

The functional architecture of the ventral temporal cortex and its role in categorization

Kalanit Grill-Spector^{1,2} and Kevin S. Weiner¹

Abstract | Visual categorization is thought to occur in the human ventral temporal cortex (VTC), but how this categorization is achieved is still largely unknown. In this Review, we consider the computations and representations that are necessary for categorization and examine how the microanatomical and macroanatomical layout of the VTC might optimize them to achieve rapid and flexible visual categorization. We propose that efficient categorization is achieved by organizing representations in a nested spatial hierarchy in the VTC. This spatial hierarchy serves as a neural infrastructure for the representational hierarchy of visual information in the VTC and thereby enables flexible access to category information at several levels of abstraction.

Visual categorization and recognition

Determining, from the visual input, what it is that we see. These processes involve multiple levels of abstraction: exemplar ('my car'); subordinate category ('Volkswagen Beetle'); basic category ('car') and superordinate category ('vehicle').

Ventral temporal cortex (VTC). An anatomical section of the human temporal lobe that includes the fusiform gyrus, parahippocampal gyrus and their bounding sulci.

Visual perception is amazingly rapid, and this enables humans to categorize the visual scene in just one-tenth of a second^{1,2}. This speed is particularly remarkable given that visual categorization and recognition require a series of processing stages over a dozen cortical regions, which together constitute the ventral visual processing stream³. This ventral stream emerges from the primary visual cortex (area V1), continues through a series of retinotopically organized visual areas (V2, V3, human V4) and eventually reaches the ventral temporal cortex (VTC) (BOX 1), where high-level visual regions reside. These high-level visual regions do not process local, low-level features of visual stimuli, such as contrast or orientation, but instead process global shape and are involved in visual perception and recognition^{4–6}. Lesions to the VTC can cause various forms of agnosia^{7–10} depending on the location and extent of the lesion, which supports the idea that the VTC has a key role in visual recognition.

A large body of research has examined the information content within the human VTC and has indicated that it contains information about colour^{11–14}, eccentricity bias^{15–17}, visual field maps^{11,18}, specific domains^{19,20}, expertise²¹, object categories^{22,23}, concepts²⁴, semantics^{25,26} and real-world object size²⁷. However, researchers still lack a computational understanding of how the human VTC anatomically organizes information and uses it for efficient categorization. Recent findings have started to uncover the anatomical features of the human VTC, including its microarchitecture^{28,29}, white matter connectivity^{30–32} and macroarchitecture^{17,33,34}. This provides a new opportunity

to examine the functional architecture of the human VTC — specifically, to directly link the structural architecture of this cortical expanse to the computations that it performs and to the information that these computations provide. Although the VTC is a large cortical expanse and is likely to be involved in more than one function, here we consider the neural mechanisms that underlie one of its key functions: visual categorization.

To understand the functional architecture of the human VTC and its role in visual categorization, we adapted David Marr's approach for understanding information-processing systems³⁵ in order to make it applicable to modern neuroscience methods. Marr proposed that, to fully understand a process such as visual categorization, it is necessary to study three levels of the system: computation, representation and neural implementation (BOX 2). The organization of this Review follows these three levels of analysis, addressing three key questions. First, what are the computational goals of the VTC? Second, what kinds of representations in the VTC support these computations? And third, how are these representations and computations physically implemented in the VTC? After examining each level of analysis separately, we apply an integrated systems perspective^{36,37} to link the information content of functional representations to their physical implementation in the cortex. We conclude by synthesizing findings across analysis levels and hypothesize that the implementation of functional representations in the human VTC is particularly optimized to support rapid and flexible visual categorization.

¹Department of Psychology, Stanford University, Stanford, California 94305, USA.

²Stanford Neuroscience Institute, Stanford University, Stanford, California 94305, USA.

Correspondence to K.G.-S.
e-mail: kalanit@stanford.edu

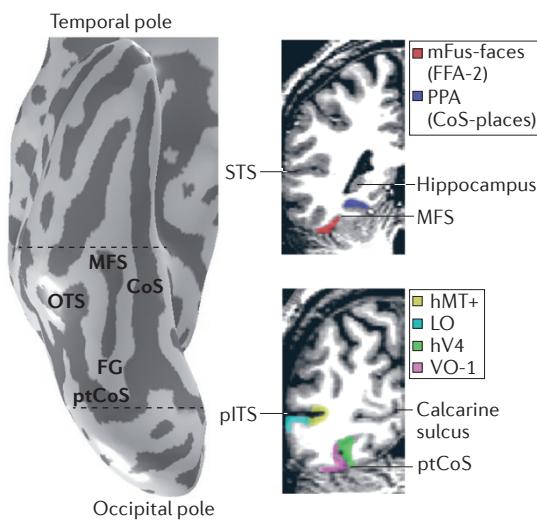
doi:10.1038/nrn3747

Published online
25 June 2014

Box 1 | The boundaries of the ventral temporal cortex

The lateral, posterior, medial and anterior boundaries of the ventral temporal cortex (VTC) are defined by the occipitotemporal sulcus (OTS), posterior transverse collateral sulcus (ptCoS), parahippocampal gyrus (PHG) and the anterior tip of the mid-fusiform sulcus (MFS), respectively (see the figure; dashed lines on the left indicate the location of the coronal slices shown on the right). The MFS bisects the fusiform gyrus (FG) longitudinally; its anterior tip is located approximately halfway between the temporal and occipital poles and aligns with the posterior end of the hippocampus (visible on the coronal slice, top right). The ptCoS is arranged transversely to the posterior end of the CoS and forms the posterior boundary of the FG. The ptCoS and MFS also serve as landmarks for functional distinctions. The ptCoS identifies the boundary between human visual cortex area V4 (hV4)¹¹⁴ and ventral occipital area VO-1 (REF. 34), whereas the anterior MFS identifies the mid-fusiform face-selective region (mFus-faces¹⁷; also known as fusiform face area-2 (FFA-2)¹⁹). The VTC is anatomically and functionally distinct from the lateral occipitotemporal cortex (LOTC)^{62,113,145}. Travelling along the cortical ribbon, the LOTC is several centimetres away from the VTC.

Although the VTC and the LOTC both contain regions that are selective for objects, faces, bodies and places, the LOTC — but not the VTC — contains regions selective for visual motion (the human motion-selective complex, hMT⁺¹⁴⁶), LO, lateral occipital (a functionally defined object-selective region)¹⁰⁰; pITS, posterior inferior temporal sulcus; PPA, parahippocampal place area²⁰; STS, superior temporal sulcus.


What are the computational goals of the VTC?

Computational theory suggests that a visual categorization system, such as the VTC, should be able to generalize across exemplars of a category while maintaining specificity to distinguish among exemplars from different categories. A visual categorization system should also provide separable category information as well as offer flexible access to category information at several levels of abstraction (FIG. 1). Below, we consider the computational requirements that are associated with each of these goals. The purpose of this section is to discuss the goals of the computations within the VTC, not to derive a mathematical formulation of these computations.

Generalization and specificity. A main computational challenge for the VTC is to achieve generalization among exemplars in order to enable robust and accurate categorization while maintaining specificity in order to distinguish between exemplars of different categories that may have similar appearances^{35,38} (FIG. 1a). To do so, the system must overcome three forms of variability: variability in the type of visual information available (for example, luminance, colour, texture, motion and stereopsis); variability in the appearance of a given exemplar that results from transformations induced by

viewing conditions (for example, changes in the illumination, distance, location and viewpoint of the object); and variability in the appearance of different exemplars of a category. Computations in the VTC should display tolerance to all three sources of variance and should be able to categorize new, never-before-seen visual stimuli. However, the categorization system should not over-generalize, as it has to determine not only which exemplars are members of the category but also which stimuli are not members of the category.

Computational modelling of object recognition has provided two important insights. First, the visual recognition system needs linear and non-linear operations^{39–41} in order to achieve specificity and generalization, respectively. Linear operators build new features, whereas non-linear operations (such as a maximum operator) increase generalization across a transformation. Second, a sequence of computations along a processing hierarchy (BOX 2) increases tolerance to transformations^{40–42} and generates more-complex features that are useful for recognition and categorization^{40,41,43}.

Efficiency through separable category information. An additional computational goal is to achieve rapid and efficient categorization. In other words, category information should be easy to determine (or to ‘read out’). One way to achieve this computational goal is to have separable representations. If representations of exemplars of different categories are separable — that is, representations are more similar within a category than across categories (FIG. 1b) — it is possible to implement a simple linear classifier (illustrated by the dashed line in FIG. 1b) to determine which exemplars belong to which category in a fast and biologically plausible way^{44,45}. One computational insight gleaned from the idea of separability is that categorical representations do not need to be completely tolerant to all of the factors that influence the variability between category exemplars but that transformation information should be independent from category information⁴⁶. One way to think about this idea is by describing the representation of all exemplars of a category as a manifold in a high-dimensional space^{46,47} (where the dimensionality is the number of independent neurons (or processing units)). If this manifold is linearly separable from manifolds that describe exemplars of other categories, then category information can be read out accurately and efficiently^{46–48}. Recent computational models have suggested that the ventral stream hierarchy ‘untangles’ representations through a cascade of processing steps, so that representations of objects and categories that are inseparable (tangled) at the initial processing stage (for example, V1) become separable (untangled) at the last stage of the hierarchy⁴⁸.

Flexible access to category information. An ideal visual recognition system would also enable flexible extraction of category information at multiple levels of abstraction. It has been suggested⁴⁹ that humans most readily extract basic-level category information (for example, cars and faces). Nevertheless, the visual system may want to extract different levels of category information from

Agnosia

A condition characterized by a loss of the ability to recognize objects, people or shapes but in which basic visual acuity and memory are preserved.

Eccentricity bias

A preference for particular eccentricities, such as the centre or periphery.

Tolerance

The ability to generalize across a transformation (such as size, position, illumination or view) that affects the appearance of an exemplar.

Separable representations

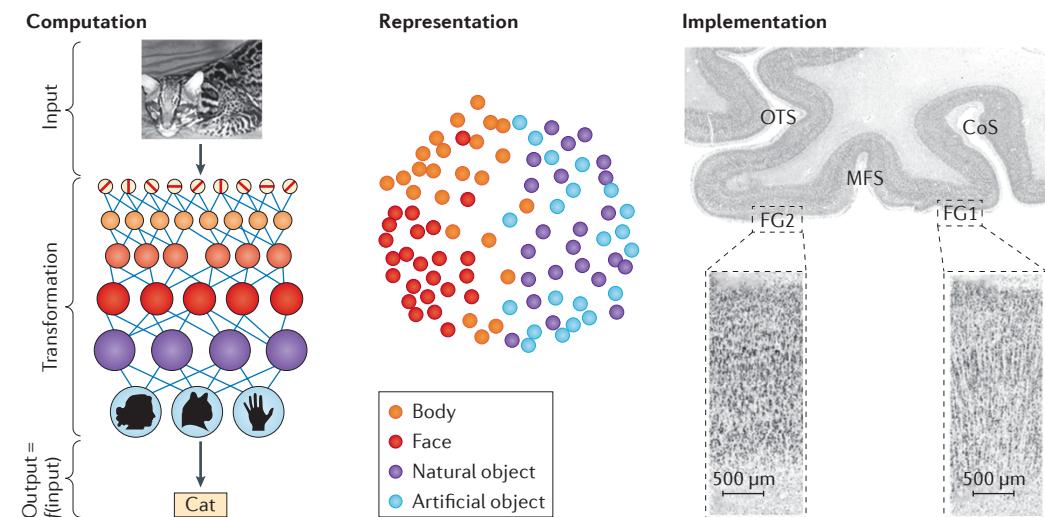
Representations that can be divided by a linear boundary.

Box 2 | David Marr's framework adapted to neural investigations of the ventral temporal cortex

David Marr³⁵ established the field of computational vision. He proposed that to understand a complex information-processing system such as the visual system, one must analyse the system at three levels. The first is the computational level (see the figure, left panel). Analysis at this level aims to establish the goal of the system; in other words, what does the system aim to do? The second level is the representational and algorithmic level (see the figure, middle panel). Analysis at this level aims to establish the method by which the system achieves its goal; in other words, what are the representations that support the computational goals and what algorithms does the system use to transform input information into output information? The third level is the implementational level (see the figure, right panel). Analysis at this level aims to establish the physical substrates of the computations (level 1) and representations (level 2); in other words, how are they implemented in the brain?

We adapt Marr's framework to accommodate data and measurements of the visual system that have been obtained by modern neuroscience methods. For example, in considering Marr's second level, we consider representations but not algorithms, as current experimental methods enable measurements of neural representations but not mathematical algorithms. The three levels are linked because the implementation (level 3) may be specialized to solve a particular computational goal (level 1) using specific representations (level 2).

CoS, collateral sulcus; FG, fusiform gyrus; MFS, mid-fusiform sulcus; OTS, occipitotemporal sulcus. Left panel is adapted with permission from REF. 41, copyright (2007) National Academy of Sciences, U.S.A. Middle panel is reprinted with permission from REF. 23, Cell Press (Elsevier). Right panel, top part, is reprinted with permission from REF. 17, Elsevier, and the bottom part is adapted with permission from REF. 28, Springer.

**Basic-level**

The mid-level (typically entry-level) of the category hierarchy; members have the most shared features and are most distinct from other categories (for example, car versus face).

Superordinate-level

The broadest level of the category hierarchy. It has a high degree of generality; members share fewer attributes than members of basic-level categories (for example, animate versus inanimate).

Subordinate-level

The most specific level of the category hierarchy; members share more features than members of basic-level categories (for example, Honda Civic versus Toyota Corolla).

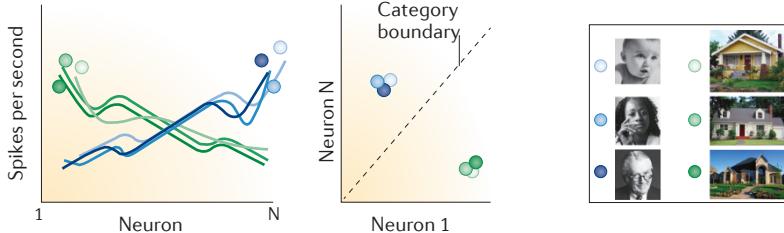
the same visual input depending on the task demands. For example, when shown the two pictures in FIG. 1c, an observer may conclude that one picture depicts an animate object and the second an inanimate object (superordinate-level classification), or determine that one picture shows a face and the other a building (basic-level categorization), or infer that the face pictured is of a man and the building pictured is of a mansion (subordinate-level classification), or identify a particular exemplar (for example, President Barack Obama and the White House). As all levels of information may be valid and relevant in different circumstances, a computational requirement of the VTC is to enable flexible access to multiple levels of class information.

Representations in the VTC

We now examine whether VTC representations support the computational goals of the VTC discussed in the previous section. Two types of category representations are evident in the VTC: clustered regions that respond more strongly to stimuli of one category as compared to stimuli of other categories (which we refer to as

category-selective regions^{19,20,50,51}) and distributed representations⁵², that is, neural patterns of responses across the VTC that are common to exemplars of a category²². We review findings showing that both clustered and distributed representations in the VTC support the computational goals of the VTC as a categorization system: neural responses generalize across format, transformation and category exemplars, and also contain separable category information at different levels of abstraction.

Generalization: format, transformations and exemplars. VTC representations show both generalization and specificity properties, as predicted by computational theory. First, VTC representations maintain their selectivity across a large spectrum of visual information, including luminance, colour, motion, texture, stereo and illusory contours^{53–56}. For example, VTC regions that preferentially respond to a particular category (such as faces or places) maintain higher responses to images of that category (compared to images of non-preferred categories) regardless of the contrast⁵⁷ and format^{58–60} of the images. Furthermore, a considerable body of evidence

a Generalization while maintaining specificity**b Separability****c Flexibility**

Superordinate	Animate	Inanimate
Basic	Face	House
Subordinate	Man	Mansion
Exemplar	President of the United States	White house

Figure 1 | Computational goals of a visual categorization system. **a** | The recognition system should generalize across a range of category exemplars — as well as across format and image transformations — while distinguishing between categories with similar features and configurations (for example, between faces of different species). **b** | To achieve efficient categorization, category information should be easy to read out. One way to achieve this efficiently is to have representations that are linearly separable. Assuming that an exemplar is represented by the distributed responses across a population of neurons, the computational constraint of separability entails that two exemplars of a category will evoke more similar distributed responses across the neural population than two exemplars of different categories (left graph). If this constraint is met, a simple linear classifier can be used to categorize stimuli (right graph). **c** | The recognition system should be able to extract several levels of information from a given input, as required by the task demands; in other words, it should enable flexible access to category information at several levels of abstraction. All photos in parts **a** and **b** courtesy of Getty/PhotoDisc. Barack Obama photo courtesy of Pictorial Press Ltd/Alamy. White House photo courtesy of Getty/PhotoDisc.

shows that VTC responses are primarily driven by the shape^{53,55,56} and content²⁴ of the stimulus rather than by low-level physical properties such as contrast, size or colour (left panel of FIG. 2a), and that they are correlated with subjects' percepts^{4–6,61}.

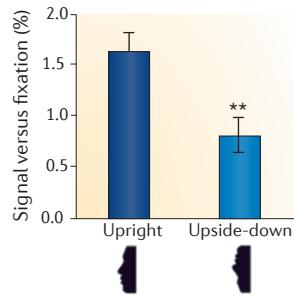
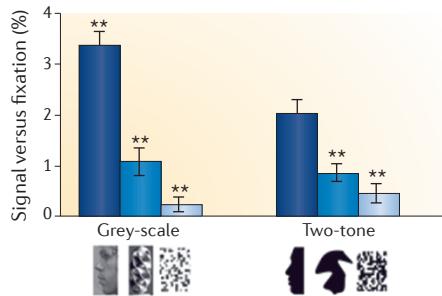
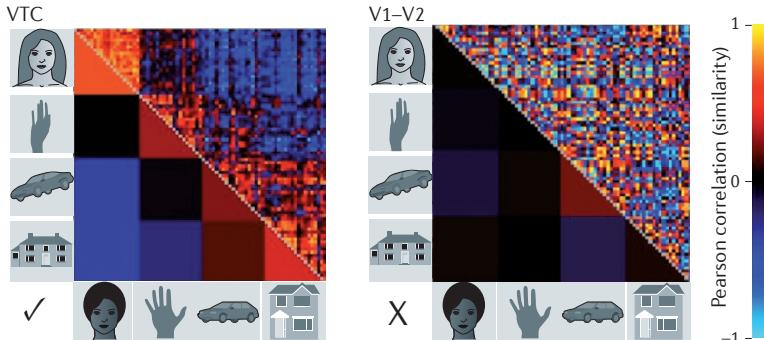
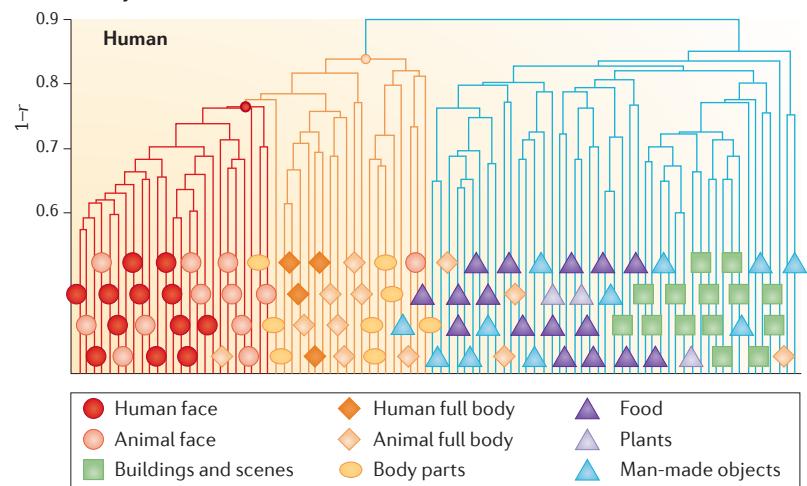
Second, category-selective responses in the VTC are less sensitive to image transformations than are responses in low-level visual areas. However, VTC responses are more tolerant to some image transformations than to others⁶². For example, VTC responses show some tolerance to transformations of position^{62–67}, size^{62,66,68,69} and mirror rotation (at least for faces^{70–72}), whereas they are sensitive to changes in illumination⁶² and viewpoint^{62,71–73}.

Third, VTC responses generalize across exemplars of a category. Functional MRI (fMRI) responses in category-selective regions, such as face-selective regions¹⁹ and place-selective regions²⁰, are consistently higher for exemplars of the preferred category than for exemplars of non-preferred categories^{74,75}. Intracranial recordings from the VTC also show generalization across exemplars, especially for face-selective responses^{76–78} and place-selective responses^{78,79}.

Separability of category information in the VTC. The separability of representations in the VTC is supported by findings showing that category information can be accurately read out from distributed VTC responses by using linear classifiers^{22,23,59,66,78,80}. Such accurate classification of category information in the VTC is possible because distributed response patterns to exemplars of the same category are similar, whereas distributed response patterns to exemplars of different categories are dissimilar (BOX 2; left panel of FIG. 2b). This is not the case in early or intermediate visual areas (V1–hV4) of the ventral stream, in which distributed response patterns do not differentiate between exemplars of different categories. For example, distributed response patterns to exemplars of the same category in early visual areas can be less similar than distributed response patterns to exemplars of different categories^{23,52,80} (right panel of FIG. 2b). Importantly, changes in the distributed response patterns in the VTC induced by transformations are independent of category information, resulting in categorical distinctions that are maintained across transformations of position, size, inversion and mirror rotation^{63,66,70,71,81,82}.

Flexibility: hierarchical category information. Representations in the VTC provide information at several levels of abstraction, ranging from shape information to category information. Intriguingly, accumulating evidence shows that VTC responses represent the perceived similarity rather than the physical similarity among shapes^{83–85}, objects⁵² and categories^{86,87}. This suggests that VTC representations may provide a substrate for perceptual and conceptual mental spaces.

Notably, distributed representations in the VTC generate a hierarchical information structure that mirrors subjects' behavioural judgements of superordinate, basic and subordinate categories⁴⁹. For example, one study²³ described a hierarchical category information structure — ranging from superordinate categories (animate versus inanimate) to basic categories (faces versus bodies) to subordinate categories (human faces versus animal faces) — in both humans and macaques (FIG. 2c). More recent

a Generalization and specificity**b Separability of category information****c Flexibility****Figure 2 | Properties of the ventral temporal cortex representations.**

a | Generalization and specificity. Stronger functional MRI responses to faces are maintained across format (grey level and silhouettes) (left bar chart). Responses are higher for upright silhouettes than for upside-down silhouettes (right bar chart). ** $P < 0.001$, significantly different from upright face silhouettes. Data from REF. 85.

b | Separability of category information in the ventral temporal cortex (VTC) but not early visual cortex (V1–V2). Correlation matrices indicating the similarity between distributed responses to pairs of images from various categories (19 images per category) in the VTC and in V1–V2. In the top triangle, each cell shows the correlation between distributed responses to a pair of images. The bottom triangle shows the average correlation across images of a category. Hot colours indicate similar distributed response patterns and cold colours indicate dissimilar distributed response patterns. Data are from REF. 78 and show electrocorticography measurements in an example subject. **c** | Flexibility. Hierarchical clustering of distributed VTC responses measured with functional MRI reveals a separation between superordinate categories (inanimate versus animate), between basic-level categories (faces versus bodies) and between subordinate categories (human faces versus animal faces). This demonstrates that multiple levels of category information are represented in the VTC. Part **c** is adapted with permission from REF. 23, Cell Press (Elsevier).

research has revealed additional category hierarchies in the VTC, including hierarchies related to biological classes⁸⁷, body parts^{88,89}, scene types^{60,82,88} and semantic information²⁶.

Implementational features of the VTC

Neither computational theory nor research of VTC representations make predictions about how category information should be physically arranged on the cortical sheet. In other words, they do not make specific predictions about the physical implementation of these representations or computations in the brain. However, implementational features are important for two reasons. First, the cortical implementation may be geared towards optimizing particular computations. As such, investigation of the features that underlie the physical implementation of category representations in the cortex will shed light on the computational strategies used by the brain. Second, if one can find a reproducible functional organization in the cortex across individuals, such organization may reflect consistencies in the underlying neural hardware and connectivity. Consequently, elucidating the factors that are responsible for a functional architecture that is consistent across subjects will reveal how specific computations are physically implemented in cortical circuits in a common manner across individuals.

Three key implementational features of the VTC are evident at multiple spatial scales. First, neurons with similar properties are clustered together (FIG. 3a); second, there is a topological organization of functional representations relative to cortical folding patterns and to other functional representations (FIG. 3b); and third, multiple functional representations are superimposed on the same cortical expanse (FIGS 3b,4a).

Clustering of neurons with similar function. A fundamental property of brain architecture is clustering of neurons that have similar properties. Clustering occurs at several spatial scales, from columns (which have a diameter of 200–400 μm in macaques⁹⁰ and 1 mm in humans⁹¹) — which are thought to reflect the basic computational unit of the brain⁹² — to patches⁹³ (0.5 cm diameter), to regions (1 cm diameter) and, finally, to maps — either within a region (for example, a retinotopic map⁹⁴) or across regions (>1 cm diameter; for example, an eccentricity band, which can span several visual areas¹⁵). Clustering is behaviourally relevant: for example, disruptions of normal neural activity by electrical stimulation in face-selective clusters⁹⁵ in the monkey inferotemporal cortex (IT)¹⁴⁸ or face-selective regions in the human fusiform gyrus⁹⁶ generate category-specific perceptual deficits.

There has been significant progress in our understanding of clustering in the VTC at the level of patches and regions. First, category-selective regions (which are generated by clustering of neurons that are selective for a particular category) exist for some but not all categories in both humans and monkeys. Research using fMRI discovered category-selective regions in the monkey IT and the human VTC for a few ecologically relevant categories: faces^{19,80,93,97,98}, body parts^{50,80,97,98}, places^{20,33,98}, and words

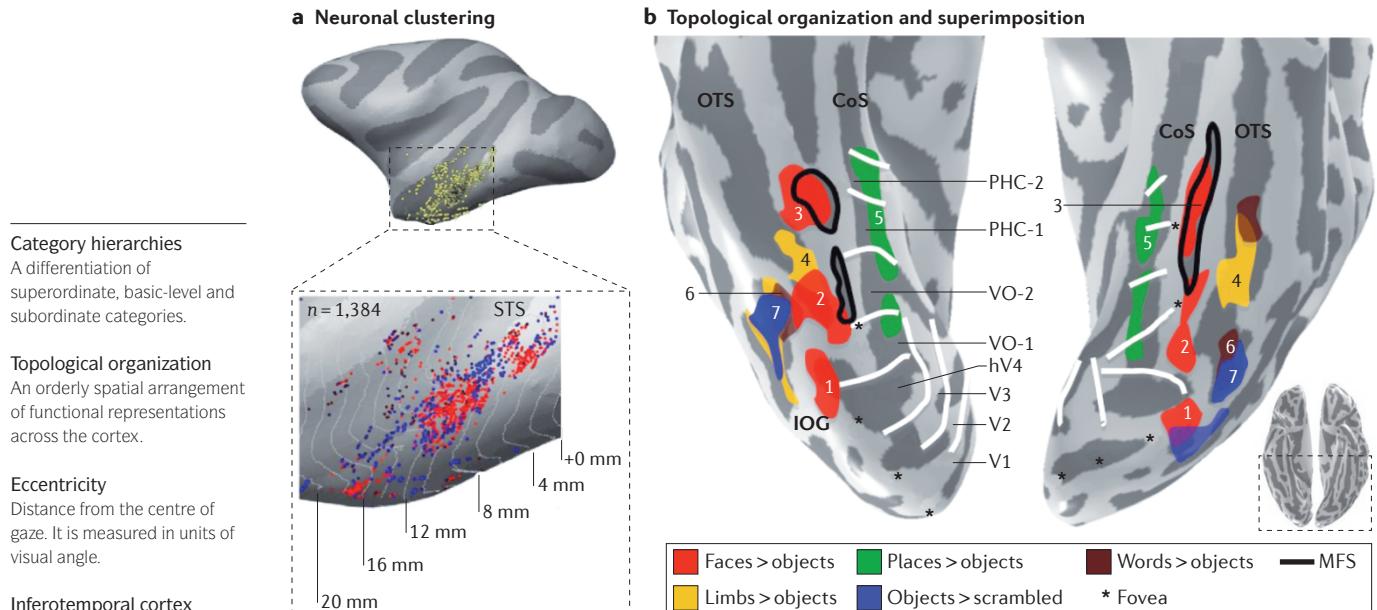


Figure 3 | Three implementational features of the ventral temporal cortex: clustering, topological organization and superimposition. **a** | Neurons with similar category selectivity are clustered together. Each yellow dot indicates the location of a neuron that was recorded. In the enlarged version, red dots indicate individual face-selective neurons, and blue dots represent individual object-selective neurons in the macaque superior temporal sulcus (STS). **b** | Clustered functional regions responding to faces (red), places (green), words (brown), body parts (yellow) and objects (blue) have a consistent topology relative to macroanatomical landmarks in the human ventral temporal cortex (VTC). The mid-fusiform sulcus (MFS) predicts the location of the mid-fusiform face-selective region (mFus-faces (3); also known as FFA-2) and the posterior fusiform face-selective region (pFus-faces (2); also known as FFA-1). The inferior occipital gyrus (IOG) predicts the location of the IOG face-selective region (IOG-faces (1); also known as the occipital face area (OFA)). The occipitotemporal sulcus (OTS) predicts the location of both the occipitotemporal body part region (OTS-limbs (4); also known as the fusiform body area) and the visual word form area (VWFA (6)). The object-selective posterior fusiform/occipitotemporal sulcus (pFus/OTS (7)) partially overlaps with the VWFA and extends more posteriorly. The collateral sulcus (CoS) predicts the location of parahippocampal place area (PPA (5); also known as the CoS place-selective region (CoS-places)). As a result of these structure–function correspondences, there is a consistent topological organization among functional activations. For example, place-selective regions are medial to face-selective regions, whereas OTS-limbs separates pFus-faces from mFus-faces. Notably, within a given macroanatomical neighbourhood in the VTC, multiple representations are superimposed. For example, place-selective representations and retinotopic representations are superimposed along the CoS. hV, human visual area; PHC, parahippocampal; VO, ventral occipital. Data are shown on the inflated right and left hemispheres of a representative subject. Data in part **b** from REFS 80,147. Part **a** is adapted with permission from REF. 102, Society for Neuroscience.

and symbols^{51,99}. In addition, there are regions that process shapes and objects more generally^{56,100} (for example, the lateral occipital complex (LOC) in humans^{56,100,101}, of which the ventral part — the posterior fusiform/occipitotemporal sulcus (pFus/OTS) — is within the VTC (BOX 1; FIG. 3b). Second, measurements in the macaque IT have shown that the proportion of neurons that is selective for a particular category within a category-selective region (as detected using fMRI) is higher than outside this region^{93,98,102} (FIG. 3a). Indeed, 29–97% of the neurons within a cluster detected by fMRI are selective for the preferred category, with the highest proportion (56–97% of neurons) being found in face-selective patches^{93,98}. Third, recent studies using high-resolution fMRI^{21,80,93,103,104} combined with retinotopy^{18,103} and analysis of single-voxel properties within these regions have shown that areas that were originally thought of as category-selective modules with homogeneous properties (for example, the parahippocampal place area (PPA) and the fusiform face area (FFA)) are in

fact composed of several finer-scale regions with different anatomical and functional properties. For example, the PPA overlaps with the peripheral representation of least two visual field maps (parahippocampal areas PHC-1 and PHC-2) that show progressively increased category preference and decreased retinotopic preference^{18,105}, and the FFA contains spatially segregated functional subdomains that differ in their adaptation properties¹⁰⁴ and in their responses to non-face categories, such as body parts⁸⁰, animals and vehicles^{21,25,106,107}.

Topology of representations is tied to cortical folding. Recent findings suggest that two types of topological organization in the VTC are consistent across subjects: first, the location of a functional representation relative to the cortical folds; and second, the spatial relationship among functional representations. Both types of topological organization have been observed for fine-scale functional clusters (FIG. 3b) and large-scale maps (FIG. 4a).

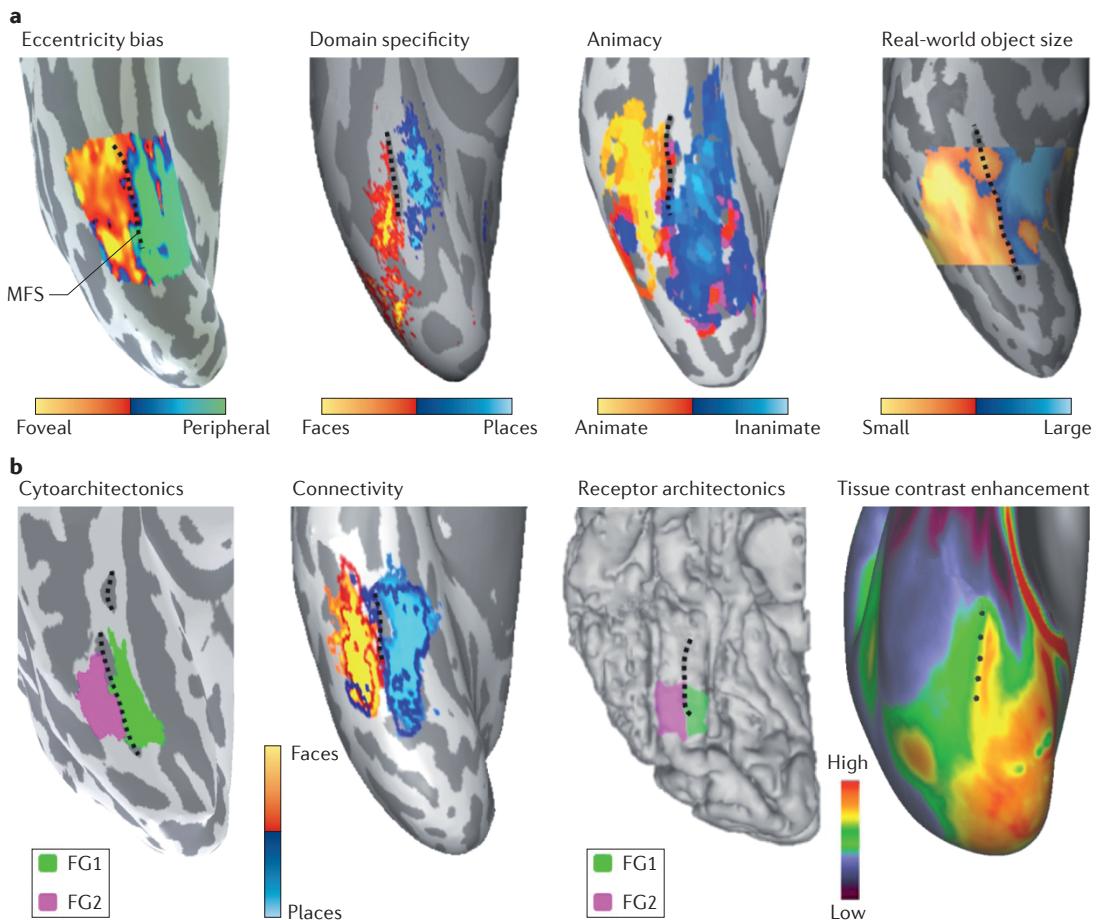


Figure 4 | Linking anatomical features to large-scale functional maps in the ventral temporal cortex. a | The mid-fusiform sulcus (MFS) predicts transitions in many large-scale functional maps in the ventral temporal cortex (VTC). Lateral-medial functional transitions in the eccentricity bias map¹⁵ (based on data from REF. 17), the domain-specificity map (based on data from REF. 33), the animacy map (based on data from REF. 116) and the real-world object-size map (based on data from REF. 27 and T. Konkle, personal communication) are all aligned to the MFS (shown by the dashed black line). Each panel shows a representative inflated right hemisphere from an individual subject, with the exception of the domain-specificity map, which was generated from ten subjects. **b** | The MFS predicts transitions of anatomical features of the VTC. Lateral-medial anatomical transitions in cytoarchitecture, in white-matter connectivity, in the density of muscarinic acetylcholine receptor type 3 and in tissue contrast enhancement (which is thought to be related to myelin content) are each aligned to the MFS. Each panel shows a representative right hemisphere, with the exception of the tissue contrast enhancement map, which is generated from 196 subjects from the Human Connectome Project (based on data from REFS 122,123). FG, fusiform gyrus. The cytoarchitectonics panel is based on data from REF. 17. The connectivity panel is based on data from REF. 30 and Z. M. Saygin, personal communication. The receptor architectonics panel is based on data from REF. 29. Receptor architectonics panel courtesy of J. Caspers, Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Germany. Tissue contrast enhancement panel courtesy of M. F. Glasser and D. C. Van Essen, Washington University, St Louis, Missouri, USA.

Fusiform face area (FFA). A region in the lateral fusiform gyrus that selectively responds to faces compared to other animate or inanimate stimuli. Recent measurements indicate anatomically and functionally distinct divisions of the FFA, which are referred to as posterior fusiform face-selective region (pFus-Faces; also known as FFA-1) and mid-fusiform face-selective region (mFus-faces; also known as FFA-2).

Posterior transverse collateral sulcus (ptCoS). A sulcus that is transverse to the posterior edge of the lateral branch of the CoS and separates the occipital lobe from the temporal lobes.

Mid-fusiform sulcus (MFS). A longitudinal sulcus that bisects the fusiform gyrus.

Consistent topology of functional representations relative to anatomical landmarks and to each other is meaningful: it reveals that particular axes of representational spaces are physically implemented as axes in anatomical space and, furthermore, that anatomical constraints might determine the topology of functional representations. The finding that representational spaces are systematically mapped across the cortical sheet suggests that a particular type of information is arranged in the cortex in a way that is easy to read out. For example, both V1 and middle temporal area MT contain neurons that selectively respond to motion in particular directions¹⁰⁹ (direction-selective neurons), but only MT contains

direction-selective columns¹⁰⁹ that are systematically arranged across the cortical sheet. This cortical topology potentially enables MT (but not V1) to represent motion direction information in an explicit manner.

Recent advancements in the delineation of the macroanatomical features of the VTC — including the morphological features of the CoS¹¹⁰, the posterior transverse collateral sulcus (ptCoS)³⁴ and the mid-fusiform sulcus (MFS)¹⁷ (BOX 1) — have improved our understanding of structure–function relationships in the VTC. For example, irrespective of inter-individual variability in MFS morphology, the anterior tip of the MFS predicts the location of the mid-fusiform face-selective region

(mFus-faces; also known as FFA-2)^{17,80} (FIG. 3b). The MFS also reliably separates face-selective regions from place-selective regions that are consistently localized to the CoS^{33,104} (BOX 1; FIG. 4a), whereas limb-selective regions^{17,80} and word-selective regions are consistently localized to the OTS^{15,111} (FIG. 3b). In addition, the ptCoS predicts a transition between the occipital and temporal eccentricity maps³⁴ as well as the boundary between hV4 and ventral occipital area VO-1 (REF. 34) (BOX 1; FIG. 3b). Together, these recent findings indicate that a surprising number of functional regions, and boundaries that indicate transitions in functional representations, can be predicted from just the cortical folding patterns of the VTC^{17,34}.

The tight coupling between functional representations and cortical folds also generates a consistent topology among functional representations. Within the lateral VTC, a body part-selective cluster on the OTS separates two face-selective regions in the fusiform gyrus, whereas object- and word-selective regions partially overlap and extend posteriorly to face-selective regions (FIG. 3b). These are just two examples, but each functional representation in the VTC — whether it is a map or a region — has consistent anatomical and functional topological boundaries. This tight structure–function coupling is most apparent when one examines these relationships at the level of individual subjects, as data acquisition (for example, voxel size) and analysis choices (for example, spatial smoothing and group analyses) can influence the measurement of structure–function coupling^{112–114}.

Superimposition of multiple functional representations. A third feature of brain architecture that is manifested in the VTC is the superimposition of multiple functional maps and fine-scale clusters within the same cortical expanse. Superimposition of functional representations may be a necessary consequence of projecting a high-dimensional representational space onto the two-dimensional cortical sheet¹¹⁵. Using movie stimuli and data-mining techniques, Haxby and colleagues¹¹⁶ estimated that the dimensionality of representations in the VTC is between 35 to 50 dimensions. Indeed, in addition to categorical representations, other dimensions of information have been uncovered in the VTC, including eccentricity bias^{15,17,120}, retinotopy^{11,18}, real-world object size²⁷, conceptual knowledge^{24,117–119} and semantics²⁵. In reviewing these findings, we observed that many of these superimposed functional maps have a regular relationship to cortical folding patterns — especially relative to the MFS, which often identifies lateral–medial transitions in these maps (FIG. 4a). Thus, there seems to be a regular structure among representations, in which some representations spatially overlap in a systematic manner (convergent representations) and other representations are consistently spatially segregated (divergent representations)³⁷. For example, lateral to the MFS, in the lateral fusiform gyrus and OTS, representations of animate, face, central and small objects converge, and these representations systematically diverge from representations that lie medial to the MFS, where inanimate, place, peripheral and large-object representations converge^{17,27,33,117,120} (FIG. 4a).

Convergent representations
Superimposition of multiple functional representations on the same cortical location.

Divergent representations
Spatially distinct functional representations in the cortex.

Cytoarchitectonic
The arrangement (for example, columnar), properties (for example, density and cell size) and characteristic layout of neuronal cell bodies in the brain.

We also observed that the spatial convergence and divergence among representations extend to finer spatial scales and can also be partial. For example, in the lateral VTC, the fine-scale clustered representations of faces and body parts^{17,80,113,121} converge with the representation of animate objects^{89,117}, thus generating a nesting of finer-scale representations within a larger-scale representation. Likewise, in the medial VTC, regions that respond more strongly to tools versus animate stimuli as well as regions that respond more strongly to places versus other stimuli^{20,33} are nested within the large-scale inanimate representation¹¹⁷, so that the regions preferring tools occupy the medial fusiform gyrus and the regions preferring places are located more medially in the CoS. In a different example, place-selective representations converge with the ventral occipital (VO-2) and parahippocampal (PHC-1 and PHC-2)¹⁸ retinotopic maps (FIG. 3b). However, this convergence only occurs on the peripheral visual field representation^{18,105}, so that there is a partial convergence between place-selectivity and retinotopic maps. fMRI measurements of convergence provide a voxel-level (2–3 mm) resolution of the relationship among representations. At present, we do not know how these representations are arranged within a voxel. Different representations converging on a voxel might be independently arranged across layers of the cortex or across adjacent columns, or they might converge at the level of a neuron.

Anatomical constraints of functional topologies. What might explain the predictable topologies of superimposed functional maps and clusters relative to cortical folding patterns? Recent evidence indicates that anatomical constraints may underlie the predictable topologies in the VTC. For example, in addition to indicating a functional transition in large-scale maps (FIG. 4a), the MFS also designates a microarchitectural transition that corresponds to the boundary between two cytoarchitectonic regions of the VTC¹⁷ on the fusiform gyrus, namely FG1 and FG2 (REF. 28) (FIG. 4b). FG1 lies medial to the MFS and has a columnar organization, whereas FG2 lies lateral to the MFS, is not columnar and has a higher cell density than FG1 (REF. 28) (BOX 2). Furthermore, transitions in the receptor organization of the fusiform gyrus²⁹, the measures of tissue contrast associated with myelin gradients in the VTC^{122,123} and differential long-range white-matter connections of VTC regions³⁰ all align with the MFS (FIG. 4b).

These data suggest that the surprisingly systematic topological arrangement of functional representations in the VTC relative to the cortical folds may be directly related to the physical implementation of underlying neural hardware properties, such as cytoarchitectonics, receptor architectonics, myelination and white-matter connectivity. It is important to note that the scale of these structure–function correspondences is not 1:1. For example, large-scale maps are larger than the cytoarchitectonic regions (for example, the surface area of the animate representation is greater than that of FG2), which in turn are larger than each functional region (for example, the surface area of FG2 is greater than that of

the pFus face-selective region (pFus-faces; also known as FFA-1). Therefore, it is likely that maps in the VTC contain additional cytoarchitectonic regions that have not yet been identified and that each of the cytoarchitectonic regions that have been identified so far in the VTC contains several functional clusters.

Linking implementation to information

The data reviewed above indicate that the three implementational features of the VTC — clustering, topology and superimposition of functional representations — generate a series of nested functional representations across multiple spatial scales, with regular convergences and divergences. We propose that this spatial hierarchy of representations in the VTC may support its hierarchical information structure (FIG. 2). Below, we develop this hypothesis by relating the convergences and divergences of the functional representations in the VTC (described above) to theories proposing general principles by which anatomical organization may be linked to information processing^{36,37,124}. We then discuss how these implementational features may be optimized to support rapid and flexible visual categorization.

Computational benefits of convergence and divergence. Based on findings in the macaque visual system^{3,36,37}, classic neuroscience theories attempted to explain the multiplicity of cortical regions in the visual system, the apparent parallel processing within^{36,37} and across regions^{3,36} revealed by functional and cytoarchitectonic partitions, as well as the arrangement of cortical connections³⁶. Specifically, these theories built a framework for linking the neuroanatomical and functional implementational features of the visual system known at that time to information processing. An important insight gained from these theories is that convergence and divergence are implementational strategies used by the brain to expedite and increase the diversity of neural computations^{36,37,124}. Importantly, these theories proposed that divergence enables segregation of information and that convergence enables integration of information. Here, we extend this theoretical framework to provide insight into the logic of large-scale organization in the human VTC.

Divergence may speed up neural processing by enabling parallel computations of independent information^{36,37}. Applying this principle to the VTC, person information is independent from place information and the cortical representations of these two types of information are physically segregated by the MFS (FIG. 4a). Divergence may also enable the generation of neural circuits that are optimized for particular computations. For example, the differential neural microarchitecture across the MFS (that is, between cytoarchitectonic regions FG1 and FG2) (FIG. 4b) may reflect dedicated neural hardware for particular computations on different sides of the MFS. In addition, connections among segregated clusters that process related information (for example, white-matter connections among clusters that are selective for stimuli of the same category^{31,32,125,126}) may expedite hierarchical processing of computations within specific networks.

Intermediate complexity features

Visual features that contain more than one low-level feature: for example, a shape with an elaborated contour or a coloured shape.

Convergence and clustering may accelerate neural communication (by placing neurons that process related information in close spatial proximity¹¹⁵) and reduce wiring cost¹²⁷. For example, clustering of face-selective neurons may enable fast communication among neurons that encode facial features, thereby expediting face categorization and recognition. Regularities among representations, even those that have partial convergence (as was shown for place and retinotopic representations¹⁸) may provide a substrate for increasing the diversity of cortical computations: namely, by enabling factorial combinations of multidimensional information^{36,124}. In turn, this may enable isolated or combined sources of information to be accessed flexibly according to task demands.

Nested spatial and representational hierarchies. The organization of VTC representations along the cortical sheet indicates that different kinds of information manifest at different spatial scales (FIG. 5). At the scale of the entire VTC (several centimetres), there is an animacy map. Within the animate component of the map in the lateral VTC, there are multiple representations of ecological categories, such as faces and body parts, each spanning ~1 cm. Within face-selective and body part-selective regions, there are clusters of face-selective and body part-selective neurons; these clusters may constitute columns⁹⁰ (~1 mm) that are differentially tuned to features such as eyes¹²⁸ and face views^{72,129,130}.

We propose that the spatial scale of the representation is linked to its level of abstraction, so that more abstract representations are implemented in larger spatial scales across the VTC (FIG. 5). According to our proposal (from small to large), cortical columns represent intermediate complexity features^{90,131,133}; columns containing features that are shared by exemplars of a category may be clustered into larger regions^{93,102} that generate basic-level representations; and these regions may be arranged together with other clusters to form large-scale maps that generate broader categorical distinctions. Thus, the spatial hierarchy of VTC representations may provide a substrate for the visual system — as well as for downstream areas involved in categorical decision making¹³² — to read out multiple levels of information from multiple spatial scales across the VTC. Access to information at a particular level of abstraction may be obtained by a top-down gating mechanism that enables the readout of information from a particular spatial scale across the VTC.

Our proposal is consistent with hierarchical computational models of visual recognition^{40,41,43}. In these models, a series of layers process increasingly more complex information — from simple features (for example, angled lines) to intermediate complexity features (for example, face or object fragments) to exemplars and categories. The categorization systems described in these hierarchical models are tolerant to changes in appearance, provide separable category information and enable the extraction of different kinds of information from various levels of the hierarchy. Although researchers have traditionally mapped these hierarchies across

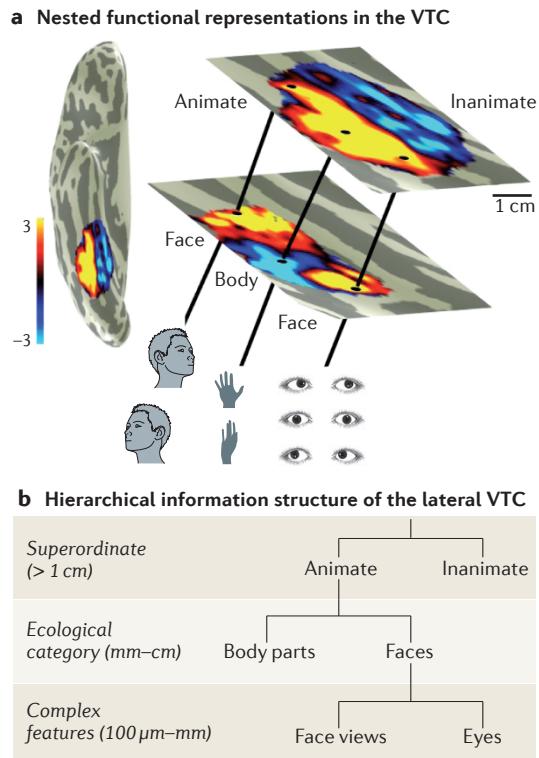


Figure 5 | The spatial structure of nested functional representations in the ventral temporal cortex supports the hierarchical information structure.

a | Superimposition of functional representations in the ventral temporal cortex (VTC) from the animacy map (top) to clustered face-selective regions and body part-selective regions (middle) to clustering of neurons with shared response properties (bottom). **b** | Schematic hierarchy linking the spatial scale of functional representations implemented in the lateral VTC to the scale of information that each level represents. We propose that more-abstract information is represented at a larger spatial scale and more-concrete information at a finer spatial scale. We illustrate this idea with animate hierarchies as an example: superordinate information (animate) is represented at the scale of the entire VTC (several centimetres); information about ecological categories such as faces and body parts is represented at the centimetre scale; and exemplar information and complex-feature information is represented at the columnar level or an even smaller spatial scale. Additional hierarchies are likely to exist in the medial VTC and the VTC more generally.

the ventral visual processing stream^{3,36,37,133}, we speculate, on the basis of the findings reviewed here, that some aspect of this computational hierarchy may be implemented at different spatial scales of the representational hierarchy within the VTC itself (FIG. 5). Indeed, a recent synthesis of anatomical connections of the macaque IT¹³⁴ pointed to the existence of additional hierarchies of nested white-matter connections within IT beyond the classical connections of the ventral pathway. These data provide empirical evidence for a connectivity substrate that might support such a hierarchical organization within the VTC.

Conclusions and future directions

In this Review, we have synthesized the current knowledge of the human VTC across three levels of analysis: computation, representation and implementation. Our synthesis has revealed three important insights regarding how the anatomical structure of the VTC may support the computations and representations that are necessary for efficient visual categorization.

First, although the VTC contains a representational space with hard-to-determine dimensions, functional representations are remarkably orderly in the VTC. We showed that both clustered and distributed functional representations in the VTC maintain a consistent spatial topology that is also reliably aligned to gyri and sulci. Perhaps most surprisingly, irrespective of stimulus dimensions, large-scale maps and fine-scale clusters align with one another. This common spatial layout suggests that when the VTC is faced with the problem of organizing information along the cortical sheet, it often implements the same solution: information is sorted and placed on different sides of the MFS, thus generating a lateral-to-medial functional gradient that is shared across representational dimensions. The common spatial gradient of many functional representations is directly linked to the underlying connectivity and microarchitecture, which also align with the MFS. This relationship was only revealed by clarifying the morphology of the MFS, which suggests that re-examining the macroanatomical features within other parts of the brain may reveal a similar tripartite relationship in other cortical systems. Advancements in *in vivo* anatomical measurements¹³⁵ may eventually enable this hypothesis to be examined in the living human brain. Although we focused on the predictability of functional regions from the cortical folding patterns within the VTC, this implementational feature has been observed in other parts of the visual system, such as V1 (REFS 136,137), V2–V3 (REF. 138), V3A^{139,140}, hV4 (REF. 34), hMT+¹⁴¹ and high-level regions that are selective for places³³, faces¹¹³ and body parts¹¹³ in the lateral occipitotemporal cortex (BOX 1). As in the VTC, each of these regions also displays predictable convergences, divergences and superimpositions with nearby functional maps and clusters. This orderly organization of functional representations across all levels of the visual system raises an important question: if the same representations were implemented physically in a disorganized manner, would this result in less efficient and less flexible computations?

Second, the spatial hierarchy of representations in the VTC generates an information hierarchy. The correspondence between the spatial and information hierarchies predicts that implementing distinct levels of categorical abstraction at different spatial scales increases the efficiency and flexibility of category processing in the VTC. This prediction can be tested empirically and computationally. Examining how the disruption of VTC function at different spatial scales affects different types of categorical decisions will test the prediction regarding efficiency. Comparing the performance of computational models that incorporate implementational features of the VTC with models lacking those features will

test the prediction regarding flexibility. Such approaches can also be used to assess how categorical distinctions in the VTC interact with processing in other regions, such as the prefrontal cortex, which is also involved in categorical decision making¹³². Thus, applying the same framework (computation, representation and implementation) to other brain regions that are involved in categorization may provide crucial knowledge of how information is relayed from the VTC and how computational transformations from one processing stage to the next might be tied to specific implementational features.

Third, the superimposition of representations in the VTC generates a substrate for information integration and segregation through different levels of convergence and divergence. These implementational features enable both fast processing of independent information through parallel computations in divergent regions and fast communication of related information in convergent regions. Although we discussed these computational benefits of segregation and integration in the context of the VTC and its role in visual categorization, convergence and divergence are general organizing principles of the brain^{36,37,124,142}. From an engineering perspective, the brain may implement the same solution across scales and systems because it is composed of the same basic units

(neurons) and faced with the same problem — accommodating a high-dimensional information space across the two-dimensional cortical sheet. As such, the principles of superimposition, convergence and divergence are general neuroanatomical solutions for processing and communicating information. How they are implemented and what computational and representational constraints they resolve may depend on the specific requirements of the individual cortical systems. For example, in V1, the stacking of representations occurs across columns¹⁴³, whereas in the entorhinal cortex the stacking occurs across cortical layers¹⁴⁴. Determining how region-specific features of superimposition accommodate and resolve computational and representational constraints is an important topic for future research.

Unravelling the organizational principles of the human brain and determining how this organization leads to a functionally relevant behaviour is a central goal of systems neuroscience. This review of the functional architecture of the human VTC brings us a step closer to understanding how computations are performed by neurons arranged in particular anatomical circuits and how this anatomical scaffolding organizes the resulting information across spatial scales for efficient visual categorization.

- Thorpe, S., Fize, D. & Marlot, C. Speed of processing in the human visual system. *Nature* **381**, 520–522 (1996).
- Grill-Spector, K. & Kanwisher, N. Visual recognition: as soon as you know it is there, you know what it is. *Psychol. Sci.* **16**, 152–160 (2005).
- Ungerleider, L. G. & Mishkin, M. In *Analysis of Visual Behaviour* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, 1982).
- Tong, F., Nakayama, K., Vaughan, J. T. & Kanwisher, N. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* **21**, 753–759 (1998).
- Grill-Spector, K., Kushnir, T., Hehler, T. & Malach, R. The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neurosci.* **3**, 837–843 (2000).
- Moutoussis, K. & Zeki, S. The relationship between cortical activation and perception investigated with invisible stimuli. *Proc. Natl Acad. Sci. USA* **99**, 9527–9532 (2002).
- Farah, M. J. *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision* (MIT Press, 1990).
- Konen, C. S., Behrmann, M., Nishimura, M. & Kastner, S. The functional neuroanatomy of object agnosia: a case study. *Neuron* **71**, 49–60 (2011).
- Schiltz, C. et al. Impaired face discrimination in acquired prosopagnosia is associated with abnormal response to individual faces in the right middle fusiform gyrus. *Cereb. Cortex* **16**, 574–586 (2006).
- Rossion, B. et al. A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain* **126**, 2381–2395 (2003).
- Brewer, A. A., Liu, J., Wade, A. R. & Wandell, B. A. Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neurosci.* **8**, 1102–1109 (2005).
- Beauchamp, M. S., Haxby, J. V., Jennings, J. E. & DeYoe, E. A. An fMRI version of the Farnsworth–Munsell 100-Hue test reveals multiple colour-selective areas in human ventral occipitotemporal cortex. *Cereb. Cortex* **9**, 257–263 (1999).
- Murphy, D. K., Yoshor, D. & Beauchamp, M. S. Perception matches selectivity in the human anterior colour center. *Curr. Biol.* **18**, 216–220 (2008).
- Bouvier, S. E. & Engel, S. A. Behavioural deficits and cortical damage loci in cerebral achromatopsia. *Cereb. Cortex* **16**, 183–191 (2006).
- Hasson, U., Levy, I., Behrmann, M., Hehler, T. & Malach, R. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* **34**, 479–490 (2002).
- Behrmann, M. & Plaut, D. C. Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends Cogn. Sci.* **17**, 210–219 (2013).
- Weiner, K. S. et al. The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *NeuroImage* **84**, 453–465 (2014). **This study provides crucial evidence for how representational axes are mapped to cortical axes in the VTC. Results show that the MFS predicts both transitions in the functional maps and boundaries between cytoarchitectonic areas. These findings underscore the importance of the MFS, which is not even mentioned in neuroanatomical atlases.**
- Arcaro, M. J., McMains, S. A., Singer, B. D. & Kastner, S. Retinotopic organization of human ventral visual cortex. *J. Neurosci.* **29**, 10638–10652 (2009).
- Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
- Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
- McGugin, R. W., Gatenby, J. C., Gore, J. C. & Gauthier, I. High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proc. Natl Acad. Sci. USA* **109**, 17063–17068 (2012).
- Haxby, J. V. et al. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
- Kriegeskorte, N. et al. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008). **A study in which a data-driven representational similarity approach is used to show that the representational hierarchy of object categories is similar in the human VTC and monkey IT.**
- Chao, L. L., Haxby, J. V. & Martin, A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neurosci.* **2**, 913–919 (1999).
- Cukur, T., Huth, A. G., Nishimoto, S. & Gallant, J. L. Functional subdomains within human FFA. *J. Neurosci.* **33**, 16748–16766 (2013).
- Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210–1224 (2012).
- Konkle, T. & Oliva, A. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* **74**, 1114–1124 (2012).
- Caspers, J. et al. Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Struct. Funct.* **218**, 511–526 (2013).
- Caspers, J. et al. Receptor architecture of visual areas in the face and word-form recognition region of the posterior fusiform gyrus. *Brain Struct. Funct.* <http://dx.doi.org/10.1007/s00429-013-0646-z> (2013).
- Saygin, Z. M. et al. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature Neurosci.* **15**, 321–327 (2012). **The authors use a novel methodology to show that functionally defined regions in the VTC can be defined from their fingerprint of white-matter connections to the rest of the brain. They show that structure–function relationships in the VTC are so consistent that connectivity in one group of subjects can predict the functional organization of the VTC in a separate group of subjects.**
- Pyles, J. A., Verstynen, T. D., Schneider, W. & Tarr, M. J. Explicating the face perception network with white matter connectivity. *PLoS ONE* **8**, e61611 (2013).
- Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D. & Vuilleumier, P. White-matter connectivity between face-responsive regions in the human brain. *Cereb. Cortex* **22**, 1564–1576 (2012).
- Nasr, S. et al. Scene-selective cortical regions in human and nonhuman primates. *J. Neurosci.* **31**, 13771–13785 (2011).
- Withoff, N. et al. Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bht092> (2013).
- Marr, D. *Vision: A Computational Approach* (Freeman & Co., 1982). **In this book, published posthumously, David Marr established the field of computational vision.**
- Zeki, S. & Shipp, S. The functional logic of cortical connections. *Nature* **335**, 311–317 (1988). **A review linking the anatomical construction of the visual system to information processing. One of the many key insights is how divergent and convergent connections enable information segregation and integration, respectively.**

37. Van Essen, D. C., Anderson, C. H. & Felleman, D. J. Information processing in the primate visual system: an integrated systems perspective. *Science* **255**, 419–423 (1992).
- A review linking anatomical organization of the visual system to information processing through an integrated systems perspective. It discusses the definition of cortical areas and processing streams, as well as modularity, distributed hierarchies and computational flexibility.**
38. Ullman, S. *High-Level Vision: Object Recognition and Visual Cognition* (Bradford Books, 1996).
39. Selfridge, O. G. in *Mechanisation of Thought Processes. Proceedings of a Symposium held at the National Physical Laboratory on 24th, 25th, 26th and 27th November 1958 Vol. 1* 513–526 (H.M. Stationery Office, 1959).
40. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**, 1019–1025 (1999).
41. Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *Proc. Natl Acad. Sci. USA* **104**, 6424–6429 (2007).
42. Fukushima, K. Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Networks* **1**, 119–130 (1982).
43. Epshtain, B., Lifshitz, I. & Ullman, S. Image interpretation by a single bottom-up top-down cycle. *Proc. Natl Acad. Sci. USA* **105**, 14298–14303 (2008).
44. Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958).
45. Vandewalle, J. & Suykens, J. A. K. Least squares support vector machine classifiers. *Neural Process. Lett.* **9**, 293–300 (1999).
46. Edelman, S. & Duvdevani-Bar, S. A model of visual recognition and categorization. *Phil. Trans. R. Soc. Lond. B* **352**, 1191–1202 (1997).
47. Poggio, T. & Girosi, F. Regularization algorithms for learning that are equivalent to multilayer networks. *Science* **247**, 978–982 (1990).
48. Rust, N. C. & DiCarlo, J. J. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* **30**, 12978–12995 (2010).
- An examination of information transformations across the ventral processing stream in macaques. Results show that as receptive fields increase in size from V4 to IT, neural responses become more selective to feature conjunctions and more tolerant to position and scale.**
49. Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. Basic objects in natural categories. *Cogn. Psychol.* **8**, 382–439 (1976).
50. Peelen, M. V. & Downing, P. E. Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* **93**, 603–608 (2005).
51. Cohen, L. et al. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* **123**, 291–307 (2000).
52. Edelman, S., Grill-Spector, K., Kusnir, T. & Malach, R. Towards direct visualization of the internal shape space by fMRI. *Psychobiology* **26**, 309–321 (1998).
53. Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y. & Malach, R. Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* **21**, 191–202 (1998).
54. Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K. & Tootell, R. B. The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *J. Neurosci.* **19**, 8560–8572 (1999).
55. Kourtzi, Z. & Kanwisher, N. Representation of perceived object shape by the human lateral occipital complex. *Science* **293**, 1506–1509 (2001).
56. Vinberg, J. & Grill-Spector, K. Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J. Neurophysiol.* **99**, 1380–1393 (2008).
57. Avidan, G. et al. Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophysiol.* **87**, 3102–3116 (2002).
58. Ishai, A., Ungerleider, L. G., Martin, A. & Haxby, J. V. The representation of objects in the human occipital and temporal cortex. *J. Cogn. Neurosci.* **12** (Suppl. 2), 35–51 (2000).
59. Spiridon, M. & Kanwisher, N. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* **35**, 1157–1165 (2002).
60. Walther, D. B., Chai, B., Caddigan, E., Beck, D. M. & Fei-Fei, L. Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc. Natl Acad. Sci. USA* **108**, 9661–9666 (2011).
61. Grill-Spector, K., Knouf, N. & Kanwisher, N. The fusiform face area subserves face perception, not generic within-category identification. *Nature Neurosci.* **7**, 555–562 (2004).
- This article shows that neural responses in face-selective regions on the fusiform gyrus are correlated with both the detection and identification of faces but not within-category identification of non-face objects.**
62. Grill-Spector, K. et al. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* **24**, 187–203 (1999).
63. Schwarzlose, R. F., Swisher, J. D., Dang, S. & Kanwisher, N. The distribution of category and location information across object-selective regions in human visual cortex. *Proc. Natl Acad. Sci. USA* **105**, 4447–4452 (2008).
64. Andrews, T. J. & Ewbank, M. P. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage* **23**, 905–913 (2004).
65. MacEvoy, S. P. & Epstein, R. A. Position selectivity in scene- and object-responsive occipitotemporal regions. *J. Neurophysiol.* **98**, 2089–2098 (2007).
66. Liu, H., Agam, Y., Madsen, J. R. & Kreiman, G. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* **62**, 281–290 (2009).
67. Kravitz, D. J., Kriegeskorte, N. & Baker, C. I. High-level visual object representations are constrained by position. *Cereb. Cortex* **20**, 2916–2925 (2010).
68. Eger, E., Schyns, P. G. & Kleinschmidt, A. Scale invariant adaptation in fusiform face-responsive regions. *Neuroimage* **22**, 232–242 (2004).
69. Vuilleumier, P., Henson, R. N., Driver, J. & Dolan, R. J. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neurosci.* **5**, 491–499 (2002).
70. Axelrod, V. & Yovel, G. Hierarchical processing of face viewpoint in human visual cortex. *J. Neurosci.* **32**, 2442–2452 (2012).
71. Kietrys, T. C., Swisher, J. D., Konig, P. & Tong, F. Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. *J. Neurosci.* **32**, 11763–11772 (2012).
72. Freiwald, W. A. & Tsao, D. Y. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**, 845–851 (2010).
73. Epstein, R., Graham, K. S. & Downing, P. E. Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* **37**, 865–876 (2003).
74. Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M. & Kanwisher, N. Domain specificity in visual cortex. *Cereb. Cortex* **16**, 1453–1461 (2006).
75. Mur, M. et al. Categorical, yet graded—single-image activation profiles of human category-selective cortical regions. *J. Neurosci.* **32**, 8649–8662 (2012).
76. McCarthy, G., Puce, A., Belger, A. & Allison, T. Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb. Cortex* **9**, 431–444 (1999).
77. Davidescu, I. et al. Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bht038> (2013).
78. Jacques, C. et al. Electrocorticography of category-selectivity in human ventral temporal cortex: spatial organization, responses to single images, and coupling with fMRI. *J. Vision* **13**, 495 (2013).
79. Bastin, J. et al. Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *J. Neurosci.* **33**, 10123–10131 (2013).
80. Weiner, K. S. & Grill-Spector, K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* **52**, 1559–1573 (2010).
81. Sayres, R. & Grill-Spector, K. Relating retinotopic and object-selective responses in human lateral occipital cortex. *J. Neurophysiol.* **100**, 249–267 (2008).
82. Walther, D. B., Caddigan, E., Fei-Fei, L. & Beck, D. M. Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* **29**, 10573–10581 (2009).
83. Haushofer, J., Livingstone, M. S. & Kanwisher, N. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol.* **6**, e187 (2008).
84. Drucker, D. M. & Aguirre, G. K. Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb. Cortex* **19**, 2269–2280 (2009).
85. Davidenko, N., Remus, D. A. & Grill-Spector, K. Face-likeness and image variability drive responses in human face-selective ventral regions. *Hum. Brain Mapp.* **33**, 2234–2249 (2012).
86. O'Toole, A. J., Jiang, F., Abdi, H. & Haxby, J. V. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* **17**, 580–590 (2005).
87. Connolly, A. C. et al. The representation of biological classes in the human brain. *J. Neurosci.* **32**, 2608–2618 (2012).
- This study examines distributed responses in the VTC and shows that there is a hierarchy of animate classes, ranging from insects, to birds, to mammals.**
88. Op de Beeck, H. P., Brants, M., Baeck, A. & Wagenaars, J. Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. *Neuroimage* **49**, 3414–3425 (2010).
89. Orlov, T., Makin, T. R. & Zohary, E. Topographic representation of the human body in the occipitotemporal cortex. *Neuron* **68**, 586–600 (2010).
90. Fujita, I., Tanaka, K., Ito, M. & Cheng, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* **360**, 343–346 (1992).
91. Adams, D. L., Sincich, L. C. & Horton, J. C. Complete pattern of ocular dominance columns in human primary visual cortex. *J. Neurosci.* **27**, 10391–10403 (2007).
92. Mountcastle, V. B. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* **20**, 408–434 (1957).
93. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006).
94. Engel, S. A. et al. fMRI of human visual cortex. *Nature* **369**, 525 (1994).
95. Afraz, S. R., Kiani, R. & Estekey, H. Microstimulation of inferotemporal cortex influences face categorization. *Nature* **442**, 692–695 (2006).
96. Parvizi, J. et al. Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14915–14920 (2012).
97. Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G. & Kastner, S. Representations of faces and body parts in macaque temporal cortex: a functional MRI study. *Proc. Natl Acad. Sci. USA* **102**, 6996–7001 (2005).
98. Bell, A. H. et al. Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J. Neurosci.* **31**, 12229–12240 (2011).
99. Srihasam, K., Mandeville, J. B., Moroz, I. A., Sullivan, K. J. & Livingstone, M. S. Behavioural and anatomical consequences of early versus late symbol training in macaques. *Neuron* **73**, 608–619 (2012).
100. Malach, R. et al. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl Acad. Sci. USA* **92**, 8135–8139 (1995).
101. Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. The lateral occipital complex and its role in object recognition. *Vision Res.* **41**, 1409–1422 (2001).
102. Issa, E. B., Papanastassiou, A. M. & DiCarlo, J. J. Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. *J. Neurosci.* **33**, 15207–15219 (2013).
103. Weiner, K. S. & Grill-Spector, K. Not one extrastriate body area: using anatomical landmarks, HMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* **56**, 2183–2199 (2011).
104. Weiner, K. S., Sayres, R., Vinberg, J. & Grill-Spector, K. fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across timescales. *J. Neurophysiol.* **103**, 3349–3365 (2010).
105. Witthoft, N., Golarai, G., Nguyen, M., Liberman, A. & Grill-Spector, K. Anatomy, retinotopy, & category selectivity in human ventral visual cortex. *J. Vision* **12**, 1177 (2012).
106. Hanson, S. J. & Schmidt, A. High-resolution imaging of the fusiform face area (FFA) using multivariate nonlinear classifiers shows diagnosticity for non-face categories. *Neuroimage* **54**, 1715–1734 (2011).

107. Grill-Spector, K., Sayres, R. & Ress, D. High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nature Neurosci.* **9**, 1177–1185 (2006).
108. Albright, T. D. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* **52**, 1106–1130 (1984).
109. Albright, T. D., Desimone, R. & Gross, C. G. Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* **51**, 16–31 (1984).
110. Huntgeburth, S. C. & Petrides, M. Morphological patterns of the collateral sulcus in the human brain. *Eur. J. Neurosci.* **35**, 1295–1311 (2012).
111. Yeatman, J. D., Rauschecker, A. M. & Wandell, B. A. Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. *Brain Lang.* **125**, 146–155 (2013).
112. Glezer, L. S. & Riesenhuber, M. Individual variability in location impacts orthographic selectivity in the “visual word form area”. *J. Neurosci.* **33**, 11221–11226 (2013).
113. Weiner, K. S. & Grill-Spector, K. Neural representations of faces and limbs neighbour in human high-level visual cortex: evidence for a new organization principle. *Psychol. Res.* **77**, 74–97 (2013).
114. Winawer, J., Horiguchi, H., Sayres, R. A., Amano, K. & Wandell, B. A. Mapping hV4 and ventral occipital cortex: the venous eclipse. *J. Vis.* **10**, 1 (2010).
115. Kohonen, T. *Self-Organization and Associative Memory* (Springer, 1983).
116. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).
117. Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. Neural correlates of category-specific knowledge. *Nature* **379**, 649–652 (1996).
118. Beauchamp, M. S., Lee, K. E., Haxby, J. V. & Martin, A. Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* **34**, 149–159 (2002).
119. Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M. & Caramazza, A. Category-specific organization in the human brain does not require visual experience. *Neuron* **63**, 397–405 (2009).
120. Levy, I., Hasson, U., Avidan, G., Hendler, T. & Malach, R. Center-periphery organization of human object areas. *Nature Neurosci.* **4**, 533–539 (2001).
121. Weiner, K. S. & Grill-Spector, K. Improbable simplicity of the fusiform face area. *Trends Cogn. Sci.* **16**, 251–254 (2012).
122. Glasser, M. F. & Van Essen, D. C. Mapping human cortical areas *in vivo* based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011).
123. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
124. Malach, R. Cortical columns as devices for maximizing neuronal diversity. *Trends Neurosci.* **17**, 101–104 (1994).
125. Moeller, S., Freiwald, W. A. & Tsao, D. Y. Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* **320**, 1355–1359 (2008).
126. Kornblith, S., Cheng, X., Ohayon, S. & Tsao, D. Y. A network for scene processing in the macaque temporal lobe. *Neuron* **79**, 766–781 (2013).
127. Van Essen, D. C. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature* **385**, 313–318 (1997).
128. Issa, E. B. & DiCarlo, J. J. Precedence of the eye region in neural processing of faces. *J. Neurosci.* **32**, 16666–16682 (2012).
129. Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. Organization and functions of cells responsive to faces in the temporal cortex. *Phil. Trans. R. Soc. Lond. B* **335**, 23–30 (1992).
130. Wang, G., Tanaka, K. & Tanifugi, M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* **272**, 1665–1668 (1996).
131. Tsunoda, K., Yamane, Y., Nishizaki, M. & Tanifugi, M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neurosci.* **4**, 832–838 (2001).
132. Jiang, X. *et al.* Categorization training results in shape- and category-selective human neural plasticity. *Neuron* **53**, 891–903 (2007).
133. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
134. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
- A review linking anatomical features to recurrent processing networks in the macaque ventral visual processing stream. Modern insights are incorporated into the classic understanding of the anatomical and functional construction of the ventral visual pathway across species.**
135. Mezer, A. *et al.* Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. *Nature Med.* **19**, 1667–1672 (2013).
136. Holmes, G. Disturbances of vision by cerebral lesions. *Br. J. Ophthalmol.* **2**, 353–384 (1918).
137. Hasnain, M. K., Fox, P. T. & Woldorff, M. G. Structure-function spatial covariance in the human visual cortex. *Cereb. Cortex* **11**, 702–716 (2001).
138. Benson, N. C., Butt, O. H., Brainard, D. H. & Aguirre, G. K. Correction of distortion in flattened representations of the cortical surface allows prediction of V1–V3 functional organization from anatomy. *PLoS Comput. Biol.* **10**, e1003538 (2014).
139. Tootell, R. B. *et al.* Functional analysis of V3A and related areas in human visual cortex. *J. Neurosci.* **17**, 7060–7078 (1997).
140. Fischl, B. *et al.* Cortical folding patterns and predicting cytoarchitecture. *Cereb. Cortex* **18**, 1973–1980 (2008).
141. Dumoulin, S. O. *et al.* A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cereb. Cortex* **10**, 454–463 (2000).
142. Braintenberg, V. & Schüz, A. *Anatomy of the Cortex* (Springer, 1991).
143. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
144. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M. B. Path integration and the neural basis of the ‘cognitive map’. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
145. O'Toole, A. J., Roark, D. A. & Abdi, H. Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* **6**, 261–266 (2002).
146. Tootell, R. B. *et al.* Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* **15**, 3215–3230 (1995).
147. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Differential rate of temporal processing across category-selective regions in human high-level visual cortex. *Vision Sci. Soc. Abstr.* 23.579 (2014).
148. Gross, C. G., Bender, D. B. & Rocha-Miranda, C. E. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* **166**, 1303–1306 (1969).

Acknowledgements

The authors thank C. Jacques for creating figure 2b, and N. Davidenko for creating figure 2a. The authors thank D. Van Essen, M. Glasser, J. Caspers, S. Nasr, T. Konkle and Z. M. Saygin for providing access to their data and contributing to figure 4. The authors thank A. Connolly, J. S. Guntupalli, J. V. Haxby, E. Issa and N. Kriegeskorte for permission to use their figures. This work was supported by the National Science Foundation, BCS grant 0920865 and National Eye Institute grant NIH 1 RO1 EY 02231801A1.

Competing interests statement

The authors declare no competing interests.

FURTHER INFORMATION

Human Connectome Project: www.humanconnectome.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF