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

Authors and affiliations: Paolo Papale\textsuperscript{1,2,*}, Andrea Leo\textsuperscript{1,2}, Giacomo Handjaras\textsuperscript{1}, Luca Cecchetti\textsuperscript{1}, Pietro Pietrini\textsuperscript{2} and Emiliano Ricciardi\textsuperscript{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: Pablo Papale: pablo.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. 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 (Neurobehaviorial 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-
A planar imaging sequence (GE-EPI) was used, with TE = 40ms, TR = 2500ms, FA = 90°, 160 volumes with four additional dummy scans, acquisition time 6’50”; 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, 1mm³ 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, $3dmerge$), 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.
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
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). To exemplify, 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., $R^2_{\text{curvature}} =
R^2_{\text{full}} - R^2_{\text{silhouette+medial-axis+inked area+identity}}$). 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).

Correlation distance was used to compute the RDM of fMRI activity patterns in each
searchlight and only voxels pertaining to the cerebral cortex with a probability higher than 50%
were included in the procedure. 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 sum of variance explained uniquely by the five models with the group-averaged sum of variance explained by their shared components for each searchlight; a higher value indicates, therefore, that a higher fraction of variance is explained by individual models, rather than being shared across them. We tested whether a linear trend between the Y coordinate and mean orthogonality in each XZ-slice 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 category as abstract ones. Then, complexity was measured by the ratio between variance uniquely explained either by low-level or abstract models. Thus, within each searchlight, group-averaged 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
Here, three competing shape models were tested. A first description was computed by extracting the silhouette, consisting of a simple stimulus vectorization. The link between shape silhouette and OTC representations has been extensively investigated in neuroimaging studies (Bracci and Op de Beeck 2016; Kaiser et al. 2016; Khaligh-Razavi and Kriegeskorte 2014; Proklova et al. 2016). Second, a skeletal representation of each stimulus was extracted by performing the *medial axis* transform (Blum 1973). It controls the spike rate of IT neurons in monkey (Hung et al. 2012), captures behavioral ratings of shape similarity (Lowet et al. 2018) and its spatiotemporal association with brain activity in humans has been described in several neuroimaging studies (Handjaras et al. 2017; Leeds et al. 2013; Lescroart and Biederman 2013; Papale et al. 2019). A third description was obtained by computing the *curvature* distribution for each object’s contour. It has been showed that V4 neurons in monkey are selective to a specific degree of curvature (Cadieu et al. 2007; Carlson et al. 2011; Connor et al. 2007). Moreover, the pivotal role of contour curvature in object perception has been extensively demonstrated both by behavioral (Elder and Velisavljevic 2009; Lawrence et al. 2016; Long et al. 2017; Wolfe et al. 1992) and neuroimaging studies in humans (Caldara et al. 2006; Long et al. 2018; Vernon et al. 2016; Yue et al. 2014). In addition, the area (in pixels) of each stimulus was computed to account for the *inked-area* bias – a problem that is almost unavoidable when using complex objects in isolation (but see Bracci and Op de Beeck 2016 for an elegant stimulus design). Finally, to get rid of high-level biases that could affect the performance of the three shape models, object *identity* was included as a further control (Khaligh-Razavi and Kriegeskorte 2014; Kriegeskorte et al. 2008). 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 show moderate-to-high degrees of collinearity (Figure 1E). Consequently, we used a method that accounts for multicollinearity before considering the significance of the association of each model with brain activity.
representations. Combining the variance partitioning analysis (Lescroart et al., 2015) and a searchlight procedure (whole brain, 6mm radius: Kriegeskorte et al. 2006), we identified group-level clusters significantly explained by three physiologically-validated shape models independently from competing representations (Figure 1D).

The human visual cortex encodes multiple orthogonal shape representations

Group-level results show both distinct and overlapping clusters of shape selectivity in OTC, mildly extending also to posterior dorsal regions (p < 0.05 one-tailed, TFCE corrected). The silhouette model (Figure 2, in red) shows a significant association with brain representations along the Calcarine sulcus (CalcS), the occipitotemporal sulcus (OTS), the right collateral sulcus (CollS), the right inferior temporal sulcus (ITS), the right fusiform gyrus (FusG), the cuneus (Cun) and in posterior portions of the middle temporal gyrus (pMTG) and intraparietal sulcus (pIPS). The medial-axis (Figure 2, in green) explains a significant portion of unique variance in the right lateral occipital area (LO) only. Finally, curvature (Figure 2, in blue) significantly explains fMRI representational geometries in the left lingual gyrus (LinG), in the bilateral FusG, along bilateral OTS and ITS, along the right CollS, in the right MTG, bilaterally in the Cun and along the right IPS. The significant clusters for the control models are also represented in Figure 2.

As all orthogonal components of our tested models show at least a significant cluster of selectivity, shape representation does not rely on a single feature, but on a multi-dimensional coding scheme.

** 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,
Cun, right pMTG and right pIPS for silhouette and curvature. Thus, those brain regions encode multiple shape features, independently from the shared variance between them.

**Figure 3 near here**
**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
pivotal role in evolving reliable selectivity in the brain (Benjamin et al. 2019). Thus, we hypothesized that shared information should become more relevant along the visual hierarchy, moving from posterior to anterior brain regions.

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

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

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

Finally, IT (Kayaert et al. 2005b; Yue et al. 2014), LO (Vernon et al. 2016) and FusG (Caldara et al. 2006) were bilaterally tuned to contour curvature (Figure 2, bottom in blue), in accordance with previous neuroimaging investigations. Actually, LO has a pivotal role in object processing (Grill-Spector et al. 2001; Grill-Spector et al. 1999; Kourtzi and Kanwisher 2001), as IT in monkeys (Brincat and Connor 2004; Desimone et al. 1984; Kayaert et al. 2005a; Op de Beeck et al. 2001; Tanaka 2003; Zoccolan et al. 2007). In addition, while we focus our discussion on the ventral stream, we also observed few significant clusters in dorsal visual regions (R pIPS; see Figure 2), both for curvature and silhouette, which confirm previous observations (Freud et al. 2017).
Closed shapes can be easily and reliably generated by combining simple elements (e.g., geons or medial axes), by connecting few salient points with acute curvature or by modulating its radial frequency. This may suggest that a unique featural dimension – and maybe a single brain region as V4 or LO - could critically account for cortical shape representations. However, the evidence that all the tested dimensions independently contribute to shape representation in the human visual cortex favors the hypothesis of a multi-dimensional coding of object shape (Silson et al. 2016; Silson et al. 2013), similarly to what observed for texture processing (Okazawa et al. 2015; Ziemba et al. 2016).

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

This observation complements what has been already observed on the two extremes of the ventral visual pathway: V1 and IT. Representations in V1 are over-complete relative to the retinal input (Olshausen and Field 1996; Vinje and Gallant 2000). In addition, inhibitory interactions in V1 are specifically targeted at neurons with similar tuning properties (Chettih and Harvey 2019). Both these factors increase V1 representational capacity and may ultimately lead to
a higher selectivity to orthogonal features, as we observed in posterior regions. On the other hand, higher sensitivity to shared information in more anterior areas may be produced by populations of neurons that are not tuned to a specific property but that encode multiple dimensions at once. Indeed, shared featural selectivity has been proposed as the mechanism responsible to achieve dimensionality reduction of the sensory input in IT (Lehky et al. 2014), where both neural density and surface are much lower than in V1 (Cahalane et al. 2012; Van Essen et al. 1992). In line with this, the highest dimensional among our three shape models (i.e., silhouette) is also represented in posterior regions (Figure 2). Relatedly, the interaction between multiple features is thought to represent the optimal solution to increase the sensitivity to their mutual changes: in this view, instead of having few neurons encoding a single feature each, it may be preferable to have most of the neurons encoding multiple features at once (Benjamin et al. 2019). It has been also suggested that interactions between features are responsible for the poor reliability of tuning curves in predicting brain responses in natural vision (Benjamin et al. 2019).

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

It should be noted that due to the low fMRI temporal resolution we cannot resolve which
mechanisms support the different tuning for shared representations. Moreover, while the
selected models capture visual transformations, many alternative descriptions exist (e.g., Khaligh-
Razavi and Kriegeskorte 2014). Overall, however, our results hint at the existence of a multi-
dimensional coding of object shape, and reveal that selectivity for shared object representations
are topographically arranged and increases along the visual hierarchy. Future research will identify
how different tasks (e.g., determining object similarity vs. extracting affordances), and alternative
descriptions impact on the observed patterns of selectivity.

Acknowledgments

This work has been supported by the Italian Ministry of Education, University and Research

Author contributions

Pa.P., and A.L. analyzed the data. All the authors discussed the results and wrote the manuscript.

References

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


