object features shapes the human visual hierarchy**
379 A) The ratio between the explained variance of low-level and abstract features (i.e. complexity)
380 along the visual hierarchy reveals an abrupt shift. Values higher than one (horizontal dashed line)
381 indicate that brain representations are better accounted for by retinotopic information while

values smaller than one that abstract representation is more relevant. The vertical dashed line represents the point where mean and slope (dashed black lines) present an abrupt change.
B) The ratio between the variance explained by the orthogonal components of the five models and the sum of the shared components between the five models (i.e. orthogonality) linearly decreases along the visual hierarchy ($\rho = -0.83$, ***: $p < 0.001$, parametric test). Values higher than one indicate that brain representations are better explained by orthogonal components of variance.

Discussion

In the present study, we assessed that object shape is not encoded by a single feature but is encoded by multiple representations (i.e., silhouette, medial-axis and curvature) that independently contribute to object processing in the human visual cortex (Figures 2-4). Moreover, we showed that the brain encodes orthogonal object representations in a topographic fashion: the early visual cortex is biased towards unique components of variance, while shared representations become progressively more relevant in more anterior regions (Figure 5).

Shape coding is multidimensional

In line with previous studies, we found that object silhouette is mainly encoded in early visual areas (Bracci and Op de Beeck 2016; Kaiser et al. 2016; Khaligh-Razavi and Kriegeskorte 2014; Proklova et al. 2016). This result can be explained by top-down figure-dependent mechanisms that modulate V1 activity both in monkeys (Poort et al. 2016; Self et al. 2019) and humans (Kok and de Lange 2014; Muckli et al. 2015), and enhances the processing of object-related information in early visual areas also during natural vision (Papale et al. 2018). However, another possibility may be that the silhouette model better captures the object physical appearance (Kubilius et al. 2016).

Instead, the variance component unique to the medial-axis model – which is the most transformation-resistant shape description (Yang et al. 2008) – was significant in a smaller extent of cortex comprising only a subset of voxels in right LO (Figure 2, middle in green). This can be due

409 to a higher spatial inter-subject variability of this representation that has been already observed
410 by Leeds *et al.* (2013), or to a higher collinearity with the low- and high-level models we employed
411 (Figure 1C) that prevents from disentangling its contribution from competing representations.
412 Nonetheless, our result fits previous evidence of medial-axis coding in monkey IT (Hung et al.
413 2012; putative homologue of human LO), human LO (while also controlling for low-level
414 properties: Ayzenberg et al. 2019b) and is consistent with our previous MEG study showing that
415 medial-axis processing is limited to a small cluster of right posterior sensors, when controlling for
416 collinearity with low-level and categorical representations (Papale et al. 2019).

417 Finally, IT (Kayaert et al. 2005b; Yue et al. 2014), LO (Vernon et al. 2016) and FusG (Caldara
418 et al. 2006) were bilaterally tuned to contour curvature (Figure 2, bottom in blue), in accordance
419 with previous neuroimaging investigations. Actually, LO has a pivotal role in object processing
420 (Grill-Spector et al. 2001; Grill-Spector et al. 1999; Kourtzi and Kanwisher 2001), as IT in monkeys
421 (Brincat and Connor 2004; Desimone et al. 1984; Kayaert et al. 2005a; Op de Beeck et al. 2001;
422 Tanaka 2003; Zoccolan et al. 2007). In addition, while we focus our discussion on the ventral
423 stream, we also observed few significant clusters in dorsal visual regions (R pIPS; see Figure 2),
424 both for curvature and silhouette, which confirm previous observations (Freud et al. 2017).

425 Overall, the evidence that all the tested dimensions independently contribute to shape
426 representation in the human visual cortex favors the hypothesis of a multi-dimensional coding of
427 object shape (Silson et al. 2016; Silson et al. 2013), similarly to what is observed for texture
428 processing (Okazawa et al. 2015; Ziemba et al. 2016). In line with this, a recent study showed that
429 behavioral shape similarity could be modeled with no less than 109 different dimensions
430 (Morgenstern et al. 2020).

431

432 *Coding of shared information increases along the visual hierarchy*

433 Long *et al.* (2018) suggested that mid-level computations, covarying with high-level semantic
434 processing (including curvature extraction), control the organization of OTC. In the present study,
435 however, we observed overlapping selectivity to orthogonal features in LO (medial-axis and
436 curvature), IT, right FusG, Cun, right pMTG and right pIPS (silhouette and curvature). Since we
437 controlled for collinearity between models, this result could not be merely ascribed to the
438 variance shared by those features. While this may apparently result in contrast with the proposal
439 by Long *et al*. (2018), here we also observed that coding of shared descriptions in OTC is
440 topographically arranged and its relevance linearly increases from posterior to anterior regions
441 (Figure 5). This observation, consistent with the core finding of Long *et al.* (2018), suggests that
442 the hierarchy of visual processing is not only shaped by specificity to increasingly complex
443 features, but also by a higher selectivity to shared representations.

444 This observation complements what has been already observed on the two extremes of
445 the ventral visual pathway: V1 and IT. Representations in V1 are over-complete relative to the
446 retinal input (Olshausen and Field 1996; Vinje and Gallant 2000). In addition, inhibitory
447 interactions in V1 are specifically targeted at neurons with similar tuning properties (Chettih and
448 Harvey 2019). Both these factors increase V1 representational capacity and may ultimately lead to
449 a higher selectivity to orthogonal features, as we observed in posterior regions. On the other
450 hand, higher sensitivity to shared information in more anterior areas may be produced by
451 populations of neurons that are not tuned to a specific property but that encode multiple
452 dimensions at once. Indeed, shared featural selectivity has been proposed as the mechanism
453 responsible to achieve dimensionality reduction of the sensory input in IT (Lehky et al. 2014),
454 where both neural density and surface are much lower than in V1 (Cahalane et al. 2012; Van Essen
455 et al. 1992). In line with this, the highest-dimensional among our three shape models (i.e.,
456 silhouette) is also represented in posterior regions (Figure 2). Relatedly, the interaction between

457 multiple features is thought to represent the optimal solution to increase the sensitivity to their
458 mutual changes: in this view, instead of having few neurons encoding a single feature each, it may
459 be preferable to have most of the neurons encoding multiple features at once (Benjamin et al.
460 2019). It has been also suggested that interactions between features are responsible for the poor
461 reliability of tuning curves in predicting brain responses in natural vision (Benjamin et al. 2019).

462 Thus, what can be concluded on the nature of object processing? On one hand, we
463 observed an abrupt shift from retinotopic to abstract representations moving anteriorly across the
464 brain (Figure 5A). However, this shift is relative: though less relevant, orthogonal retinotopic
465 information spreads also to OTC, explaining a significant portion of its variance, in line with
466 previous work and suggesting a link between low-level and object selectivity (Rajimehr et al. 2011;
467 Rice et al. 2014). On the other hand, we found a linear dependency between the posterior-to-
468 anterior axis and the variance explained by shared information (Figure 5B), in line with previous
469 research showing an increase in coding category-orthogonal information from V4 to IT in non-
470 human primates (Hong et al. 2016). As stated earlier, this property describes the linear cascade of
471 computations in the visual hierarchy better than complexity: optimizing the coding of shared
472 variance between behaviorally relevant features may represent a key factor in shaping the
473 architecture of our visual cortex and achieving reliable, view-point invariant object
474 representations. In this light, the next step should be to move from modeling representational
475 geometries to more direct modulations of brain responses, so to control also for nonlinear
476 interactions between features (Benjamin et al. 2019).

477

478 *Limits and conclusions*

479 It should be noted that due to the low fMRI temporal resolution, we cannot resolve which
480 mechanisms support the different tuning for shared representations. Moreover, while the

481 selected models capture visual transformations, many alternative descriptions exist (e.g., Khaligh-
482 Razavi and Kriegeskorte 2014). Overall, however, our results hint at the existence of a multi-
483 dimensional coding of object shape, and reveal that selectivity for shared object representations
484 are topographically arranged and increases along the visual hierarchy. Future experiments will
485 identify how different tasks (e.g., determining object similarity vs. extracting affordances), and
486 alternative descriptions impact on the observed patterns of selectivity. Finally, we described what
487 qualitatively appears to be a shape coding topography in right OTC: further research would be
488 necessary to understand how strong is the link between structural and functional organizations.
489

490 **Acknowledgments**

491 This work has been supported by the Italian Ministry of Education, University and Research
492 grants PRIN 2015WXAXJF and 2015AR52F9.
493

494 **Author contributions**

495 Pa.P., A.L. G.H., L.C. and E.R. conceived the study. A.L., G.H. and L.C. performed experiments.
496 Pa.P., and A.L. analyzed the data. All the authors discussed the results and wrote the manuscript.
497

498 **References**

499

500 **Ayzenberg V, Chen Y, Yousif SR, and Lourenco SF**. Skeletal representations of shape in human
501 vision: Evidence for a pruned medial axis model. *J Vis* 19: 6, 2019a.

502 **Ayzenberg V, Kamps FS, Dilks DD, and Lourenco SF**. A dual role for shape skeletons in human
503 vision: perceptual organization and object recognition. *bioRxiv* 799650, 2019b.

504 **Ayzenberg V, and Lourenco SF**. Skeletal descriptions of shape provide unique perceptual
505 information for object recognition. *Sci Rep* 9: 9359, 2019.

506 **Benjamin AS, Ramkumar P, Fernandes H, Smith MA, and Kording KP**. Hue tuning curves in V4
507 change with visual context. *bioRxiv* 780478, 2019.

508 **Blum H**. Biological shape and visual science. I. *J Theor Biol* 38: 205-287, 1973.

509 **Bracci S, and Op de Beeck H**. Dissociations and Associations between Shape and Category
510 Representations in the Two Visual Pathways. *J Neurosci* 36: 432-444, 2016.

511 **Bracci S, Ritchie JB, Kalfas I, and Op de Beeck H**. The ventral visual pathway represents animal
512 appearance over animacy, unlike human behavior and deep neural networks. *J Neurosci* 2019.

513 **Brincat SL, and Connor CE**. Underlying principles of visual shape selectivity in posterior
514 inferotemporal cortex. *Nat Neurosci* 7: 880-886, 2004.

515 **Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, and Poggio T**. A model of V4 shape
516 selectivity and invariance. *J Neurophysiol* 98: 1733-1750, 2007.

517 **Cahalane DJ, Charvet CJ, and Finlay BL**. Systematic, balancing gradients in neuron density and
518 number across the primate isocortex. *Front Neuroanat* 6: 28, 2012.

519 **Caldara R, Seghier ML, Rossion B, Lazeyras F, Michel C, and Hauert CA**. The fusiform face area is
520 tuned for curvilinear patterns with more high-contrasted elements in the upper part. *Neuroimage*
521 31: 313-319, 2006.

522 **Carlson ET, Rasquinha RJ, Zhang K, and Connor CE**. A sparse object coding scheme in area V4. *Curr*
523 *Biol* 21: 288-293, 2011.

524 **Chettih SN, and Harvey CD**. Single-neuron perturbations reveal feature-specific competition in V1.
525 *Nature* 567: 334-340, 2019.

526 **Connor CE, Brincat SL, and Pasupathy A**. Transformation of shape information in the ventral
527 pathway. *Curr Opin Neurobiol* 17: 140-147, 2007.

528 **Cox RW**. AFNI: software for analysis and visualization of functional magnetic resonance
529 neuroimages. *Comput Biomed Res* 29: 162-173, 1996.

530 **de Heer WA, Huth AG, Griffiths TL, Gallant JL, and Theunissen FE**. The Hierarchical Cortical
531 Organization of Human Speech Processing. *J Neurosci* 37: 6539-6557, 2017.

532 **Desimone R, Albright TD, Gross CG, and Bruce C**. Stimulus-selective properties of inferior
533 temporal neurons in the macaque. *J Neurosci* 4: 2051-2062, 1984.

534 **Elder JH, and Velisavljevic L**. Cue dynamics underlying rapid detection of animals in natural
535 scenes. *J Vis* 9: 7, 2009.

536 **Freud E, Culham JC, Plaut DC, and Behrmann M**. The large-scale organization of shape processing
537 in the ventral and dorsal pathways. *Elife* 6: 2017.

538 **Gao JS, Huth AG, Lescroart MD, and Gallant JL**. Pycortex: an interactive surface visualizer for fMRI.
539 *Front Neuroinform* 9: 23, 2015.

540 **Grill-Spector K, Kourtzi Z, and Kanwisher N**. The lateral occipital complex and its role in object
541 recognition. *Vision Res* 41: 1409-1422, 2001.

542 **Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, and Malach R**. Differential processing
543 of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:
544 187-203, 1999.

545 **Grill-Spector K, and Weiner KS**. The functional architecture of the ventral temporal cortex and its
546 role in categorization. *Nature Reviews Neuroscience* 15: 536-548, 2014.

547 **Groen, II, Greene MR, Baldassano C, Fei-Fei L, Beck DM, and Baker CI**. Distinct contributions of
548 functional and deep neural network features to representational similarity of scenes in human
549 brain and behavior. *Elife* 7: 2018.

550 **Handjaras G, Leo A, Cecchetti L, Papale P, Lenci A, Marotta G, Pietrini P, and Ricciardi E**.
551 Modality-independent encoding of individual concepts in the left parietal cortex.
552 *Neuropsychologia* 105: 39-49, 2017.

553 **Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, and Pietrini P**. Distributed and overlapping
554 representations of faces and objects in ventral temporal cortex. *Science* 293: 2425-2430, 2001.

555 **Hong H, Yamins DL, Majaj NJ, and DiCarlo JJ**. Explicit information for category-orthogonal object
556 properties increases along the ventral stream. *Nat Neurosci* 19: 613-622, 2016.

557 **Hung CC, Carlson ET, and Connor CE**. Medial axis shape coding in macaque inferotemporal cortex.
558 *Neuron* 74: 1099-1113, 2012.

559 **Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, and Smith SM**. Fsl. *Neuroimage* 62: 782-
560 790, 2012.

561 **Kaiser D, Azzalini DC, and Peelen MV**. Shape-independent object category responses revealed by
562 MEG and fMRI decoding. *J Neurophysiol* 115: 2246-2250, 2016.

563 **Kay KN**. Understanding visual representation by developing receptive-field models. *Visual*
564 *population codes: Towards a common multivariate framework for cell recording and functional*
565 *imaging* 133-162, 2011.

566 **Kayaert G, Biederman I, Op de Beeck HP, and Vogels R**. Tuning for shape dimensions in macaque
567 inferior temporal cortex. *Eur J Neurosci* 22: 212-224, 2005a.

568 **Kayaert G, Biederman I, and Vogels R**. Representation of regular and irregular shapes in macaque
569 inferotemporal cortex. *Cereb Cortex* 15: 1308-1321, 2005b.

570 **Khaligh-Razavi S-M, and Kriegeskorte N**. Deep supervised, but not unsupervised, models may
571 explain IT cortical representation. 2014.

572 **Kok P, and de Lange FP**. Shape perception simultaneously up- and downregulates neural activity in
573 the primary visual cortex. *Curr Biol* 24: 1531-1535, 2014.

574 **Kourtzi Z, and Kanwisher N**. Representation of perceived object shape by the human lateral
575 occipital complex. *Science* 293: 1506-1509, 2001.

576 **Kriegeskorte N, Goebel R, and Bandettini P**. Information-based functional brain mapping.
577 *Proceedings of the National Academy of Sciences of the United States of America* 103: 3863-3868,
578 2006.

579 **Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, and Bandettini PA**.
580 Matching categorical object representations in inferior temporal cortex of man and monkey.
581 *Neuron* 60: 1126-1141, 2008.

582 **Kubilius J, Bracci S, and Op de Beeck HP**. Deep Neural Networks as a Computational Model for
583 Human Shape Sensitivity. *PLOS Comput Biol* 12: e1004896, 2016.

584 **Lawrence SJ, Keefe BD, Vernon RJ, Wade AR, McKeefry DJ, and Morland AB**. Global shape
585 aftereffects in composite radial frequency patterns. *J Vis* 16: 17, 2016.

586 **Leeds DD, Seibert DA, Pyles JA, and Tarr MJ**. Comparing visual representations across human fMRI
587 and computational vision. *J Vis* 13: 25, 2013.

588 **Lehky SR, Kiani R, Esteky H, and Tanaka K**. Dimensionality of object representations in monkey
589 inferotemporal cortex. *Neural Comput* 26: 2135-2162, 2014.

590 **Lescroart MD, and Biederman I**. Cortical representation of medial axis structure. *Cereb Cortex* 23:
591 629-637, 2013.

592 **Lescroart MD, Stansbury DE, and Gallant JL**. Fourier power, subjective distance, and object
593 categories all provide plausible models of BOLD responses in scene-selective visual areas. *Front*
594 *Comput Neurosci* 9: 135, 2015.

Long B, Stormer VS, and Alvarez GA. Mid-level perceptual features contain early cues to animacy. *J Vis* 17: 20, 2017.

Long B, Yu CP, and Konkle T. Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc Natl Acad Sci U S A* 115: E9015-E9024, 2018.

Lowet AS, Firestone C, and Scholl BJ. Seeing structure: Shape skeletons modulate perceived similarity. *Atten Percept Psychophys* 80: 1278-1289, 2018.

Monroy A, Eigenstetter A, and Ommer B. Beyond straight lines—object detection using curvature. IEEE, 2011, p. 3561-3564.

Morgenstern Y, Hartmann F, Schmidt F, Tiedemann H, Prokott E, Maiello G, and Fleming RW. An image-computable model of human visual shape similarity. *bioRxiv* 2020.

Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, and Yacoub E. Contextual Feedback to Superficial Layers of V1. *Curr Biol* 25: 2690-2695, 2015.

Nimon K, Lewis M, Kane R, and Haynes RM. An R package to compute commonality coefficients in the multiple regression case: an introduction to the package and a practical example. *Behav Res Methods* 40: 457-466, 2008.

Nimon KF, and Oswald FL. Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods* 16: 650-674, 2013.

612 **Okazawa G, Tajima S, and Komatsu H**. Image statistics underlying natural texture selectivity of
613 neurons in macaque V4. *Proc Natl Acad Sci U S A* 112: E351-360, 2015.

614 **Olshausen BA, and Field DJ**. Emergence of simple-cell receptive field properties by learning a
615 sparse code for natural images. *Nature* 381: 607-609, 1996.

616 **Op de Beeck H, Wagemans J, and Vogels R**. Inferotemporal neurons represent low-dimensional
617 configurations of parameterized shapes. *Nat Neurosci* 4: 1244-1252, 2001.

618 **Papale P, Betta M, Handjaras G, Malfatti G, Cecchetti L, Rampinini A, Pietrini P, Ricciardi E,**
619 **Turella L, and Leo A**. Common spatiotemporal processing of visual features shapes object
620 representation. *Sci Rep* 9: 7601, 2019.

621 **Papale P, Leo A, Cecchetti L, Handjaras G, Kay KN, Pietrini P, and Ricciardi E**. Foreground-
622 Background Segmentation Revealed during Natural Image Viewing. *eNeuro* 5: 2018.

623 **Poort J, Self MW, van Vugt B, Malkki H, and Roelfsema PR**. Texture Segregation Causes Early
624 Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex. *Cereb*
625 *Cortex* 26: 3964-3976, 2016.

626 **Proklova D, Kaiser D, and Peelen MV**. Disentangling Representations of Object Shape and Object
627 Category in Human Visual Cortex: The Animate-Inanimate Distinction. *J Cogn Neurosci* 28: 680-
628 692, 2016.

Rajimehr R, Devaney KJ, Bilenko NY, Young JC, and Tootell RBH. The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys. *PLoS Biol* 9: e1000608, 2011.

Rice GE, Watson DM, Hartley T, and Andrews TJ. Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *The Journal of Neuroscience* 34: 8837-8844, 2014.

Riesenhuber M, and Poggio T. Models of object recognition. *Nature neuroscience* 3: 1199-1204, 2000.

Sebastian TB, Klein PN, and Kimia BB. Recognition of shapes by editing their shock graphs. *IEEE Trans Pattern Anal Mach Intell* 26: 550-571, 2004.

Self MW, Jeurissen D, van Ham AF, van Vugt B, Poort J, and Roelfsema PR. The Segmentation of Proto-Objects in the Monkey Primary Visual Cortex. *Curr Biol* 29: 1019-1029 e1014, 2019.

Silson EH, Groen, II, Kravitz DJ, and Baker CI. Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J Vis* 16: 14, 2016.

Silson EH, McKeefry DJ, Rodgers J, Gouws AD, Hymers M, and Morland AB. Specialized and independent processing of orientation and shape in visual field maps LO1 and LO2. *Nat Neurosci* 16: 267-269, 2013.

647 **Smith SM, and Nichols TE**. Threshold-free cluster enhancement: addressing problems of
648 smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44: 83-98,
649 2009.

650 **Tanaka K**. Columns for complex visual object features in the inferotemporal cortex: clustering of
651 cells with similar but slightly different stimulus selectivities. *Cereb Cortex* 13: 90-99, 2003.

652 **Van Eede M, Macrini D, Telea A, Sminchisescu C, and Dickinson SS**. Canonical skeletons for shape
653 matching. IEEE, 2006, p. 64-69.

654 **Van Essen DC, Anderson CH, and Felleman DJ**. Information processing in the primate visual
655 system: an integrated systems perspective. *Science* 255: 419-423, 1992.

656 **Vernon RJ, Gouws AD, Lawrence SJ, Wade AR, and Morland AB**. Multivariate Patterns in the
657 Human Object-Processing Pathway Reveal a Shift from Retinotopic to Shape Curvature
658 Representations in Lateral Occipital Areas, LO-1 and LO-2. *J Neurosci* 36: 5763-5774, 2016.

659 **Vinje WE, and Gallant JL**. Sparse coding and decorrelation in primary visual cortex during natural
660 vision. *Science* 287: 1273-1276, 2000.

661 **Wolfe JM, Yee A, and Friedman-Hill SR**. Curvature is a basic feature for visual search tasks.
662 *Perception* 21: 465-480, 1992.

663 **Yang M, Kpalma K, and Ronsin J**. *A survey of shape feature extraction techniques*. In-Tech, 2008.

664 **Yue X, Pourladian IS, Tootell RB, and Ungerleider LG**. Curvature-processing network in macaque
665 visual cortex. *Proc Natl Acad Sci U S A* 111: E3467-3475, 2014.

666 **Ziemba CM, Freeman J, Movshon JA, and Simoncelli EP**. Selectivity and tolerance for visual
667 texture in macaque V2. *Proc Natl Acad Sci U S A* 113: E3140-3149, 2016.

668 **Zoccolan D, Kouh M, Poggio T, and DiCarlo JJ**. Trade-off between object selectivity and tolerance
669 in monkey inferotemporal cortex. *J Neurosci* 27: 12292-12307, 2007.

670