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

Authors and affiliations: Paolo Papale[1,2,*], Andrea Leo[1,2], 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. *These authors contributed equally*

Running head: Shape coding in natural vision

***Corresponding author**: Paolo Papale: *paolo.papale@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 representations are thought to occur at specific stages of the visual hierarchy, many of them are correlated (e.g., object shape and identity) and can be retrieved from the activity of several 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 objects, and employed a statistical approach that partition orthogonal and shared object representations to reveal their relative impact on brain processing. Orthogonal shape representations (i.e., silhouette, curvature and medial-axis) independently explain distinct and overlapping clusters of selectivity in occitotemporal (OTC) and parietal cortex. Moreover, we showed 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

While we always have available a general sense of what 'a shape is', what is the computational counterpart of this immediate percept? 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

Since the advent of neuroimaging, much effort has been devoted to characterizing object-selectivity patterns in the human occipito-temporal cortex (OTC; Haxby et al., 2001). Several possible organizing principles have been proposed to explain the large-scale topography of OTC, ranging from the tuning to low-level visual features (e.g., contrast and spatial frequencies - Papale et al. 2018; Rajimehr et al. 2011; Rice et al. 2014), to the processing of broad semantic dimensions, such as object size or the animate-inanimate distinction (Coggan et al. 2016; Julian et al. 2017; Konkle and Caramazza 2013).

There is little doubt, however, that these distinct visual dimensions, ranging from local orientation to identity, may equally contribute to the striking coherency of our object perception (Figure 1A). Thus, to establish the origins of the intrinsic organization in human visual cortex, we would need to understand how these dimensions are coded, and how they mutually interact.

Nonetheless, remarkable evidence from previous studies suggests that visual dimensions in natural vision are indeed highly correlated (Bracci and Op de Beeck 2016; Kay 2011; Papale et al. 2019). Thus, addressing the extent to which brain regions represent different dimensions along the visual hierarchy has so far proven challenging: how can we disentangle the role of different object properties (e.g., shape and category) if they likely covary together?

Notably, Bracci and Op de Beeck (2016) employed a set of stimuli in which shape silhouette and category were dissociated (i.e., by selecting objects similar in shape but pertaining to different categories), and demonstrated that object-selectivity in OTC cannot be merely ascribed just to a specific visual property, such as shape silhouette. Conversely, Long *et al.* (2018) showed that mid-level features, such as texture and curvature, covary with high-level semantic dimensions, and are capable to explain the representations in OTC, even when using synthetic and unidentifiable

stimuli that hinder object recognition. Hence, even if we acknowledge that visual dimensions, such as shape silhouette (Bracci and Op de Beeck, 2016) or curvature (Long et al. 2018), are relevant to OTC, what is their relative contribution in explaining its activity patterns? In this regard, shape is an elusive object property: while a general sense of 'what a shape is' is always available to us, what is the computational counterpart of this immediate percept? For instance, silhouette and curvature capture different features of object shape, as exemplified in Figure 1B.

Another question emerges from the existing literature. Both orthogonal (Bracci and Op de Beeck, 2016) and shared (Long et al., 2018) representations between different visual dimensions explain to a large extent the patterns of brain responses evoked by viewed objects. However, are different brain regions encoding more orthogonal or shared representations? As a matter of fact, the brain focuses on specific aspects of object along different brain regions of the visual hierarchy. Consequently, high level associative regions may encode shared object representations, in order to integrate fragmented descriptions into coherent percepts, while the opposite may hold for early sensory regions, aimed at representing the incoming signal with the highest fidelity.

To answer these questions, we recorded functional MRI (fMRI) activity while participants passively attended to object pictures. We employed a statistical approach that partitions orthogonal and shared shape representations revealing their relative impact on brain processing, while controlling at the same time for low- and high-level confounds (Figure 1; Lescroart et al. 2015). 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: Figure 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 **

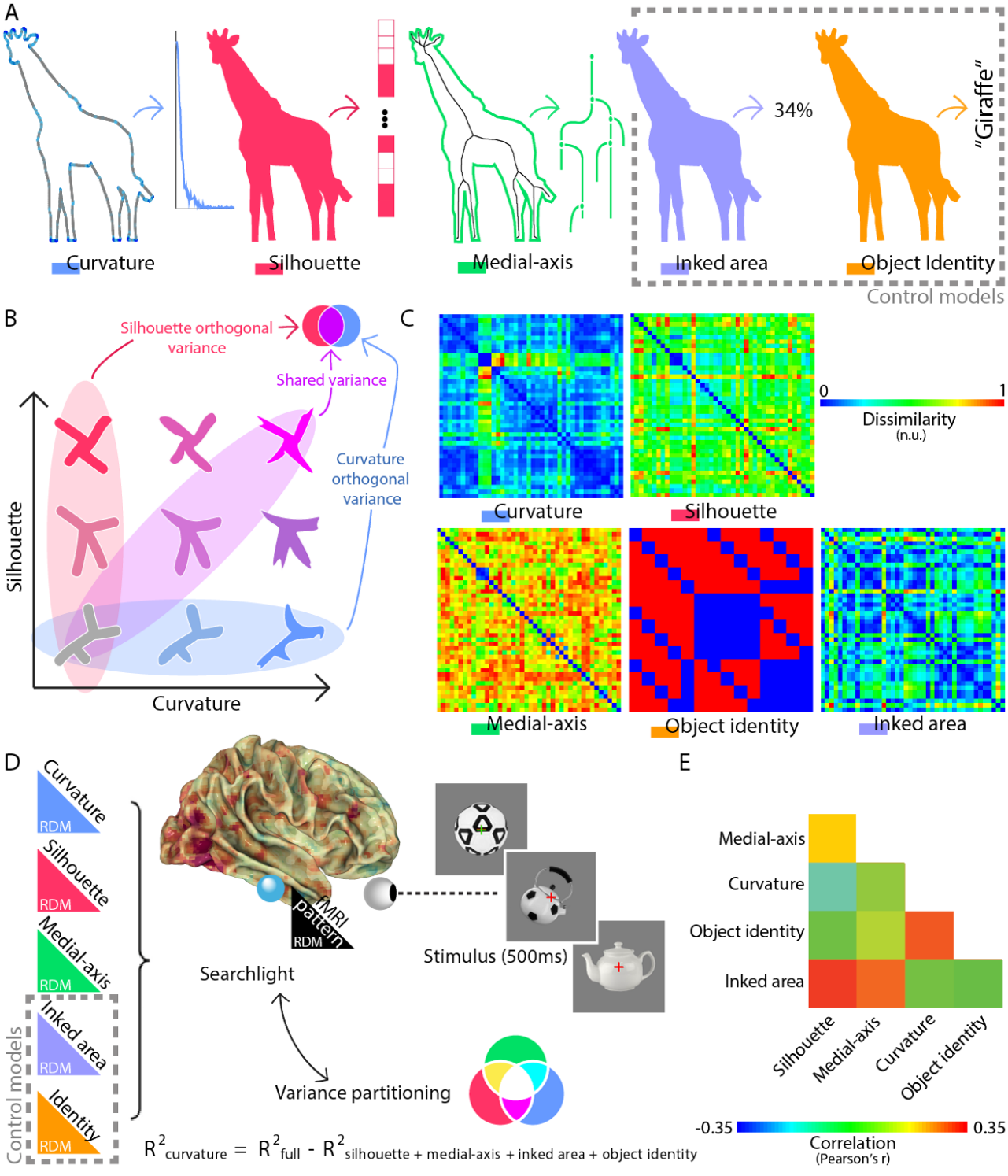


Figure 1. **Schematic of the shape models and experiment.**
A) Five different object representations are employed: three shape models and two further controls. From left: silhouette, medial axis, curvature, inked area (low-level control) and object identity (high-level control).

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 a duration of 500ms 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: within each searchlight, the brain activity RDM was correlated with a combination of the impact of the five models and of their shared variance.
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 a 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. 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 which consisted of 500ms 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 7:20 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-

146 planar imaging sequence (GE-EPI) was used, with TE = 40ms, TR = 2500ms, FA = 90°, 160 volumes
147 with four additional dummy scans, acquisition time 6'50"; image geometry parameters were:
148 Field-Of-View 258x258mm, 128x128 in-plane matrix, voxel size 2.03x2.03x4mm, 37 axial slices for
149 total brain coverage (z-axis extent = 148mm). To acquire detailed information of subject anatomy,
150 a 3D Fast Spoiled Gradient Echo T1-weighted sequence was also acquired (TE = 3.18ms, TR =
151 8.16ms, FA = 12°, Field-Of-View 256x256mm, 256x256 matrix size, $1mm^3$ isotropic voxels, 256 axial
152 slices, z-axis extent 256mm).
153
154 *Functional MRI data processing*
155 Data preprocessing was carried out with AFNI (Cox 1996) and FSL 5.0 (Jenkinson et al.
156 2012). Preprocessing of functional data comprised slice timing correction with Fourier method
157 (*3dtshift*), rigid-body motion correction using the first volume of the third run as reference
158 (*3dvolreg*), spike removal (elimination of outliers in the functional time series, *3dDespike*),
159 smoothing with a Gaussian filter (fixed FWHM 4 mm, *3dmerge*), scaling of BOLD time series to
160 percentage of the mean of each run (*3dTstat*, *3dcalc*). Processing of anatomical images consisted
161 of brain extraction (*bet*), segmentation for bias-field estimation and removal (*FAST*, *fslmaths*),
162 linear (*FLIRT*) and nonlinear registration (*FNIRT*) to MNI152 standard space.
163 For each subject, data from the six concatenated runs (960 time points) were used for a
164 GLM analysis (*3dDeconvolve*) with the responses for each stimulus – modeled with 1 seconds-long
165 block functions convolved with a canonical HRF – as predictors of interest, and the six motion
166 parameters plus polynomial trends up to 4th order as predictors of no-interest.
167 Responses for individual stimuli were converted to MNI152 space by applying the
168 transformation matrices estimated as explained above, and resampled to a resolution of
169 2x2x2mm.

Shape models and controls

Five different representations of the 42 stimuli were developed: three shape-based descriptions of interest and two further controls. 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 control RDMs were built. For the inked-area bias, pairwise dissimilarity was computed as the Euclidean distance 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

194 variance of brain representations (computed using Pearson's correlation distance), as computed in
195 6 mm-radius spherical searchlights (Kriegeskorte et al. 2006). To this aim, explained variance
196 coefficient (R^2) was computed for each model RDM in independent linear regressions, and then all
197 the different combinations of models were tested in further multiple linear regressions. The final
198 statistic reporting the partial goodness of fit for unique and shared components was computed
199 following the work by Nimon and colleagues (2008). To exemplify, the unique variance explained
200 by the curvature model in a specific searchlight was determined as the difference between the
201 full-model R^2 and the variance explained by the combination of all other models (i.e., $R^2_{\text{curvature}}$ =
202 $R^2_{\text{full}} - R^2_{\text{silhouette + medial-axis +inked area + identity}}$). In the context of multiple linear regression, this approach
203 is better known as 'commonality analysis' (Nimon and Oswald 2013), and its popularity is growing
204 in neuroimaging (de Heer et al. 2017; Groen et al. 2018; Lescroart et al. 2015).

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

216

217 *Orthogonality and complexity testing*

218 Orthogonality was computed by dividing the group-averaged sum of variance explained
219 uniquely by the five models with the group-averaged sum of variance explained by their shared
220 components for each searchlight; a higher value indicates, therefore, that a higher fraction of
221 variance is explained by individual models, rather than being shared across them. We tested
222 whether a linear trend between the Y coordinate and mean orthogonality in each XZ-slice was
223 present by searching for abrupt changes in the slope, as high as 50% of the maximum value. As we
224 found no significant changes, the strength of the linear dependency between orthogonality and
225 the posterior-to-anterior direction was calculated using the Spearman's correlation (Figure 5A)
226 and significance was then computed with a parametric test.

227 Following (Vernon et al., 2016), two different groups of features were identified: low-level
228 representations, sensitive to retinotopic information, and abstract representations, that are
229 independent of the extent of retinotopic cortex stimulated. Inked-area and silhouette were
230 labeled as low-level models, and medial-axis, curvature and category as abstract ones. Then,
231 complexity was measured by the ratio between variance uniquely explained either by low-level or
232 abstract models. Thus, within each searchlight, group-averaged sum of variance explained
233 uniquely by the low-level models was divided by the group-averaged sum of variance explained by
234 the abstract ones. Linearity was tested as for the orthogonality index.

235

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

239

240 **Results**

241 Here, three competing shape models were tested. A first description was computed by
242 extracting the *silhouette*, consisting of a simple stimulus vectorization. The link between shape
243 silhouette and OTC representations has been extensively investigated in neuroimaging studies
244 (Bracci and Op de Beeck 2016; Kaiser et al. 2016; Khaligh-Razavi and Kriegeskorte 2014; Proklova
245 et al. 2016). Second, a skeletal representation of each stimulus was extracted by performing the
246 *medial axis* transform (Blum 1973). It controls the spike rate of IT neurons in monkey (Hung et al.
247 2012), captures behavioral ratings of shape similarity (Lowet et al. 2018) and its spatiotemporal
248 association with brain activity in humans has been described in several neuroimaging studies
249 (Handjaras et al. 2017; Leeds et al. 2013; Lescroart and Biederman 2013; Papale et al. 2019). A
250 third description was obtained by computing the *curvature* distribution for each object's contour.
251 It has been showed that V4 neurons in monkey are selective to a specific degree of curvature
252 (Cadieu et al. 2007; Carlson et al. 2011; Connor et al. 2007). Moreover, the pivotal role of contour
253 curvature in object perception has been extensively demonstrated both by behavioral (Elder and
254 Velisavljevic 2009; Lawrence et al. 2016; Long et al. 2017; Wolfe et al. 1992) and neuroimaging
255 studies in humans (Caldara et al. 2006; Long et al. 2018; Vernon et al. 2016; Yue et al. 2014). In
256 addition, the area (in pixels) of each stimulus was computed to account for the *inked-area* bias – a
257 problem that is almost unavoidable when using complex objects in isolation (but see Bracci and Op
258 de Beeck 2016 for an elegant stimulus design). Finally, to get rid of high-level biases that could
259 affect the performance of the three shape models, object *identity* was included as a further
260 control (Khaligh-Razavi and Kriegeskorte 2014; Kriegeskorte et al. 2008).

261 As expected from both theoretical and experimental investigations on this topic (Kay, 2011;
262 Bracci and Op de Beeck, 2016; Papale et al., 2019), the five models show moderate-to-high
263 degrees of collinearity (Figure 1E). Consequently, we used a method that accounts for
264 multicollinearity before considering the significance of the association of each model with brain

265 representations. Combining the variance partitioning analysis (Lescroart et al., 2015) and a
266 searchlight procedure (whole brain, 6mm radius: Kriegeskorte et al. 2006), we identified group-
267 level clusters significantly explained by three physiologically-validated shape models
268 independently from competing representations (Figure 1D).
269
270 *The human visual cortex encodes multiple orthogonal shape representations*
271 Group-level results show both distinct and overlapping clusters of shape selectivity in OTC,
272 mildly extending also to posterior dorsal regions ($p < 0.05$ one-tailed, TFCE corrected). The
273 silhouette model (Figure 2, in red) shows a significant association with brain representations along
274 the Calcarine sulcus (CalcS), the occipitotemporal sulcus (OTS), the right collateral sulcus (CollS),
275 the right inferior temporal sulcus (ITS), the right fusiform gyrus (FusG), the cuneus (Cun) and in
276 posterior portions of the middle temporal gyrus (pMTG) and intraparietal sulcus (pIPS). The
277 medial-axis (Figure 2, in green) explains a significant portion of unique variance in the right lateral
278 occipital area (LO) only. Finally, curvature (Figure 2, in blue) significantly explains fMRI
279 representational geometries in the left lingual gyrus (LinG), in the bilateral FusG, along bilateral
280 OTS and ITS, along the right CollS, in the right MTG, bilaterally in the Cun and along the right IPS.
281 The significant clusters for the control models are also represented in Figure 2.
282 As all orthogonal components of our tested models show at least a significant cluster of
283 selectivity, shape representation does not rely on a single feature, but on a multi-dimensional
284 coding scheme.
285
286 ** Figure 2 near here **

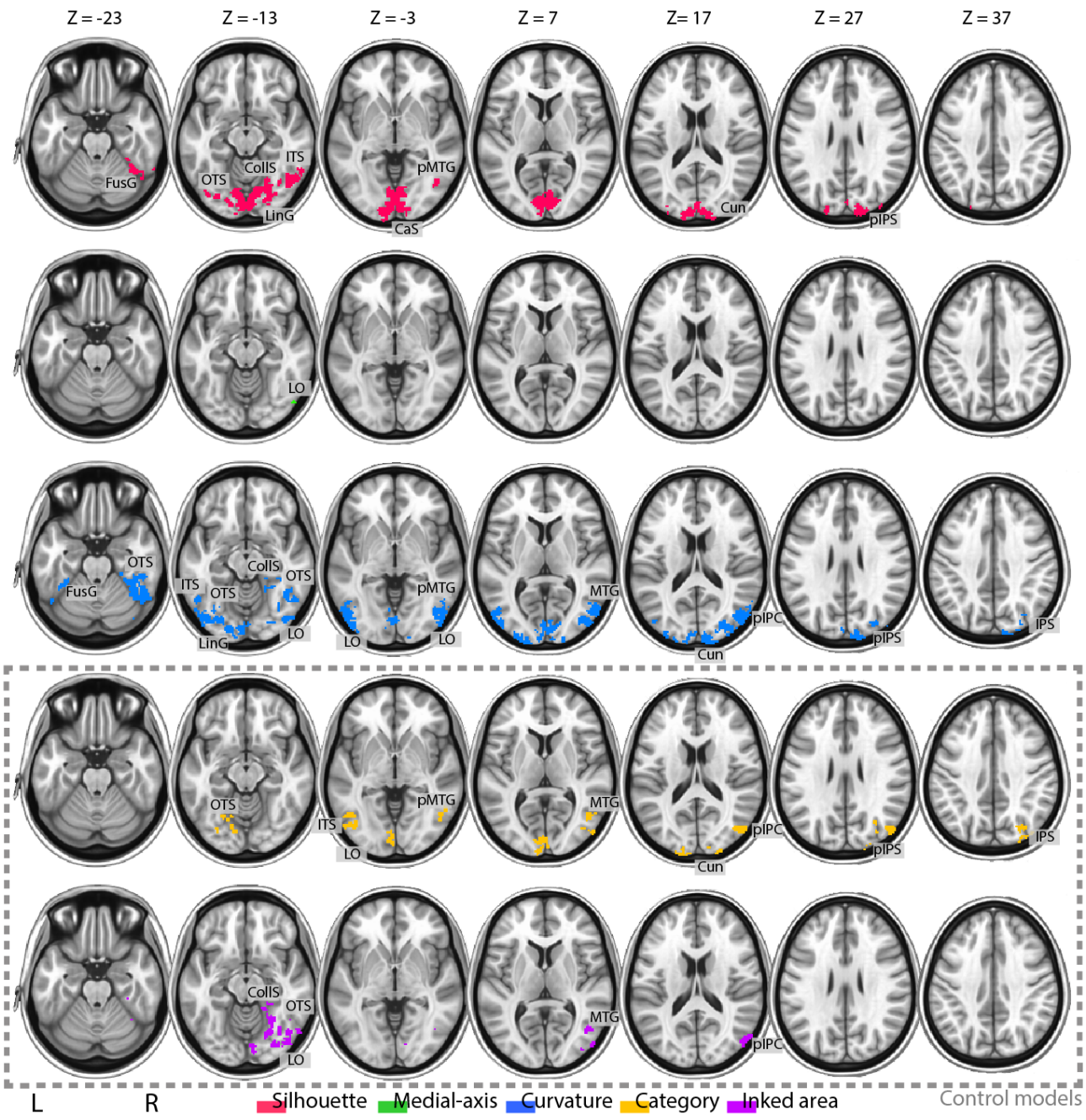
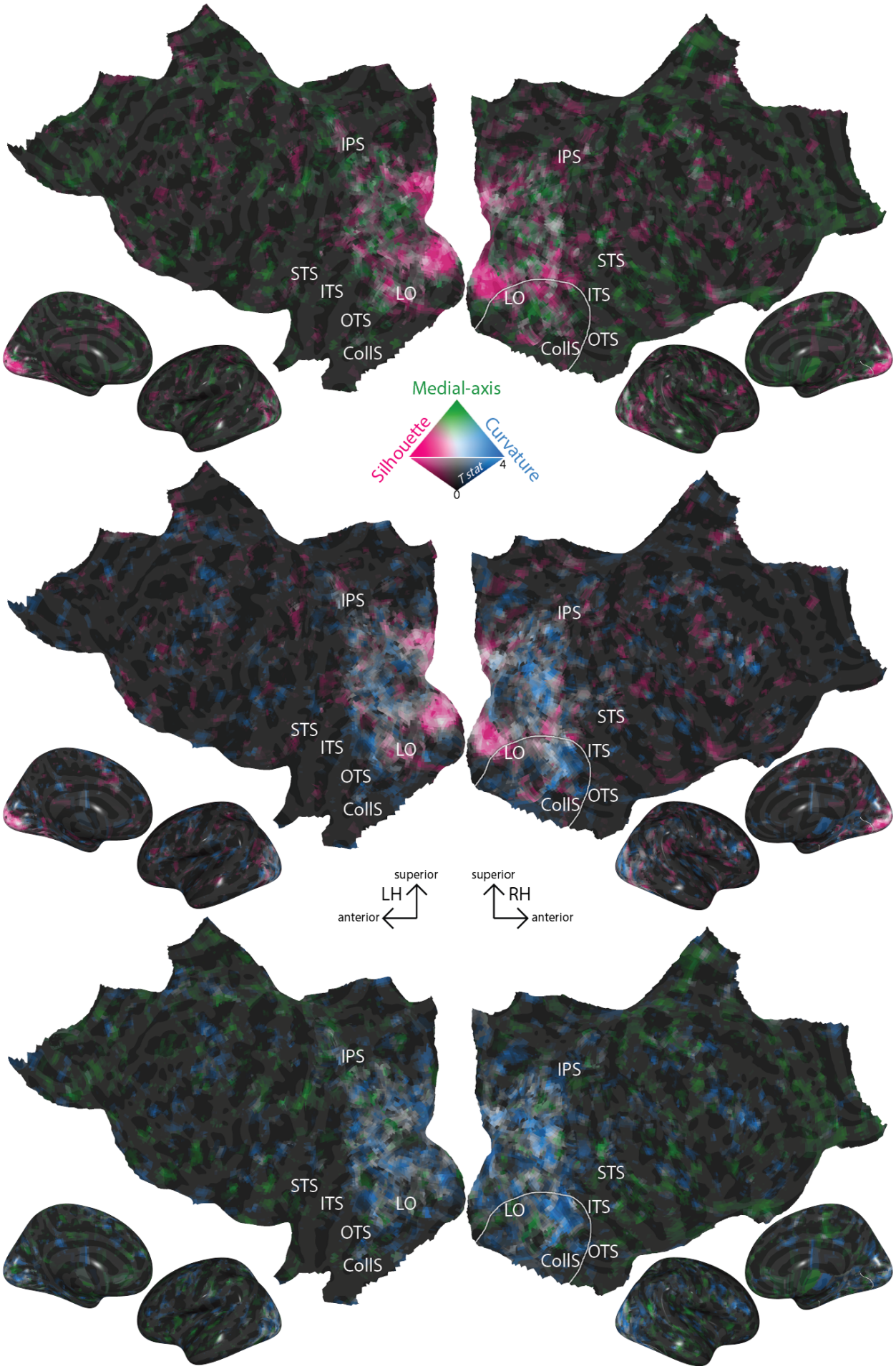


Figure 2. The human visual cortex encodes orthogonal shape representations. Group-level maps showing significant clusters of shape selectivity in OTC and in posterior dorsal regions (one-tailed $p < 0.05$, TFCE corrected). Selectivity to orthogonal components of silhouette (red), medial-axis (green)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

Selectivity to orthogonal shape representations coexist in the same cortical regions

We looked further at the overlap between the selectivity to orthogonal shape representations. Figure 3 depicts the pairwise comparisons between the three shape models in our study. A stronger overlap is observed in LO for medial-axis and curvature, and in IT, right FusG,

297 Cun, right pMTG and right pIPS for silhouette and curvature. Thus, those brain regions encode
298 multiple shape features, independently from the shared variance between them.
299
300 ** Figure 3 near here **
301


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

302

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, only within right OTC all the three models are significant (enclosed by a white line in Figure 3). Figure 4 depicts right OTC in isolation with 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 results 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 **

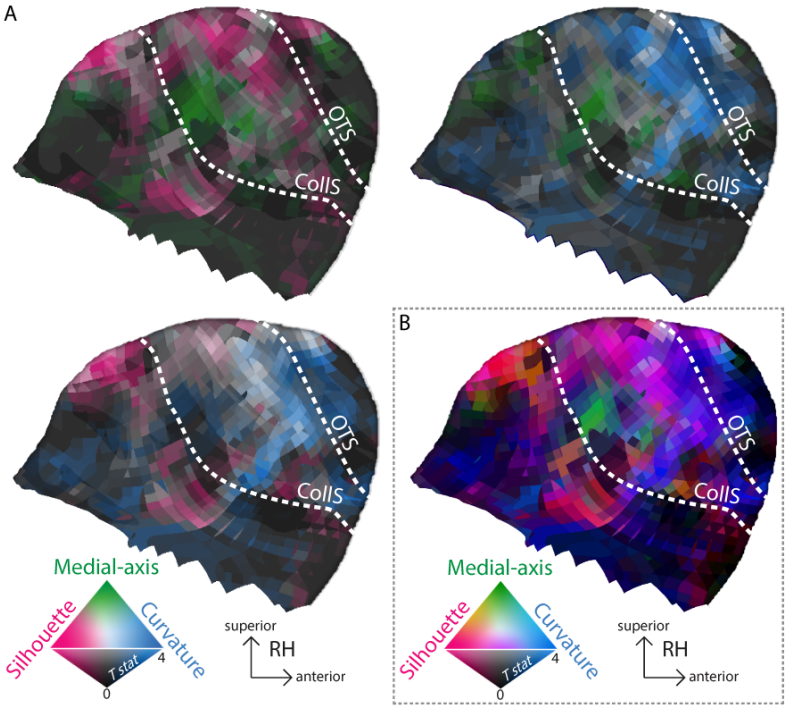


Figure 4. Topographic organization of object shape in right OTC

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).

Coding of orthogonal object representations decreases from posterior to anterior regions

In a previous study, Vernon and colleagues (2016) explored the relationship between retinotopic and more abstract object representations, including contour curvature. They defined two orthogonal components enclosing low-level and complex features, and found a shift between retinotopic and more abstract features in LO. Here, we further tested this aspect by looking also at the relative weight of orthogonal and shared components. Indeed, the tuning to increasingly complex features is considered as the cornerstone of hierarchical object processing (Riesenhuber and Poggio 2000). However, it has been proposed that interaction between features plays a

348 pivotal role in evolving reliable selectivity in the brain (Benjamin et al. 2019). Thus, we
349 hypothesized that shared information should become more relevant along the visual hierarchy,
350 moving from posterior to anterior brain regions.

351 Similarly to Vernon et al (2016), we defined two independent components, one for the low-
352 level features and one for the abstract ones. The first comprised the orthogonal variance of
353 silhouette and inked-area, since both are linked to the local retinotopic arrangement and to the
354 extent of retinotopic cortex stimulated. The second includes the orthogonal variance of medial-
355 axis, curvature (both insensitive to differences in object orientation and size) and object identity
356 models. We computed the ratio between the explained variance of low-level and abstract features
357 (i.e., complexity): values higher than one indicate that brain representations are better accounted
358 for by retinotopic information, while values smaller than one that abstract representations are
359 more relevant. When looking at the slope of complexity along the posterior-to-anterior axis, we
360 observed an abrupt shift from retinotopic to abstract features around Y_{MNI} = -72 (Figure 5A). Of
361 note, the shift occurs at the limit between occipital and temporal or parietal cortex. Indeed,
362 previous studies on ventral temporal cortex selectivity constrained their analysis between Y_{MNI} = -
363 70 and Y_{MNI} = -20 (e.g., Haxby et al, 2001; Rice et al., 2014).

364 Then, we looked at the ratio between orthogonal and shared variance components (i.e.,
365 orthogonality). We summed the variance explained by the orthogonal components of the five
366 models and divided it by the sum of the shared components between the five models: here, values
367 higher than one indicate that brain representations are better explained by orthogonal
368 components of variance. We found no shifts along the posterior-to-anterior axis, instead
369 orthogonality linearly decreases (ρ = -0.83, $p < 0.001$, parametric test; Figure 5B). Thus, while
370 orthogonal information is always more represented than shared variance (min = 2.15), it becomes
371 less relevant proceeding along the visual hierarchy.

** Figure 5 near here **

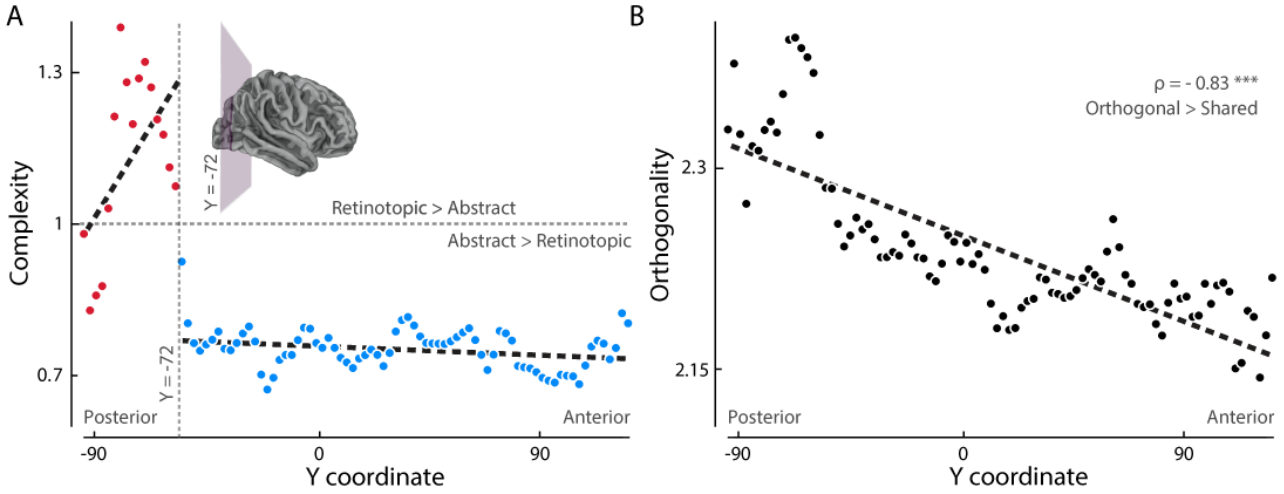


Figure 5. The link between object features shapes the human visual hierarchy
A) The ratio between the explained variance of low-level and abstract features (i.e. complexity) along the visual hierarchy reveals an abrupt shift. Values higher than one (horizontal dashed line) indicate that brain representations 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 found 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 (Figure 2). 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 anterior regions (Figure 5A).

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

395 2014; Proklova et al. 2016). This result can be explained by top-down figure-dependent
396 mechanisms that modulate V1 activity both in monkeys (Poort et al. 2016; Self et al. 2019) and
397 humans (Kok and de Lange 2014; Muckli et al. 2015), and enhances the processing of object-
398 related information in early visual areas also during natural vision (Papale et al. 2018). However,
399 another possibility may be that the silhouette model better captures the object physical
400 appearance (Kubilius et al. 2016).

401 Instead, the variance component unique to the medial-axis model – which is the most
402 transformation-resistant shape description (Yang et al. 2008) – was significant in a smaller extent
403 of cortex comprising only a subset of voxels in right LO (Figure 2, middle in green). This can be due
404 to a higher spatial inter-subject variability of this representation that has been already observed
405 by Leeds *et al.* (2013), or to a higher collinearity with the control models we employed (Figure 1C)
406 that prevents from disentangling its contribution from competing representations. Nonetheless,
407 our result fits previous evidence of medial-axis coding in monkey IT (Hung et al. 2012; putative
408 homologue of human LO), and is consistent with our previous MEG study showing that medial-axis
409 processing is limited to a small cluster of right posterior sensors, when controlling for collinearity
410 with low-level and categorical representations (Papale et al. 2019).

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

419 Closed shapes can be easily and reliably generated by combining simple elements (e.g., geons
420 or medial axes), by connecting few salient points with acute curvature or by modulating its radial
421 frequency. This may suggest that a unique featural dimension – and maybe a single brain region as
422 V4 or LO - could critically account for cortical shape representations. However, the evidence that
423 all the tested dimensions independently contribute to shape representation in the human visual
424 cortex favors the hypothesis of a multi-dimensional coding of object shape (Silson et al. 2016;
425 Silson et al. 2013), similarly to what observed for texture processing (Okazawa et al. 2015; Ziemba
426 et al. 2016).

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

438 This observation complements what has been already observed on the two extremes of
439 the ventral visual pathway: V1 and IT. Representations in V1 are over-complete relative to the
440 retinal input (Olshausen and Field 1996; Vinje and Gallant 2000). In addition, inhibitory
441 interactions in V1 are specifically targeted at neurons with similar tuning properties (Chettih and
442 Harvey 2019). Both these factors increase V1 representational capacity and may ultimately lead to

443 a higher selectivity to orthogonal features, as we observed in posterior regions. On the other
444 hand, higher sensitivity to shared information in more anterior areas may be produced by
445 populations of neurons that are not tuned to a specific property but that encode multiple
446 dimensions at once. Indeed, shared featural selectivity has been proposed as the mechanism
447 responsible to achieve dimensionality reduction of the sensory input in IT (Lehky et al. 2014),
448 where both neural density and surface are much lower than in V1 (Cahalane et al. 2012; Van Essen
449 et al. 1992). In line with this, the highest dimensional among our three shape models (i.e.,
450 silhouette) is also represented in posterior regions (Figure 2). Relatedly, the interaction between
451 multiple features is thought to represent the optimal solution to increase the sensitivity to their
452 mutual changes: in this view, instead of having few neurons encoding a single feature each, it may
453 be preferable to have most of the neurons encoding multiple features at once (Benjamin et al.
454 2019). It has been also suggested that interactions between features are responsible for the poor
455 reliability of tuning curves in predicting brain responses in natural vision (Benjamin et al. 2019).

456 Thus, what can be concluded on the nature of object processing? On one hand, we
457 observed an abrupt shift from retinotopic to abstract representations moving anteriorly across the
458 brain (Figure 5A). However, this shift is relative: though less relevant, orthogonal retinotopic
459 information spreads also to OTC, explaining a significant portion of its variance, in line with
460 previous work and suggesting a link between low-level and object selectivity (Rajimehr et al.,
461 2011; Rice et al., 2014). On the other hand, we found a linear dependency between the anterior-
462 to-posterior axis and the variance explained by shared information (Figure 5B). As stated earlier,
463 this property describes the linear cascade of computations in the visual hierarchy better than
464 complexity: optimizing the coding of shared variance between behaviorally relevant features may
465 represent a key factor in shaping the architecture of our visual cortex and achieving reliable, view-
466 point invariant object representations. In this light, the next step should be to move from

467 modeling representational geometries to more direct modulations of brain responses, so to
468 control also for nonlinear interactions between features (Benjamin et al. 2019).
469 It should be noted that due to the low fMRI temporal resolution we cannot resolve which
470 mechanisms support the different tuning for shared representations. Moreover, while the
471 selected models capture visual transformations, many alternative descriptions exist (e.g., Khaligh-
472 Razavi and Kriegeskorte 2014). Overall, however, our results hint at the existence of a multi-
473 dimensional coding of object shape, and reveal that selectivity for shared object representations
474 are topographically arranged and increases along the visual hierarchy. Future research will identify
475 how different tasks (e.g., determining object similarity vs. extracting affordances), and alternative
476 descriptions impact on the observed patterns of selectivity.
477

478 **Acknowledgments**

479 This work has been supported by the Italian Ministry of Education, University and Research
480 grants PRIN 2015WXAXJF and 2015AR52F9.
481

482 **Author contributions**

483 Pa.P., A.L. G.H., L.C. and E.R. conceived the study. A.L., G.H. and L.C. performed experiments.
484 Pa.P., and A.L. analyzed the data. All the authors discussed the results and wrote the manuscript.
485

486 **References**

487

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

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

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

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

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

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

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

501 **Caldara R, Seghier ML, Rossion B, Lazeyras F, Michel C, and Hauert CA**. The fusiform face area is
502 tuned for curvilinear patterns with more high-contrasted elements in the upper part. *Neuroimage*
503 31: 313-319, 2006.

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

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

508 **Coggan DD, Liu W, Baker DH, and Andrews TJ**. Category-selective patterns of neural response in
509 the ventral visual pathway in the absence of categorical information. *Neuroimage* 135: 107-114,
510 2016.

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

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

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

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

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

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

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

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

527 **Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, and Malach R**. Differential processing
528 of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:
529 187-203, 1999.

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

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

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

538 **Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, and Smith SM**. Fsl. *Neuroimage* 62: 782-
539 790, 2012.

540 **Julian JB, Ryan J, and Epstein RA**. Coding of Object Size and Object Category in Human Visual
541 Cortex. *Cereb Cortex* 27: 3095-3109, 2017.

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

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

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

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

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

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

555 **Konkle T, and Caramazza A**. Tripartite organization of the ventral stream by animacy and object
556 size. *J Neurosci* 33: 10235-10242, 2013.

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

559 **Kriegeskorte N, Goebel R, and Bandettini P**. Information-based functional brain mapping.
560 *Proceedings of the National Academy of Sciences of the United States of America* 103: 3863-3868,
561 2006.

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

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

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

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

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

573 **Lescroart MD, and Biederman I**. Cortical representation of medial axis structure. *Cereb Cortex* 23:
574 629-637, 2013.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

616 **Riesenhuber M, and Poggio T**. Models of object recognition. *Nature neuroscience* 3: 1199-1204,
617 2000.

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

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

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

625 **Silson EH, McKeefry DJ, Rodgers J, Gouws AD, Hymers M, and Morland AB**. Specialized and
626 independent processing of orientation and shape in visual field maps LO1 and LO2. *Nat Neurosci*
627 16: 267-269, 2013.

628 **Smith SM, and Nichols TE**. Threshold-free cluster enhancement: addressing problems of
629 smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44: 83-98,
630 2009.

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

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

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

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

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

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

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

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

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

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