ANNUAL REVIEWS

The Organization and Operation of Inferior Temporal Cortex

Bevil R. Conway[1,2]

[1]Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 28092, USA; email: bevil@nih.gov

[2]National Institutes of Mental Health, National Institute of Neurological Disease and Stroke, National Institutes of Health, Bethesda, Maryland 28092, USA

Keywords

ventral visual pathway, functional organization, brain homologies, face processing, color processing, inferior temporal cortex, inferotemporal cortex

Abstract

Inferior temporal cortex (IT) is a key part of the ventral visual pathway implicated in object, face, and scene perception. But how does IT work? Here, I describe an organizational scheme that marries form and function and provides a framework for future research. The scheme consists of a series of stages arranged along the posterior-anterior axis of IT, defined by anatomical connections and functional responses. Each stage comprises a complement of subregions that have a systematic spatial relationship. The organization of each stage is governed by an eccentricity template, and corresponding eccentricity representations across stages are interconnected. Foveal representations take on a role in high-acuity object vision (including face recognition); intermediate representations compute other aspects of object vision such as behavioral valence (using color and surface cues); and peripheral representations encode information about scenes. This multistage, parallel-processing model invokes an innately determined organization refined by visual experience that is consistent with principles of cortical development. The model is also consistent with principles of evolution, which suggest that visual cortex expanded through replication of retinotopic areas. Finally, the model predicts that the most extensively studied network within IT—the face patches—is not unique but rather one manifestation of a canonical set of operations that reveal general principles of how IT works.

1. INTRODUCTION

> [A] common mechanism of evolution is the replication of body parts due to genetic mutation in a single generation which is then followed in subsequent generations by the gradual divergence of structure and functions of the duplicated parts (Gregory 1935). The replication of sensory representations in the central nervous system may have followed a similar course. This possibility is an alternative to the view that topographically organized sensory representations gradually differentiated from 'unorganized' cortex or that individual topographically organized areas gradually differentiated into additional topographical representations. (Allman & Kaas 1971, p. 100)

The part of the brain implicated in determining the content of vision, such as the objects in a scene, is called the ventral visual pathway (VVP). A key component of the VVP is inferior temporal cortex (IT). In macaque monkeys, IT is on the lateral and ventral surface of the occipital and temporal cortex, tucked behind the ears (**Figure 1**). In humans, IT is pushed around the lateral margin of the temporal lobe onto the ventral and medial surface [see figure 9 of (Lafer-Sousa et al. 2016)]. There are many differences between human and monkey brains (Van Essen et al. 2016). But the visual systems of the two species are similar in terms of psychophysical performance (Gagin et al. 2014, Rajalingham et al. 2015, Stoughton et al. 2012) and anatomy (Passingham 2009). Moreover, the organization of areas early in the cortical visual-processing hierarchy [primary visual cortex (V1), visual area 2 (V2), visual area 3 (V3)] are almost identical in monkeys and humans (Brewer et al. 2002), and the organization of IT across the two species is comparable (Lafer-Sousa et al. 2016, Tsao et al. 2008, Vanduffel et al. 2014). In both humans and monkeys, there exists a distinct transition from regions early in the putative visual-processing hierarchy, which have clear retinotopic organization, and IT, which does not (**Figure 1**).

Like the VVP in humans, IT in macaque monkeys is a large region, encompassing about as much tissue as areas V1, V2, V3, and V4 combined. There is broad consensus that IT is
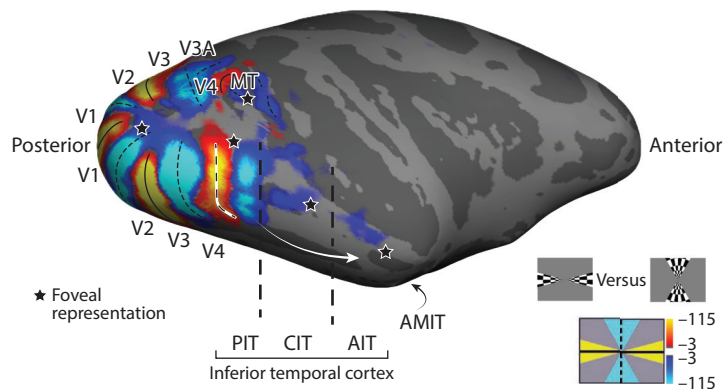

Figure 1

Visual-field meridian representations in alert macaque cortex. The white arrow shows the ventral stream through inferior temporal cortex (IT). The organizational structure of early visual areas [primary visual cortex (V1), visual area 2 (V2), and visual area 3 (V3)], but not higher-order regions (IT), is clearly uncovered by meridian representations. Orange shows regions that responded more strongly to checkered stimuli restricted to a wedge along the horizontal meridian; blue shows regions that responded more strongly to checkered stimuli restricted to a wedge along the vertical meridian (*bottom right* shows the stimuli). The dashed line shows the anterior border of V4. The foveal representation (*stars*; see **Figure 4**) are shown for reference. Figure generated with data from Lafer-Sousa & Conway (2013). Additional abbreviations: AIT, anterior inferior temporal cortex; AMIT, anteromedial inferior temporal cortex; CIT, central inferior temporal cortex; MT, middle temporal area; PIT, posterior inferior temporal cortex.

involved in object vision, face recognition, and scene perception. The region shows strong blood-oxygen-level-dependent signals [measured with functional magnetic resonance imaging (fMRI)] when subjects are shown pictures of objects, faces, and scenes (Grill-Spector & Weiner 2014, Kanwisher 2010). Neurons within IT show stimulus selectivity for images, with some neurons showing striking selectivity for all exemplars of a given category (Freiwald & Tsao 2010, Gross 2008, Lehky & Tanaka 2016). And patients with brain damage to the region often show deficits in object, face, color, or scene vision (Farah 2004, Martinaud 2017). But it has not been clear whether IT comprises one area or many areas or even if the term "area" applies (DiCarlo et al. 2012, Yamins & DiCarlo 2016a). In this review, I develop the idea that IT comprises parallel multistage processing networks governed by a repeated eccentricity template inherited from V1. This idea was first suggested by fMRI experiments that exploited faces and color as tools for assessing brain organization, combined with eccentricity mapping (Lafer-Sousa & Conway 2013). The organizational scheme builds on the importance of eccentricity representations in determining how extrastriate cortex works (Levy et al. 2001) and has accumulated support from several labs, which I describe. The scheme is not mathematical. But insofar as the scheme provides a plan for how IT is organized that gives rise to testable predictions, one might refer to it as a model.

A successful model of IT should show how IT accomplishes its computational objectives, rather than provide just a descriptive account of functional patterns of activation. In Section 2, I discuss the computational objective of IT (what IT is for), so that we can evaluate the explanatory power of various IT models. In Section 3, I describe the history of retinotopy as a fundamental organizing principle of V1—this history gave birth to the idea of visual areas, and it inspired the concept of eccentricity templates as a governing principle. In Section 4, I provide a survey of current alternative cartoons of the organization of IT, which sets the stage for the description in Section 5 of the new scheme. In Section 6, I outline three implications of the new organizational scheme regarding the search for canonical operations in IT, how IT computes semantic categories from visual-shape features, and what determines which complex features are represented in IT. And finally, in Section 6, I touch upon brain mechanisms that decode activity of IT.

2. WHAT IS INFERIOR TEMPORAL CORTEX FOR?

To understand how IT works, it is useful to remind ourselves what IT cortex is probably for, because the brain is likely adapted not only to the statistics of the world (Geisler 2008) but also to the tasks required of it. IT is often said to be important for object vision, although its role is broader than this statement implies. Besides enabling object recognition, IT is likely involved in determining the relevance of objects to behavior in nonverbal (social) communication (face recognition) and in navigating the environment.

What is an object? The term captures both scale—a banana is an object; the sky is not—and active engagement. Essential to the concept of an object and its scale dependence is an interaction between the agent and the visual world (Weisberg et al. 2007). Objects can be grasped. They are things we want, or want to avoid. Visual signals associated with objects tell us about both identity and value, information that is often separable. Consider a banana: As it ripens, its shape (identity) changes little, but its color (behavioral relevance) changes substantially (**Figure 2a**) (Conway 2016). Faces are a kind of object that are obviously important for social species such as primates. As with other objects, information about identity is separable from information about value: Identity remains relatively fixed over time, while emotion, health status, and desirability can change (**Figure 2b**). Scenes are collections of objects defined by their spatial relationships. An understanding of scenes is essential for us to obtain our goals and respond in appropriate ways to the environment (Malcolm et al. 2016).
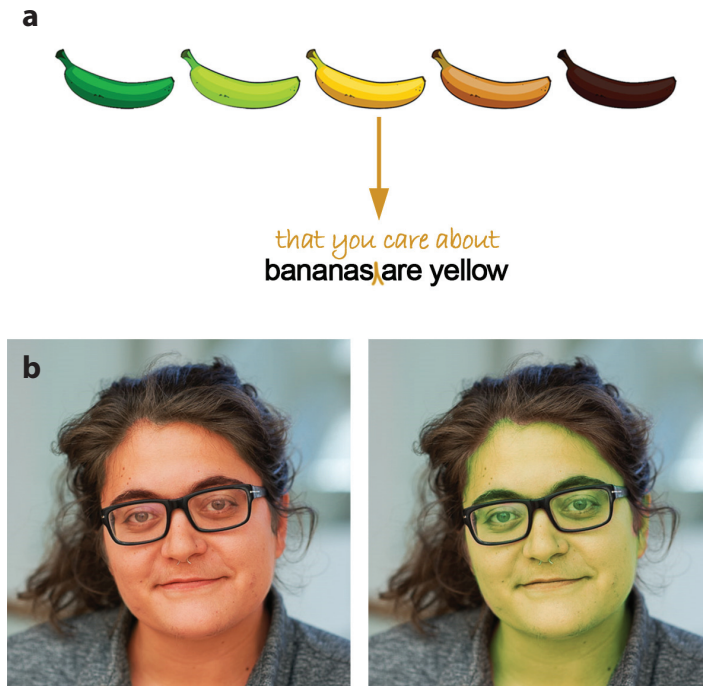
a

that you care about
bananas are yellow

b

One of the biggest challenges in understanding how IT works has been to develop a theory for how the visual system transforms retinal images into seemingly abstract representations (semantic ideas) such as object categories like cat, vehicle, animal, and tool (Huth et al. 2016). The paradox is that vision depends on retinal signals, yet object knowledge appears to be divorced from these signals. For example, we can recognize an object from many vantage points, an ability called view-invariant recognition. Object vision is even more remarkable because we can group together many different exemplars into a common category even though each exemplar causes an obviously different pattern of retinal activity (hammer, saw, and pliers are all tools, but the retinal image caused by each is unique). How does the visual system use retinal images to create representations that are not reflexively yoked to retinal images? Do categorizations have a fundamental perceptual basis that reflects the structure of the world? Or do categorizations reflect a cognitive (semantic) operation that is independent from visual shape parameters?

My goal here is to develop a heuristic for the organization and operation of IT to address these questions, incorporating principles of neural circuit function, development, evolution, and behavior. Brain regions that enable behavior must, presumably, reflect not simply visual image statistics but what we want to do with visual information—our behavioral goals. One hope is that an understanding of the neural mechanisms and organization of IT, including the inputs and outputs, will help us better understand what objects are, what object vision is for, and how the brain systems that support object vision evolved (Mahon & Caramazza 2011, Malcolm et al. 2016). For example, the rich connectivity between IT and the basal ganglia provides clues for how objects acquire value (Griggs et al. 2017).

3. PRINCIPLES OF VISUAL-INFORMATION PROCESSING: VISUAL AREAS AND RETINOTOPY

The classical, simplified idea of information processing by the ventral visual pathway depends on a feedforward hierarchical network that connects a set of visual areas, culminating in high-level visual representations in IT (DiCarlo et al. 2012, Serre 2016). The original notion of a visual area arose following the observation that the cerebral cortex contains multiple, spatially distinct retinotopic representations of the visual field. The use of a new kind of gun during the Russo-Japanese War in the early 1900s led to many soldiers who suffered gunshot wounds but did not die. The powerful bullets damaged the brain with piercing lesions that caused scotomas in the visual field whose locations were correlated with the bullet's trajectory through the cortex (Inouye 1909). By correlating the location of brain damage with the location within the visual field of vision loss, Inouye uncovered the systematic spatial arrangement of V1 neurons within the cortical sheet, in which the fovea is represented on the posterior pole of the occipital lobes and the visual-field periphery is represented by cortex radiating medially. Our understanding of the retinotopic map of V1 was refined with microelectrode recordings and anatomical studies in monkeys in the 1960s and 1970s (Hubel & Wiesel 1977, Van Essen & Zeki 1978) and fMRI in the 1990s (Wandell & Winawer 2011). With this refinement came the discovery of additional retinotopically organized regions—the idea of a second and third visual area (V2, V3) was born. The importance of retinotopy in determining cortical organization is reflected in how the cortex folds itself (Rajimehr & Tootell 2009). Folding patterns correlate with retinotopic maps, apparently to minimize the tension on the cortical sheet connecting corresponding portions of the retinotopic map across areas (Parker 1896, Van Essen 1997).

Why would we need multiple, separate cortical representations of the visual field? Consideration of this question spawned the idea of a visual area: a region of cortex containing a single, more-or-less complete representation of the visual field devoted to performing some fundamental computation. The multiple visual areas came to be understood as connected in series, although the actual connectivity of the neurons between the areas is much more complicated, involving reciprocal feedback connections and feedforward pathways that skip areas: Major pathways include V1-V3-MT (skipping V2); V1-V2-V4 (skipping V3); and V1-V4 (skipping V2 and V3) (see **Figure 3**). Direct connections between V1 and areas beyond V2 undermine the idea that the visual pathway is strictly hierarchical. The notion that each visual area has a single unified function has also been undermined by observations of subcompartments within areas associated with distinct functions from the rest of the area, such as the blobs in V1 (Horton & Hubel 1981), the stripes in V2 (Hubel & Livingstone 1985), disparity columns in V3/V3A (Hubel et al. 2013), and globs in V4 (Conway et al. 2007).

How many retinotopic visual areas are there? The question was first addressed by mapping out the pattern of degenerating axons following resection of the corpus callosum (Van Essen & Zeki 1978). A band of degenerated tissue demarcated the vertical meridian representation that separates V1 from V2. Another band of degenerating tissue provided evidence for two more visual areas, V3 and V4 (the boundary between V2 and V3 is a horizontal meridian). Anterior to V4, degenerating tissue was found as islands that did not form a clear band but were nonetheless unmistakable evidence of a connection between the hemispheres, demarcating a fragmented representation between V4 and IT. These results are confirmed by fMRI in alert monkeys, which reveal a clear anterior V4 boundary (Lafer-Sousa et al. 2012, Vanduffel et al. 2014). There is also a weak meridian representation within posterior IT (Lafer-Sousa & Conway 2013). But disagreements persist about the retinotopy of V4 and the extent to which comparable regions can be found in humans and monkeys (Winawer et al. 2010). Regardless, there is consensus that in both monkeys and humans,

the organizational rules of IT do not depend on strict retinotopic maps like those found in V1, V2, and V3, at least not at a spatial scale that can be recovered with current imaging techniques (Nishimoto et al. 2017, Vernon et al. 2016).

Is IT one area or many? Does the concept of area even make sense when considering IT? It would seem unlikely that IT is one area, simply because the scale of the region is so much larger than any other visual area. Some schematics of IT carve the region into two chunks (Boussaoud et al. 1991, Tanaka 1996), while others divide it into three (DiCarlo et al. 2012) or more components (Rolls 2000). The basis for the number of partitions and their boundaries is not entirely objective but depends on subjective assessments of the variation in cytoarchitecture, together with the patterns recovered in tract-tracing experiments. In the two-chunk scheme, IT comprises a posterior region [sometimes called TEO (temporal occipital cortex)] and an anterior region [TE (temporal cortex)]; in the three-chunk scheme, IT comprises a posterior region (PIT), a central region (CIT), and an anterior region (AIT). A recent synthesis of anatomical data advances the idea that IT comprises not two or three regions but four main nodes (Kravitz et al. 2013); to anticipate the discussion below, this anatomical framework is most consistent with the functional organization discovered using fMRI. But in any of these schemes, how does the anatomy relate to functional organization? A comprehensive theory of the organization of IT needs to be mechanistic, not just descriptive. We seek to know the rules that determine the organizational structure of IT and how these rules relate to principles of neural systems found elsewhere. The theory must be consistent with developmental mechanisms, as well as evolutionary processes that brought about the expansion of the cerebral cortex (Allman 1999, Krubitzer 2009). Importantly, the organizational structure of IT needs to reflect what IT is for.

4. CURRENT MODELS OF INFERIOR TEMPORAL CORTEX ORGANIZATION

The abrupt loss of clear retinotopic maps from V4 to PIT reflects a computational operation: the transformation of retinal representations into object representations that are not yoked to retinal projections. The location of the transition implicates V4/PIT as a key player in the transformation and may explain the involvement of these regions in visual attention: they are the gateway that regulates how retinal information guides behavior.

Two competing ideas for how IT is organized have emerged that draw upon an extensive fMRI literature documenting the pattern of responses to faces, visual words, places, bodies, shapes, motion, color, movies, and disparity (Bartels & Zeki 2000, Grill-Spector & Weiner 2014, Huth et al. 2012, Kanwisher 2010). Advances in recording technology promise an even clearer description of the different scales of IT organization (Sato et al. 2013).

The first idea is that IT comprises specialized modules for certain highly relevant categories such as faces and bodies (Kanwisher 2010, Pitcher et al. 2009). According to this idea, specialized islands of tissue are distributed within general-purpose tissue. This idea has been called the Swiss Army knife model because it considers the brain a collection of instruments, some of which are designed to tackle a very specific task (face perception; the cork screw) and others of which have a more general function (object categorization; the knife) (Kanwisher 2010). The model has been enormously useful in guiding research. But it is descriptive. There is no a priori rationale for why some tasks get a special instrument; and it is not clear what determines where in the cortex the specialized modules are located. The model suggests that knowledge gleaned from studying any specialized instrument need not generalize to an understanding of how other aspects of IT operate. The second model describes the representations within IT as distributed (Haxby et al. 2011, Huth et al. 2012, Kriegeskorte et al. 2008). The distributed model acknowledges that the

responses to different object categories are not uniform across the cortex and argues that the brain exploits the whole distributed pattern, including information from both the increase and decrease in activity across the cortex, to encode objects. This model is also descriptive and provides no mechanistic account of what governs the response patterns.

It seems likely that IT should relate to the rules of cortical organization of V1 for two reasons. First, evolution is a gradual process that builds upon what exists and V1 is likely homologous across mammals. Second, the rules of V1 shape the organization of other extrastriate areas, yielding an organization of V1/V2/V3 that is consistent with evolutionary processes [see opening quote (Allman & Kaas 1971)]. An alternative model advances the idea that IT consists of a single area governed by an eccentricity template, in which the foveal representation has taken on the role of processing faces, while the peripheral representation has taken on the role of processing places (Levy et al. 2001, Malach et al. 2002). This scheme is attractive because it relates in an obvious way to the retinotopic organization of early visual cortex. Moreover, the association of foveal representations (the center of gaze) with face perception links cortical organization with behavior (we tend to look at faces). The idea that the functional organization of extrastriate visual cortex is determined by retinotopic principles of V1 as they relate to behavior has been extended to consider asymmetries in upper and lower visual fields (Groen et al. 2017, Silson et al. 2015). Konkle & Oliva (2012) have advanced a related model, in which real-world size, rather than eccentricity, governs the organizational plan of IT. Others (Kriegeskorte et al. 2008, Naselaris et al. 2012, Sha et al. 2015) have uncovered a spatial representation of animate versus inanimate objects, which also maps coarsely onto the eccentricity template. All these models treat IT as one large area, which neglects studies of anatomical connectivity and neurophysiology that suggest subdivisions within IT along the posterior-anterior axis.

5. A NEW MODEL FOR THE ORGANIZATION OF INFERIOR TEMPORAL CORTEX

We revisited the question of the organization of IT by mapping the responses to a large battery of images and visual stimuli using fMRI in alert monkeys (Lafer-Sousa & Conway 2013, Verhoef et al. 2015) and humans (Lafer-Sousa et al. 2016). There were two key features of the experiments. First, responses to all stimuli were obtained in the same individual subjects, so we could recover the location of the functional domains relative to each other. Thus the experiment controlled for individual differences in the absolute location of functional domains relative to anatomical landmarks (Glasser et al. 2016). Second, we determined responses to many different object cues, including classes of stimuli defined by independent low-level features, such as colored gratings and achromatic faces. As discussed above, prior work has shown that IT contains domains that are specialized for faces, bodies, visual words, and places. The extent to which these domains are distinct from each other has been hard to pin down because in all cases, the stimuli rely on the same low-level visual features: luminance contrast, lines, and edges. Body-responsive regions reside next to face patches (Grill-Spector & Weiner 2014, Lafer-Sousa & Conway 2013), but there is the nagging possibility that this relationship derives from similarity in the low-level features of the stimuli used and not the underlying functional organization. Our experiments uncovered a series of color-biased regions along the temporal lobe and showed that these regions avoid disparity-sensitive regions and are located between face patches and place-biased regions (Lafer-Sousa & Conway 2013, Verhoef et al. 2015). The location of the series of functionally biased domains is predicted by the nodes in the anatomical connectivity map (**Figure 3**). Anatomical connectivity studies have also uncovered several parallel networks running along the posterior-anterior axis of IT (Kravitz et al. 2013). This anatomical form presumably accounts for the pattern of functional
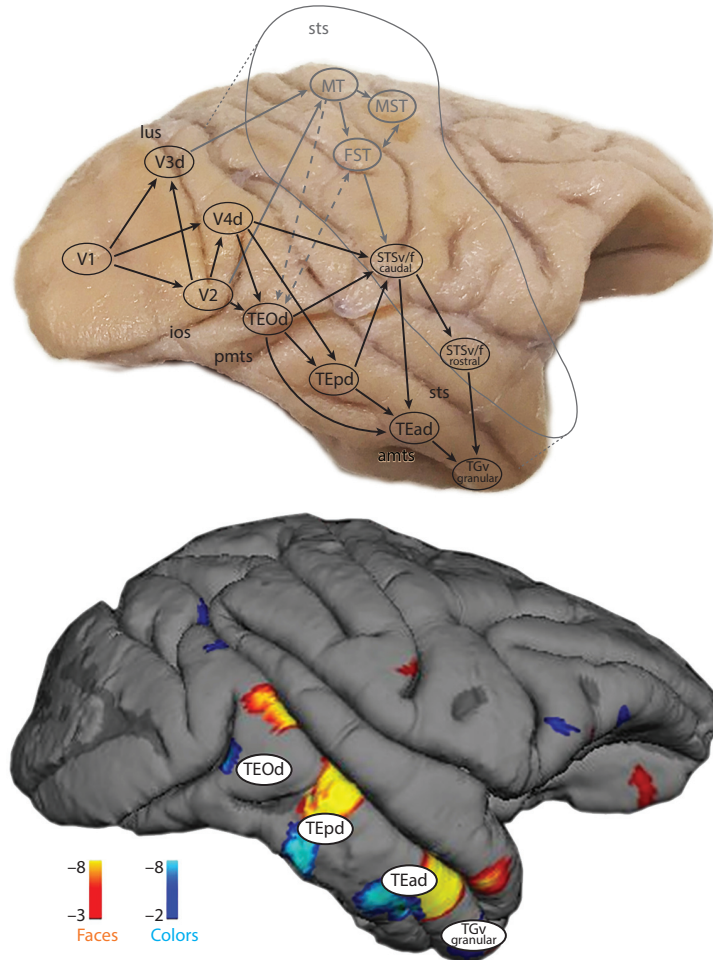
Figure 3

The location of the functional activation patterns in inferior temporal cortex are predicted
by nodes in the anatomical connectivity map. Top panel shows the serial stages of the major anatomical
projections through the ventral visual pathway. The anatomy also uncovers several parallel pathways along
the posterior-to-anterior axis. Regions of the dorsal pathway are shown in gray text. Bottom panel shows the
functional activation patterns to faces (*orange*) and colors (*blue*) in an example macaque monkey. Top panel
adapted from Kravitz et al. (2013) and bottom panel from Lafer-Sousa & Conway (2013). Abbreviations:
amts, anterior middle temporal sulcus; d, dorsal; f, fundus; FST, floor of superior temporal sulcus; ios, inferior
occipital sulcus; lus, lunate sulcus; MST, medial superior temporal area; MT, middle temporal area; pmts,
posterior middle temporal sulcus; sts, superior temporal sulcus; STSf, fundus of the superior temporal sulcus;
STSv, lower (ventral) bank of the superior temporal sulcus; TEad, dorsal subregion of anterior TE; TEOd,
area TEO, dorsal part; TEpd, dorsal subregion of posterior TE; TGv, ventral temporal pole; v, ventral.

responses. And the systematic organization suggests that all of IT is subject to one organizing
principle, in which the processing operations within each parallel stream are similar. A radical
prediction from this organization is that the most studied system, the face-patch system, is not
unique, but rather one manifestation of a common set of rules that determines how all of IT
operates. The results in **Figure 3** are from macaque monkeys. If the organization is fundamental,
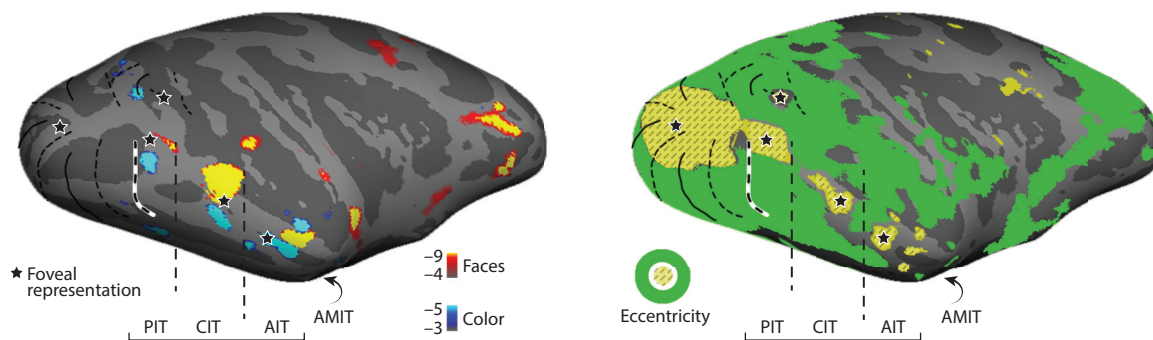
Figure 4

IT contains a series of eccentricity representations that correspond to the location of the sets of color-biased and face-biased domains. Cyan-blue shows regions with higher responses to drifting equiluminant colored gratings compared to drifting achromatic gratings; orange-red shows regions with higher response to black and white pictures of faces compared to pictures of bodies (*left panel*). Yellow hashing shows regions with higher responses to checkers restricted to a 3° diameter disc centered at the fovea, and green shows responses to checkers in the periphery (the screen was 40° across; *right panel*). Figure created using data from Lafer-Sousa & Conway (2013) and Verhoef et al. (2015). Abbreviations: AIT, anterior inferior temporal cortex; AMIT, anteromedial inferior temporal cortex; CIT, central inferior temporal cortex; IT, inferior temporal cortex; PIT, posterior inferior temporal cortex.

we predicted the same organizational scheme in humans. Indeed, we found that the ventral visual pathway in humans also consists of a series of color-biased regions sandwiched between face and place regions (Lafer-Sousa et al. 2016), supporting the idea that this tripartite organization emerged at least as early as our common primate ancestor about 25 million years ago.

To determine the rules that govern the organizational structure, we mapped fMRI responses to checkerboards shown at different visual-field eccentricities (Lafer-Sousa & Conway 2013). Even though IT does not have a clear retinotopic organization, surprisingly it does have a set of coarse eccentricity representations. Several separate eccentricity foci along the temporal lobe carve up IT into several more-or-less equal-sized chunks, each about the same size as a classical retinotopic visual area (**Figure 4**): Three of the IT regions (PIT, CIT, AIT) have their own clear foveal representation; the fourth region, which we refer to as anteromedial IT (AMIT), following the PIT/CIT/AIT naming convention, is located at the ventral-medial pole of the temporal lobe. Face patches align with foveal representations; color-biased regions are in mid-periphery representations; and place patches reside in peripheral representations. The systematic relationship of the functional regions to each other, and their relationship to the eccentricity biases, has been confirmed (Arcaro et al. 2017, Rajimehr et al. 2014) and suggests that each chunk of IT comprises a complete set of functionally biased regions. The results provide the contours for a new model for the organization of IT: parallel, multistage processing governed by repeated eccentricity templates (**Figure 5**). The physical proximity of color-biased regions and face patches makes some surprising predictions, and it explains a paradox in the human-lesion literature. It has long been known that acquired achromatopsia is accompanied by prosopagnosia (Bouvier & Engel 2006). This observation has been perplexing because face perception and color perception are largely independent perceptual/cognitive processes. The organizational scheme proposed here provides a mechanism that resolves the paradox. Moreover, the scheme predicts that acquired prosopagnosia should be accompanied by acquired achromatopsia in brain-damaged patients. To my knowledge, this prediction has yet to be tested.
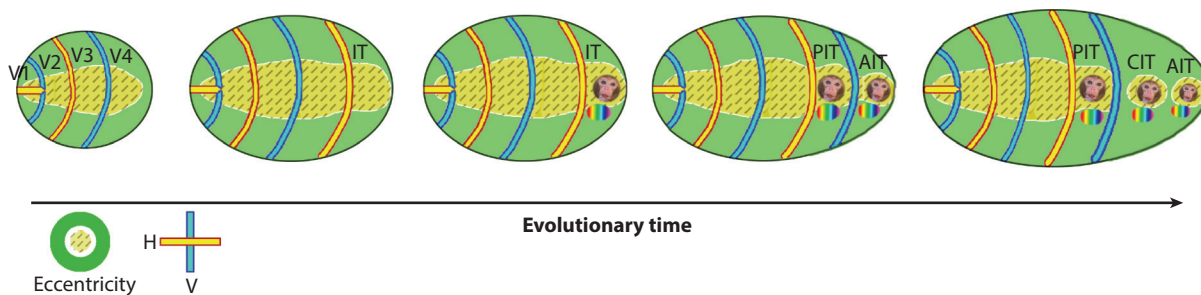
Figure 5

Model for the evolutionary expansion of IT through duplication of retinotopic visual areas. Abbreviations: AIT, anterior inferior temporal cortex; CIT, central inferior temporal cortex; H, horizontal; IT, inferior temporal cortex; PIT, posterior inferior temporal cortex; V, vertical; V1, primary visual cortex; V2, visual area 2.

The model is attractive because it not only accounts for empirical evidence but also acknowledges the importance of retinotopy as a fundamental organizing property of visual areas. The results are consistent with the idea that IT comprises four areas arranged along the posterior-anterior axis and suggest that these areas evolved by the same mechanism that has been suggested for the evolution of V1, V2, and V3: reduplication of a proto-area that itself is governed by eccentricity [see the quote at the beginning of the article (Allman & Kaas 1971)]. The model is parsimonious not only with principles of cortical expansion found in other domains (such as the motor system) but also with developmental mechanisms known to wire up cortical areas. Waves of spontaneous activity across the retina are essential for establishing the organization of V1 (Firth et al. 2005, Warland et al. 2006, Wong 1999). It is plausible that spontaneous waves of activity propagating from V1 to V2 and V3, yoked by a common eccentricity template, provide the instructions to establish the series of areas in IT (Baldassano et al. 2012, Liu et al. 2015, Striem-Amit et al. 2015) and more broadly across the cortex (Griffis et al. 2017).

The idea that eccentricity serves as a template that governs the functional maps has since found support in work showing that retinotopic biases in IT are present very early in postnatal development (Arcaro & Livingstone 2017). If the structural relationship between cortical folding patterns and retinotopic maps described in Section 3 extends to IT, one might expect a relationship between anatomical landmarks and the functional domains in IT. Indeed, IT regions relate systematically to anatomical landmarks as well as each other (Grill-Spector & Weiner 2014, Osher et al. 2016). Moreover, the topography of patterns of activity in IT are stable across time and behavioral training (Op de Beeck et al. 2008, Srihasam et al. 2014). The importance of eccentricity as the template of IT organization makes one other surprising prediction: that object coding should depend on spatial information across the visual scene, which appears to be true (Connor & Knierim 2017, Kim & Biederman 2011, Kravitz et al. 2010, Monosov et al. 2010, Sereno et al. 2014). Finally, the model predicts that visual behaviors depending on the upper versus lower visual field should be encoded by parts of IT that inherit the corresponding representation (Malcolm et al. 2016, Vaziri & Connor 2016).

The basic architecture of IT is comparable in monkeys and humans (Lafer-Sousa et al. 2016, Schalk et al. 2017), but establishing firm homologies will require further work. For example, humans have sets of functionally defined visual regions on the lateral cortical surface, including a face region on the lateral surface [posterior portion of the superior temporal sulcus (pSTS)]. Do macaque monkeys have corresponding regions? We cannot expect the functional regions to be located with the same relationship to anatomical landmarks, because there are substantial

differences in gross cortical anatomy between primate species; moreover, large evolutionary changes in other parts of the cortex, such as in auditory and language regions, have presumably perturbed the functional layout of the cortex in humans compared to macaque monkeys. One possibility is that pSTS in humans corresponds to the middle dorsal (MD) and middle fundus (MF) face patches in monkeys, identified using dynamic face stimuli (Fisher & Freiwald 2015). According to this scheme, the region separating the ventral and dorsal sets of face regions has expanded disproportionately in humans. Quantitative comparisons of functional activation patterns across the species, using a large range of stimuli and paradigms, will be needed to further determine which computations in IT are conserved across species and which are species specific (Lafer-Sousa et al. 2016, Nasr et al. 2011, Tsao et al. 2008, Yovel & Freiwald 2013). Behavioral experiments comparing the perceptual and cognitive abilities of monkeys and humans will also be needed to test the hypothesis that functional architecture reflects adaptations of the cortex to species-specific behavior.

6. THREE IMPLICATIONS OF THE MODEL

6.1. Canonical Operations Along the Parallel Streams of Inferior Temporal Cortex

The systematic relationship of the functional domains, together with anatomical evidence, suggests that the operations of the face-patch system are not unique but rather one manifestation of a set of canonical operations implemented in parallel streams through IT. The model raises the possibility that knowledge obtained by studying operations implemented in one stream through IT generalize to other streams. This idea could be powerful as we seek to understand what computations are carried out in the various functional domains of IT. How many parallel networks are there through IT? The anatomy and functional data suggest perhaps half a dozen parallel circuits in IT (Kravitz et al. 2013, Verhoef et al. 2015). What are they for? And how are the canonical operations manifest in each of them?

The operations within the face-patch system have received by far the most attention of any part of IT (Freiwald et al. 2016, Leopold et al. 2006). The work suggests that face cells are linked in a causal way to face perception (Afraz et al. 2006, Moeller et al. 2017, Sadagopan et al. 2017), and, moreover, that the face patches comprise a hierarchy in which posterior patches encode primitive features, such as eyes (Issa & DiCarlo 2012), central regions calculate context relationships that underlie the configuration of a face (Freiwald & Tsao 2010), and anterior regions represent memory for specific individuals (Landi & Freiwald 2017). This simple heuristic is mirrored in the body-patch network (Kumar et al. 2018) and in the overall increase in category selectivity and view invariance from V4 through IT (Hong et al. 2016, Rust & DiCarlo 2010). The heuristic represents a substantial departure from the conventional notion of a functional area as defined by Zeki, in which all aspects of a given operation are housed in one localized patch of cortex. In the same way that face processing depends on a network extending throughout IT, there does not appear to be a single so-called color area but rather a network of regions along the ventral visual pathway that together give rise to the various aspects of color perception. Color was the tool used to identify these regions, but we do not know what operations the cells in these regions perform. Hence the names for these regions are intentionally descriptive (color-biased regions). It seems unlikely that the regions are contributing simply to a representation of hue, because a complete representation of color is already evident within V4/PIT (Bohon et al. 2016), which is earlier than IT in the putative visual-processing hierarchy. Rather, the existence of color-biased regions in IT, and their association with the face patches, suggests that the IT color-biased regions are using color toward some high-level object representation.

Color is rarely mentioned in computational work on object recognition (Gauthier & Tarr 2016, Serre 2016); stimuli in object-recognition experiments are often achromatic; and influential object-recognition models ignore color (Riesenhuber & Poggio 1999). The lack of color in the object-vision literature reflects the widespread belief that color is not important for most object-recognition tasks—although careful quantitative studies have shown that color does afford a modest benefit for object memory and recognition (Gegenfurtner & Rieger 2000). The discovery of an extensive network of color-biased regions in IT, the part of the brain implicated in high-level object vision, compels a reexamination of the role of color (Beauchamp et al. 1999, Lafer-Sousa & Conway 2013). Perhaps the main contribution of color is not to facilitate object recognition but rather to provide information independent of shape (or identity) about the relevance of objects to behavior (**Figure 2**). Our working hypothesis is that color is a trainable system that allows rapid detection of objects of likely behavioral relevance, which promotes the efficiency of vision by directing attention to those aspects of the visual world that are most likely to contain useful information (Conway 2016). In support of this hypothesis, we find that the color statistics of objects are not a random, uniform sampling of color space. Instead, object colors are systematically biased toward warm colors, which correlates with biases in neural color tuning found in IT (Rosenthal et al. 2018). The importance of warm coloring is reflected in the informativity of color-naming systems: Warm colors are communicated more efficiently than cool colors across all languages (Gibson et al. 2017). Thus color naming, and the color-vision system more generally, reflects the usefulness of objects. The idea that color contributes in an important way to object knowledge should have implications for artificial vision systems that aim to capture how human vision works. And while color was the tool that enabled the discovery of this system, it is unlikely to be the sole feature exploited by the system. Other surface and material properties, such as texture, likely contribute to the operations within it (Nishio et al. 2014).

Using the face system as a guide, we can speculate on the operations within the various stages of the color-biased network. Object-color memories constitute high-level color behavior reflecting specific learned associations. These may be analogous to the representation of behaviorally relevant faces; such representations are found in anterior IT, so we might expect also to find representations for object-color memories in anterior IT. Consistent with this idea, anterior IT is richly connected with subcortical structures, including the striatum and amygdala, and with medial temporal cortex (Kravitz et al. 2013), including the hippocampus—brain structures that are likely called upon to mediate the memory and rewarding aspects of color. Meanwhile, computations such as color grouping share some homology with the contextual calculations performed by the middle face patches. Finally, a representation of specific hues is arguably comparable to a representation of specific face features: Both are primitive building blocks for more complex representations. Consistent with the multistage parallel framework, eye features and specific hues appear to be processed in parallel, at the same early stage along IT, in PIT (Bohon et al. 2016, Conway & Tsao 2009, Namima et al. 2014). But a full test of these hypotheses will require measuring cell activity across functionally defined domains in the context of relevant tasks.

The model outlines a series of stages but does not call for a strict hierarchy among the regions. This is important because operations carried out by each of the stages along IT can function somewhat independently through redundant inputs from across the network of stages. Anatomical tracing shows that the stages are richly connected through feedback and feedforward connections that link nonadjacent stages (**Figure 3**, *top*) (Kravitz et al. 2013). Behavioral evidence from the literature on human patients provides support for this aspect of the model. Although color perception would seem to be a precursor for color imagery, patients suffering brain damage can lose color perception and color naming but retain color imagery (Bartolomeo et al. 1997).

Other patients can retain the ability to see and name colors but lose the ability to associate the appropriate color with objects (Luzzatti & Davidoff 1994, Miceli et al. 2001), and other patients can retain color perception and object-shape perception but lose the ability to link the color with the name of a common object of that color (Tanaka et al. 2001). Measuring activity across multiple domains simultaneously will be necessary to determine how the whole network of regions operates in concert to carry out behavior. Knowledge acquired through this approach will likely trigger a revision of the model, in much the same way as simultaneous single-cell recordings with arrays have revolutionized our understanding of neural-circuit function and population coding.

6.2. How Shape Features Become Semantic Categories

IT responses are often interpreted through a feedforward hierarchical model, an extension of the Hubel-Wiesel approach. An alternative is that IT responses represent semantic concepts decoupled from visual shape signals (Henriksson et al. 2015, Huth et al. 2012). The thrust of this idea is that object categories are bound by behavioral relevance (semantic meaning), not by common visual features (Carlson et al. 2014, Mahon & Caramazza 2011). The presumption is that top-down signals reflecting accumulated experience (abstract knowledge about the world) are instrumental in organizing IT. The idea has been probed by analyzing responses to stimuli categorized as animate versus inanimate (Liu et al. 2013, Naselaris et al. 2012). The importance of the animate/inanimate distinction finds support in studies of brain-damaged patients (Caramazza & Shelton 1998). Representational similarity analyses comparing human fMRI results and macaque neurophysiology data (Kriegeskorte et al. 2008) suggest that representing animacy is a main organizing principle of IT. The work suggests not only that monkeys and humans share the same functional operations within IT but also that these operations cannot easily be explained by object features.

But the extent to which IT is organized by top-down (semantic categories) or bottom-up (hierarchical feedforward) rules is not settled. Other neurophysiological data argue that apparent semantic categorical responses in IT, including animacy, can be explained by similarity among category members in terms of visual features such as object area, luminance, curvature, aspect ratio, and the extent to which the object has a star shape (elements radiating from a common point) (Baldassi et al. 2013, DiCarlo et al. 2012, Op de Beeck et al. 2001). Consistent with these results, color features not only distinguish objects from backgrounds (Gibson et al. 2017) but also provide information that separates animate and inanimate objects (Rosenthal et al. 2018). The notion that shape or color features, which are encoded by stages in visual processing prior to IT, can explain categorical responses supports the feedforward hierarchical model.

Deep convolutional neural networks (CNNs) have attracted attention as another way to address the extent to which neural representations encode visual features or semantic (categorical) meaning (Cadieu et al. 2014, Khaligh-Razavi & Kriegeskorte 2014, Yamins & DiCarlo 2016b). These networks are trained on massive numbers of images and can perform categorization tasks very well (He et al. 2015). Models trained using only visual features perform less well in categorizing stimuli, as compared with the CNNs; models trained using categorical labels perform comparably to the CNNs (Jozwik et al. 2017). These results support the idea that at some stage in the nervous system, representations of object categories are invoked to guide object-similarity judgments. The results have been taken to support a model in which IT organization is determined by semantic (top-down) signals. But I think the results support the alternative hypothesis: The performance of the CNNs is determined only by the images fed to the network and image labels made by humans. If the CNN recovers category (or semantic or any other kind of) information, it must be because that information is embedded somehow in the image statistics. The visual features or similarity

judgments used in the models must not be an accurate reflection of the visual features available to the CNN (or IT). Figuring out how the brain turns visual data into semantic concepts remains a hard problem.

Taken together, the results across a diverse range of studies suggest that semantic categories cannot be entirely decoupled from the visual features that make up objects. Semantic categories appear to be bound up with visual shape features (Gaffan & Heywood 1993, Lescroart et al. 2015, Watson et al. 2017). The idea that categorization depends on visual features recalls behavioral work showing that basic-level categories are defined by distinctive shape features (birds, defined by wings and feathers)—and, as one would predict if visual features define categories, category members that do not share the defining feature are less reliably categorized (Logothetis & Sheinberg 1996). The new model shows how shape features could become semantic categories: The tuning in IT necessarily depends on sensitivity to shape features inherited from early visual areas but reflects the computational objectives of IT to support visual processing toward behaviorally relevant objectives. Until now, the major challenge for experiments on IT has been to define an objective, comprehensive stimulus space. The new model allows us to constrain the stimulus space to a relevant domain (e.g., faces, color, places) and facilitates the integration of information across microelectrode studies by way of functional landmarks.

6.3. Determining Which Complex Visual Features Are Represented in Inferior Temporal Cortex

Some IT neurons appear to respond to only a single exemplar of an object category (e.g., the face of a specific individual person), while other neurons seem to respond to many exemplars of a category. It is tempting to interpret the sparseness of IT responses as evidence in favor of a strict hierarchical feedforward model that culminates with what are referred to as grandmother cells (Rolls 2000). Task performance is surprisingly resilient after local ablations of IT (Matsumoto et al. 2016), so IT cannot be strictly hierarchical. The notion that recognition of an object is handled by a small, dedicated neural population has other limitations—it is hard to imagine how known principles of neural circuit development would produce such a system and enable it to learn new categories and exemplars. Instead, it seems likely that the apparent sparseness reflects sensitivity of the neurons to abstract properties that are both diagnostic of a category and useful for behavior and that each object, and category, is necessarily represented by a distributed population of neurons. This framework extends the representational principle we find in V1: Each single V1 neuron responds sparsely to only a limited range of orientations and only at a given location of the visual field. Yet we would not conclude that V1 comprises grandmother cells for orientation. Instead, orientation is the nature of a representation encoded by V1. It is not clear what the abstract representation is in IT. One idea suggested by the model is that each stage along IT has a computational objective such as detecting members of a category (PIT) or recognizing members within a category (AIT), and these operations involve tuning to the stimulus dimensions that capture the most variance for those problems. Neurons that might appear to be object-identity specific—so-called grandmother cells—are likely part of an ensemble of cells that works together to code identity, rather than each cell serving as a feature detector (Chang & Tsao 2017).

The model described here stipulates that the way in which IT samples and recombines input signals is determined by the behavioral relevance of the information (Kourtzi & Connor 2011, Martin 2016), reflecting pressures exerted during evolution, development, and life experience. A strong test of the importance of experience was provided by an experiment in which monkeys were reared without exposing them to faces, and cortical responses were evaluated throughout development using fMRI (Arcaro et al. 2017). The animals did not develop IT face patches or

look preferentially at faces, suggesting that neural and behavioral selectivity for faces requires experience. These results suggest that experience is the dominant organizing force in IT. But the issue is not settled. In similar experiments, Sugita (2008) reared monkeys without exposing them to faces and showed that the animals nonetheless had a preference for human or monkey faces. After experience with the faces, Sugita's animals developed a species-specific face preference, suggesting that IT has an innate organizational plan that is refined by experience. It is not yet clear how the results of the two experiments can be reconciled. But both sets of results are consistent with the idea that foveal representations in IT inherit a bias for high spatial frequencies and overt attention associated with where you look.

A key aspect of the model described here is the role of active behavior—such as attention, object selection, and behavioral goals—in shaping how IT works. Behavioral context is often not considered in evaluating the core computations of IT, but it should be. The notion of objects implies active behavior, as discussed in Section 2. And ample evidence shows that IT activity is dependent on task performance. The issue has been addressed theoretically (Walther & Koch 2007) and is supported by four lines of evidence. First, shape tuning of IT neurons in monkeys increases in a way that is determined by the task the monkeys perform (Lim et al. 2015, McKee et al. 2014, Sheinberg & Logothetis 2001, Sigala & Logothetis 2002). Second, IT responses predict task performance (Harel et al. 2014, Mruczek & Sheinberg 2007). Third, the response of IT neurons typically follows the responses of neurons in prefrontal cortex (PFC) (Monosov et al. 2010, Tomita et al. 1999); is altered by PFC inactivation (Monosov et al. 2011); and tracks perception of multi-stable images (Sheinberg & Logothetis 1997), uncovering top-down signals to IT from PFC. And finally, IT cells acquire sensitivity to sets of visually dissimilar stimuli if the animal thinks all the stimuli are the same (Li & DiCarlo 2008). It seems inescapable that the categorization behavior of IT cells reflects an interaction of visual statistics and behavioral relevance. This interaction is made explicit by the model: The IT regions in the eccentricity template that are connected to foveal representations in early visual cortex exploit the tuning to high-spatial frequencies to take on a role in high-acuity object vision (including face recognition); intermediate representations take on a role in establishing behavioral valence (using color and surface cues); and peripheral representations encode spatial information such as places. The functional role of the regions in IT is thus dependent both on the predisposition of each part of the eccentricity template to compute certain kinds of information and how those representations are used, as reflected by connectivity patterns to other parts of the brain (Mahon & Caramazza 2011).

The model described here requires both nature (eccentricity templates) and nurture (experience) to establish the functional architecture of IT. An extreme hypothesis, in which experience is the sole determinant, might suppose that cells are clustered in the cortex based only on the emotion, reward, or action elicited by their activation. This theory takes seriously the notion of an object as an active construction that reflects the interaction of agent and environment. According to this theory, the observed clustering of cells in IT would arise because of commonalities in the reward contingencies of objects and not because of similarities in shape features. Thus cells tuned to vegetables and fruits would be grouped together in IT but separately from cells tuned for tools, because vegetables and fruits are edible and tools are not. There is some behavioral evidence that objects are categorized according to ecological experience (Ghazizadeh et al. 2016). Moreover, shape tuning of IT cells can be altered by changing the temporal contingency of visual stimuli during natural search (Li & DiCarlo 2008), which is consistent with the idea that behavioral relevance (and not just visual shape features) influences the responses of IT cells. As of now, the role of ecological experience in determining the overall organization of IT has not been tested with fMRI or neurophysiology. The weight of evidence suggests that similarities in shape features among members of a category play the dominant role in organizing IT (Op de Beeck et al. 2008).

7. HOW DOES THE BRAIN DECODE ACTIVITY IN INFERIOR TEMPORAL CORTEX?

The bulk of this review has addressed how categorical information is encoded in IT. But how are representations in IT decoded to drive behavior? The population dynamics of neurons in prefrontal cortex suggest that they can flexibly select and integrate information from different domains to carry out a common behavioral plan (Mante et al. 2013), making PFC a candidate as a decoder of IT activity. Working out how different PFC regions contribute to this process is an active area of research. The functional organization of prefrontal cortex remains unsettled (O'Reilly 2010, Passingham & Wise 2012, Riley et al. 2017, Romanski 2004), but there is some evidence for stimulus-driven functional organization: PFC contains a set of color-biased regions whose absolute locations within any given hemisphere are somewhat variable but are nonetheless found in a systematic relationship to face patches (Lafer-Sousa & Conway 2013, Romero et al. 2014). The stimulus-driven organization of PFC suggests that different parts of PFC may contribute in distinct ways to decoding the parallel representations of IT. It will be useful to address these questions using an approach that bridges the gap between microelectrode recordings, which provide high cellular resolution of neural mechanisms but suffer severe sampling limitations, and systems-level measurements such as those from fMRI, which provide information on global organization but lack spatial and temporal resolution.

8. SUMMARY

Converging evidence from experiments in macaque monkeys and humans suggests that IT is organized by a sequence of eccentricity templates into about four stages, each of which might be considered a distinct visual area. The functional organization of IT, and the activity of IT cells, reflects an interaction of innate programs that link corresponding parts of the eccentricity templates, stimulus features, and behavioral goals. Working out how corresponding regions along each parallel stream through IT engage with one another to carry out behavioral programs promises to uncover fundamental truths not only about how IT works but about how the brain turns sense data into perceptions, thoughts, and ideas.

LITERATURE CITED

Afraz SR, Kiani R, Esteky H. 2006. Microstimulation of inferotemporal cortex influences face categorization. *Nature* 442:692–95

Allman JM. 1999. *Evolving Brains*. New York: Sci. Am. Libr.

Allman JM, Kaas JH. 1971. A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (*Aotus trivirgatus*). *Brain Res*. 31:85–105

Arcaro MJ, Livingstone MS. 2017. A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife* 6:e26196

Arcaro MJ, Schade PF, Vincent JL, Ponce CR, Livingstone MS. 2017. Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* 20:1404–12

Baldassano C, Iordan MC, Beck DM, Fei-Fei L. 2012. Voxel-level functional connectivity using spatial regularization. *NeuroImage* 63:1099–106

Baldassi C, Alemi-Neissi A, Pagan M, DiCarlo JJ, Zecchina R, Zoccolan D. 2013. Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLOS Comput. Biol.* 9:e1003167

Bartels A, Zeki S. 2000. The architecture of the colour centre in the human visual brain: new results and a review. *Eur. J. Neurosci.* 12:172–93

Bartolomeo P, Bachoud-Levi AC, Denes G. 1997. Preserved imagery for colours in a patient with cerebral achromatopsia. *Cortex* 33:369–78

Beauchamp MS, Haxby JV, Jennings JE, DeYoe EA. 1999. An fMRI version of the Farnsworth–Munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cereb. Cortex* 9:257–63

Bohon KS, Hermann KL, Hansen T, Conway BR. 2016. Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *eNeuro* 3:e0039-16.2016

Boussaoud D, Desimone R, Ungerleider LG. 1991. Visual topography of area TEO in the macaque. *J. Comp. Neurol.* 306:554–75

Bouvier SE, Engel SA. 2006. Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cereb. Cortex* 16:183–91

Brewer AA, Press WA, Logothetis NK, Wandell BA. 2002. Visual areas in macaque cortex measured using functional magnetic resonance imaging. *J. Neurosci.* 22:10416–26

Cadieu CF, Hong H, Yamins DL, Pinto N, Ardila D, et al. 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLOS Comput. Biol.* 10:e1003963

Caramazza A, Shelton JR. 1998. Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *J. Cogn. Neurosci.* 10:1–34

Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR. 2014. The emergence of semantic meaning in the ventral temporal pathway. *J. Cogn. Neurosci.* 26:120–31

Chang L, Tsao DY. 2017. The code for facial identity in the primate brain. *Cell* 169:1013–28.e14

Connor CE, Knierim JJ. 2017. Integration of objects and space in perception and memory. *Nat. Neurosci.* 20:1493–503

Conway BR. 2016. Processing. In *Experience: Culture, Cognition, and the Common Sense*, ed. CA Jones, D Mather, R Uchill, pp. 86–109. Cambridge, MA: MIT Press

Conway BR, Moeller S, Tsao DY. 2007. Specialized color modules in macaque extrastriate cortex. *Neuron* 56:560–73

Conway BR, Tsao DY. 2009. Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *PNAS* 106:18034–39

DiCarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? *Neuron* 73:415–34

Farah M. 2004. *Visual Agnosia*. Cambridge, MA: MIT Press. 2nd ed.

Firth SI, Wang CT, Feller MB. 2005. Retinal waves: mechanisms and function in visual system development. *Cell Calcium* 37:425–32

Fisher C, Freiwald WA. 2015. Contrasting specializations for facial motion within the macaque face-processing system. *Curr. Biol.* 25:261–66

Freiwald W, Duchaine B, Yovel G. 2016. Face processing systems: from neurons to real-world social perception. *Annu. Rev. Neurosci.* 39:325–46

Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330:845–51

Gaffan D, Heywood CA. 1993. A spurious category-specific visual agnosia for living things in normal human and nonhuman primates. *J. Cogn. Neurosci.* 5:118–28

Gagin G, Qu J, Hu Y, Lafer-Sousa R, Conway BR. 2014. Color-detection thresholds in monkeys and humans. *J. Vis.* 14(8):12

Gauthier I, Tarr MJ. 2016. Visual object recognition: Do we (finally) know more now than we did? *Annu. Rev. Vis. Sci.* 2:377–96

Gegenfurtner KR, Rieger J. 2000. Sensory and cognitive contributions of color to the recognition of natural scenes. *Curr. Biol.* 10:805–8

Geisler WS. 2008. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59:167–92

Ghazizadeh A, Griggs W, Hikosaka O. 2016. Ecological origins of object salience: reward, uncertainty, aversiveness, and novelty. *Front. Neurosci.* 10:378

Gibson E, Futrell R, Jara-Ettinger J, Mahowald K, Bergen L, et al. 2017. Color naming across languages reflects color use. *PNAS* 114:10785–90

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, et al. 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536:171–78

Griffis JC, Elkhetali AS, Burge WK, Chen RH, Bowman AD, et al. 2017. Retinotopic patterns of functional connectivity between V1 and large-scale brain networks during resting fixation. *NeuroImage* 146:1071–83

Griggs WS, Kim HF, Ghazizadeh A, Gabriela Costello M, Wall KM, Hikosaka O. 2017. Flexible and stable value coding areas in caudate head and tail receive anatomically distinct cortical and subcortical inputs. *Front. Neuroanat.* 11:106

Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15:536–48

Groen IIA, Silson EH, Baker CI. 2017. Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philos. Trans. R. Soc. B* 372:20160102

Gross CG. 2008. Single neuron studies of inferior temporal cortex. *Neuropsychologia* 46:841–52

Harel A, Kravitz DJ, Baker CI. 2014. Task context impacts visual object processing differentially across the cortex. *PNAS* 111:E962–71

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, et al. 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72:404–16

He KM, Zhang XY, Ren SQ, Sun J. 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In *2015 IEEE International Conference on Computer Vision*, pp. 1026–34. Los Alamitos, CA: IEEE Comp. Soc.

Henriksson L, Khaligh-Razavi SM, Kay K, Kriegeskorte N. 2015. Visual representations are dominated by intrinsic fluctuations correlated between areas. *NeuroImage* 114:275–86

Hong H, Yamins DL, Majaj NJ, DiCarlo JJ. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* 19:613–22

Horton JC, Hubel DH. 1981. Regular patchy distribution of cytochrome oxidase staining in primary visual cortex of macaque monkey. *Nature* 292:762–64

Hubel DH, Livingstone MS. 1985. Complex-unoriented cells in a subregion of primate area 18. *Nature* 315:325–27

Hubel DH, Wiesel TN. 1977. Ferrier lecture—functional architecture of macaque monkey visual cortex. *Proc. R. Soc. B* 198:1–59

Hubel DH, Wiesel TN, Yeagle EM, Lafer-Sousa R, Conway BR. 2013. Binocular stereoscopy in visual areas V-2, V-3, and V-3A of the macaque monkey. *Cereb. Cortex* 25:959–71

Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532:453–58

Huth AG, Nishimoto S, Vu AT, Gallant JL. 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76:1210–24

Inouye T. 1909. *Die Sehstörungen bei Schussverletzungen der kortikalen Sehsphäre nach Beobachtungen an Versundeten der letzten Japanische Kriege.* Leipzig, Ger.: W. Engelmann

Issa EB, DiCarlo JJ. 2012. Precedence of the eye region in neural processing of faces. *J. Neurosci.* 32:16666–82

Jozwik KM, Kriegeskorte N, Storrs KR, Mur M. 2017. Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Front. Psychol.* 8:1726

Kanwisher N. 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. *PNAS* 107:11163–70

Khaligh-Razavi SM, Kriegeskorte N. 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Comput. Biol.* 10:e1003915

Kim JG, Biederman I. 2011. Where do objects become scenes? *Cereb. Cortex* 21:1738–46

Konkle T, Oliva A. 2012. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74:1114–24

Kourtzi Z, Connor CE. 2011. Neural representations for object perception: structure, category, and adaptive coding. *Annu. Rev. Neurosci.* 34:45–67

Kravitz DJ, Kriegeskorte N, Baker CI. 2010. High-level visual object representations are constrained by position. *Cereb. Cortex* 20:2916–25

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17:26–49

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, et al. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–41

Krubitzer L. 2009. In search of a unifying theory of complex brain evolution. *Ann. N. Y. Acad. Sci.* 1156:44–67

Kumar S, Popivanov ID, Vogels R. 2018. Transformation of visual representations across ventral stream body–selective patches. *Cereb. Cortex.* In press

Lafer-Sousa R, Conway BR. 2013. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat. Neurosci.* 16:1870–78

Lafer-Sousa R, Conway BR, Kanwisher NG. 2016. Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *J. Neurosci.* 36:1682–97

Lafer-Sousa R, Liu YO, Lafer-Sousa L, Wiest MC, Conway BR. 2012. Color tuning in alert macaque V1 assessed with fMRI and single-unit recording shows a bias toward daylight colors. *J. Opt. Soc. Am. A* 29:657–70

Landi SM, Freiwald WA. 2017. Two areas for familiar face recognition in the primate brain. *Science* 357:591–95

Lehky SR, Tanaka K. 2016. Neural representation for object recognition in inferotemporal cortex. *Current Opin. Neurobiol.* 37:23–35

Leopold DA, Bondar IV, Giese MA. 2006. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442:572–75

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

Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat. Neurosci.* 4:533–39

Li N, DiCarlo JJ. 2008. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* 321:1502–7

Lim S, McKee JL, Woloszyn L, Amit Y, Freedman DJ, et al. 2015. Inferring learning rules from distributions of firing rates in cortical neurons. *Nat. Neurosci.* 18:1804–10

Liu N, Kriegeskorte N, Mur M, Hadj-Bouziane F, Luh WM, et al. 2013. Intrinsic structure of visual exemplar and category representations in macaque brain. *J. Neurosci.* 33:11346–60

Liu X, Yanagawa T, Leopold DA, Fujii N, Duyn JH. 2015. Robust long-range coordination of spontaneous neural activity in waking, sleep and anesthesia. *Cereb. Cortex* 25:2929–38

Logothetis NK, Sheinberg DL. 1996. Visual object recognition. *Annu. Rev. Neurosci.* 19:577–621

Luzzatti C, Davidoff J. 1994. Impaired retrieval of object-colour knowledge with preserved colour naming. *Neuropsychologia* 32:933–50

Mahon BZ, Caramazza A. 2011. What drives the organization of object knowledge in the brain? *Trends Cogn. Sci.* 15:97–103

Malach R, Levy I, Hasson U. 2002. The topography of high-order human object areas. *Trends Cogn. Sci.* 6:176–84

Malcolm GL, Groen II, Baker CI. 2016. Making sense of real-world scenes. *Trends Cogn. Sci.* 20:843–56

Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503:78–84

Martin A. 2016. GRAPES—Grounding representations in action, perception, and emotion systems: how object properties and categories are represented in the human brain. *Psychon. Bull. Rev.* 23:979–90

Martinaud O. 2017. Visual agnosia and focal brain injury. *Rev. Neurol.* 173:451–60

Matsumoto N, Eldridge MA, Saunders RC, Reoli R, Richmond BJ. 2016. Mild perceptual categorization deficits follow bilateral removal of anterior inferior temporal cortex in rhesus monkeys. *J. Neurosci.* 36:43–53

McKee JL, Riesenhuber M, Miller EK, Freedman DJ. 2014. Task dependence of visual and category representations in prefrontal and inferior temporal cortices. *J. Neurosci.* 34:16065–75

Miceli G, Fouch E, Capasso R, Shelton JR, Tomaiuolo F, Caramazza A. 2001. The dissociation of color from form and function knowledge. *Nat. Neurosci.* 4:662–67

Moeller S, Crapse T, Chang L, Tsao DY. 2017. The effect of face patch microstimulation on perception of faces and objects. *Nat. Neurosci.* 20:743–52

Monosov IE, Sheinberg DL, Thompson KG. 2010. Paired neuron recordings in the prefrontal and inferotemporal cortices reveal that spatial selection precedes object identification during visual search. *PNAS* 107:13105–10

Monosov IE, Sheinberg DL, Thompson KG. 2011. The effects of prefrontal cortex inactivation on object responses of single neurons in the inferotemporal cortex during visual search. *J. Neurosci.* 31:15956–61

Mruczek RE, Sheinberg DL. 2007. Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *J. Neurosci.* 27:2825–36

Namima T, Yasuda M, Banno T, Okazawa G, Komatsu H. 2014. Effects of luminance contrast on the color selectivity of neurons in the macaque area v4 and inferior temporal cortex. *J. Neurosci.* 34:14934–47

Naselaris T, Stansbury DE, Gallant JL. 2012. Cortical representation of animate and inanimate objects in complex natural scenes. *J. Physiol.* 106:239–49

Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, et al. 2011. Scene-selective cortical regions in human and nonhuman primates. *J. Neurosci.* 31:13771–85

Nishimoto S, Huth AG, Bilenko NY, Gallant JL. 2017. Eye movement-invariant representations in the human visual system. *J. Vis.* 17(1):11

Nishio A, Shimokawa T, Goda N, Komatsu H. 2014. Perceptual gloss parameters are encoded by population responses in the monkey inferior temporal cortex. *J. Neurosci.* 34:11143–51

O'Reilly RC. 2010. The what and how of prefrontal cortical organization. *Trends Neurosci.* 33:355–61

Op de Beeck HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ. 2008. A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cereb. Cortex* 18:1676–94

Op de Beeck HP, Wagemans J, Vogels R. 2001. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* 4:1244–52

Osher DE, Saxe RR, Koldewyn K, Gabrieli JD, Kanwisher N, Saygin ZM. 2016. Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cereb. Cortex* 26:1668–83

Parker AJ. 1896. Morphology of the cerebral convolutions with special reference to the order of primates. *J. Acad. Nat. Sci. Phila.* 10:247–362

Passingham R. 2009. How good is the macaque monkey model of the human brain? *Curr. Opin. Neurobiol.* 19:6–11

Passingham RE, Wise SP. 2012. *The Neurobiology of the Prefrontal Cortex*. New York: Oxford Univ. Press

Pitcher D, Charles L, Devlin JT, Walsh V, Duchaine B. 2009. Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curr. Biol.* 19:319–24

Rajalingham R, Schmidt K, DiCarlo JJ. 2015. Comparison of object recognition behavior in human and monkey. *J. Neurosci.* 35:12127–36

Rajimehr R, Bilenko NY, Vanduffel W, Tootell RB. 2014. Retinotopy versus face selectivity in macaque visual cortex. *J. Cogn. Neurosci.* 26:2691–700

Rajimehr R, Tootell RB. 2009. Does retinotopy influence cortical folding in primate visual cortex? *J. Neurosci.* 29:11149–52

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2:1019–25

Riley MR, Qi XL, Constantinidis C. 2017. Functional specialization of areas along the anterior-posterior axis of the primate prefrontal cortex. *Cereb. Cortex* 27:3683–97

Rolls ET. 2000. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27:205–18

Romanski LM. 2004. Domain specificity in the primate prefrontal cortex. *Cogn. Affect. Behav. Neurosci.* 4:421–29

Romero MC, Bohon KS, Lafer-Sousa L, Conway BR. 2014. *Functional organization of colors, places and faces in alert macaque frontal cortex*. Poster presented at Neuroscience 2014, Washington, DC, Nov. 15

Rosenthal I, Ratnasingam S, Haile T, Eastman S, Fuller-Deets J, Conway BR. 2018. Color statistics of objects, and color tuning of object cortex in macaque monkey. *J. Vis.* In press

Rust NC, DiCarlo JJ. 2010. Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J. Neurosci.* 30:12978–95

Sadagopan S, Zarco W, Freiwald WA. 2017. A causal relationship between face-patch activity and face-detection behavior. *eLife* 6:e18558

Sato T, Uchida G, Lescroart MD, Kitazono J, Okada M, Tanifuji M. 2013. Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *J. Neurosci.* 33:16642–56

Schalk G, Kapeller C, Guger C, Ogawa H, Hiroshima S, et al. 2017. Facephenes and rainbows: causal evidence for functional and anatomical specificity of face and color processing in the human brain. *PNAS* 114:12285–90

Sereno AB, Sereno ME, Lehky SR. 2014. Recovering stimulus locations using populations of eye-position modulated neurons in dorsal and ventral visual streams of non-human primates. *Front. Integr. Neurosci.* 8:28

Serre T. 2016. Models of visual categorization. *WIREs Cogn. Sci.* 7:197–213

Sha L, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, et al. 2015. The animacy continuum in the human ventral vision pathway. *J. Cogn. Neurosci.* 27:665–78

Sheinberg DL, Logothetis NK. 1997. The role of temporal cortical areas in perceptual organization. *PNAS* 94:3408–13

Sheinberg DL, Logothetis NK. 2001. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* 21:1340–50

Sigala N, Logothetis NK. 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415:318–20

Silson EH, Chan AW, Reynolds RC, Kravitz DJ, Baker CI. 2015. A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *J. Neurosci.* 35:11921–35

Srihasam K, Vincent JL, Livingstone MS. 2014. Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17:1776–83

Stoughton CM, Lafer-Sousa R, Gagin G, Conway BR. 2012. Psychophysical chromatic mechanisms in macaque monkey. *J. Neurosci.* 32:15216–26

Striem-Amit E, Ovadia-Caro S, Caramazza A, Margulies DS, Villringer A, Amedi A. 2015. Functional connectivity of visual cortex in the blind follows retinotopic organization principles. *Brain* 138:1679–95

Sugita Y. 2008. Face perception in monkeys reared with no exposure to faces. *PNAS* 105:394–98

Tanaka J, Weiskopf D, Williams P. 2001. The role of color in high-level vision. *Trends Cogn. Sci.* 5:211–15

Tanaka K. 1996. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19:109–39

Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y. 1999. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401:699–703

Tsao DY, Moeller S, Freiwald WA. 2008. Comparing face patch systems in macaques and humans. *PNAS* 105:19514–19

Van Essen DC. 1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature* 385:313–18

Van Essen DC, Donahue C, Dierker DL, Glasser MF. 2016. Parcellations and connectivity patterns in human and macaque cerebral cortex. In *Micro-, Meso- and Macro-Connectomics of the Brain*, ed. H Kennedy, DC Van Essen, Y Christen, pp. 89–106. New York: Springer

Van Essen DC, Zeki SM. 1978. The topographic organization of rhesus monkey prestriate cortex. *J. Physiol.* 277:193–226

Vanduffel W, Zhu Q, Orban GA. 2014. Monkey cortex through fMRI glasses. *Neuron* 83:533–50

Vaziri S, Connor CE. 2016. Representation of gravity-aligned scene structure in ventral pathway visual cortex. *Curr. Biol.* 26:766–74

Verhoef BE, Bohon KS, Conway BR. 2015. Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. *J. Neurosci.* 35:6952–68

Vernon RJ, Gouws AD, Lawrence SJ, Wade AR, Morland AB. 2016. Multivariate patterns in the human object-processing pathway reveal a shift from retinotopic to shape curvature representations in lateral occipital areas, LO-1 and LO-2. *J. Neurosci.* 36:5763–74

Walther DB, Koch C. 2007. Attention in hierarchical models of object recognition. *Prog. Brain Res.* 165:57–78

Wandell BA, Winawer J. 2011. Imaging retinotopic maps in the human brain. *Vis. Res.* 51:718–37

Warland DK, Huberman AD, Chalupa LM. 2006. Dynamics of spontaneous activity in the fetal macaque retina during development of retinogeniculate pathways. *J. Neurosci.* 26:5190–97

Watson DM, Andrews TJ, Hartley T. 2017. A data driven approach to understanding the organization of high-level visual cortex. *Sci. Rep.* 7:3596

Weisberg J, van Turennout M, Martin A. 2007. A neural system for learning about object function. *Cereb. Cortex* 17:513–21

Winawer J, Horiguchi H, Sayres RA, Amano K, Wandell BA. 2010. Mapping hV4 and ventral occipital cortex: the venous eclipse. *J. Vis.* 10(5):1

Wong RO. 1999. Retinal waves and visual system development. *Annu. Rev. Neurosci.* 22:29–47

Yamins DL, DiCarlo JJ. 2016a. Eight open questions in the computational modeling of higher sensory cortex. *Curr. Opin. Neurobiol.* 37:114–20

Yamins DL, DiCarlo JJ. 2016b. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19:356–65

Yovel G, Freiwald WA. 2013. Face recognition systems in monkey and human: Are they the same thing? *F1000Prime Rep.* 5:10