

What has fMRI Taught us About Object Recognition?

Kalanit Grill-Spector

Department of Psychology & Neuroscience Institute

Stanford University, Stanford, CA, 94305

In *Object Categorization: Computer and Human Vision Perspectives*" edited by Sven Dickinson, Aleš Leonardis, Bernt Schiele, and Michael J. Tarr. In Press, 2008.

Introduction

Humans can effortlessly recognize objects in a fraction of a second despite large variability in the appearance of objects (Thorpe et al., 1996). What are the underlying representations and computations that enable this remarkable human ability? One way to answer these questions is to investigate the neural mechanisms of object recognition in the human brain. With the advent of functional magnetic resonance imaging (fMRI) about 15 years ago, neuroscientists and psychologists began to examine the neural bases of object recognition in humans. fMRI is an attractive method because it is a non-invasive technique that allows multiple measurements of brain activation in the same awake behaving human. Among noninvasive techniques, it provides the best spatial resolution currently available, enabling us to localize cortical activations in the spatial resolution of millimeters (as fine as 1mm) and at a reasonable time scale (in the order of seconds).

Before the advent of fMRI knowledge about the function of the ventral stream was based single unit electrophysiology measurements in monkeys and on lesion studies. These studies showed that neurons in the monkey inferotemporal (IT) cortex respond to

shapes (Fujita et al., 1992) and complex objects such as faces (Desimone et al., 1984), and that lesions to the ventral stream can produce specific deficits in object recognition such as agnosia (inability to recognize objects), and prosopagnosia (inability to recognize faces) see (Farah, 1995). However, interpreting lesion data is complicated because lesions are typically diffuse (usually more than one region is damaged), lesions typically disrupt both a cortical region and its connectivity and they are not replicable across patients. Therefore, the primary knowledge gained from fMRI research was which cortical sites in the normal human brain are involved in object recognition. The first set of fMRI studies of object and face recognition in humans identified the regions in the human brain that respond selectively to objects and faces (Malach et al., 1995; Kanwisher et al., 1997; Grill-Spector et al., 1998b). Then a series of studies demonstrated that activation in object and face selective regions correlates with success at recognizing object and faces, respectively, providing striking evidence for the involvement of these region in recognition (Grill-Spector et al., 2000; Bar et al., 2001; Grill-Spector et al., 2004). Once researchers found which regions in cortex are involved in object recognition the focus of research shifted to examining the nature of representations and computations that are implemented in these regions to understand how they enable efficient object recognition in humans.

In this chapter I will review fMRI research that provided important knowledge about the nature of object representations in the human brain. I chose to focus on this topic because results from these experiments provide important insights that can be used by computer scientists when they design artificial object recognition systems. For example, one of the fundamental problems in recognition is how to recognize an object

across variations in its appearance (invariant object recognition). Understanding how a biological system has solved this problem may give hints for how to build a robust artificial recognition system. Further, fMRI is more adequate for measuring object representations than the temporal sequence of computations en route to object recognition because the time scale of fMRI measurements is longer than the time scale of the recognition process (the temporal resolution of fMRI is in the order of seconds, whereas object recognition takes about 100-250ms). Nevertheless, combining psychophysics with fMRI may give us some clues about what kind of visual processing is implemented in distinct cortical regions. For example, finding regions whose activation is correlated with success at some tasks, but not others, may suggest the involvement of particular cortical regions in one computation, but not another.

In discussing how fMRI has impacted our current understanding of object representations I will focus on results pertaining to two aspects of object representation:

- How do the underlying representations provide for invariant object recognition?
- How is category information represented in the ventral stream?

I have chosen these topics because they are central topics in object recognition for which fMRI has substantially advanced our understanding, some findings related to these topics stirred considerable debate, (see *Debates About the Nature of Functional Organization in the Human Ventral Stream*) and some of the fMRI findings in humans are surprising given prior knowledge from single-unit electrophysiology in monkeys. In organizing the chapter, I will begin with a brief introduction of the functional organization of the human ventral stream and definition of object-selective cortex. Then I will describe research that elucidated the properties of these regions with respect to basic

coding principles, I will continue with findings related to invariant object recognition and end with research and theories regarding category representation and specialization in the human ventral stream.

The Functional Organization of the Human Ventral Stream

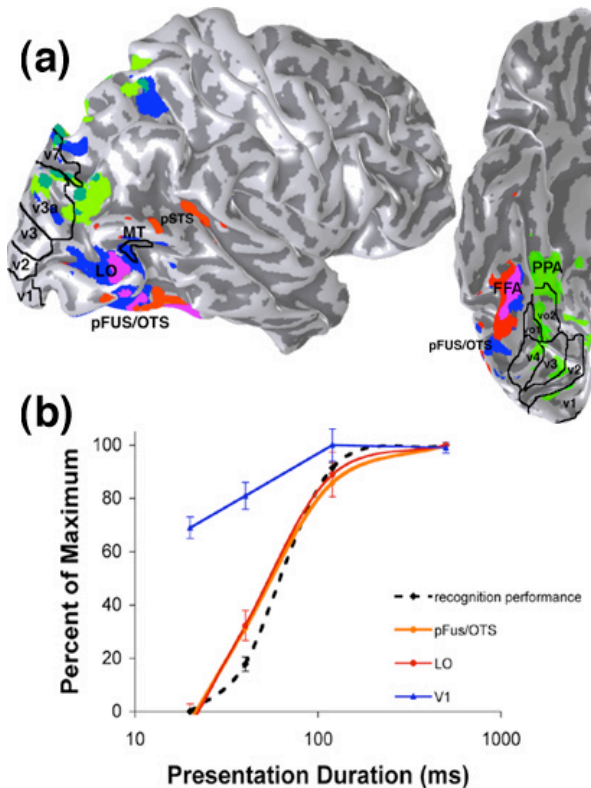


Figure 1: Object-, face- and place-selective cortex. (a) Data of one representative subject shown on her partially inflated right hemisphere. Left: lateral view; Right: ventral view. Dark gray: sulci; Light gray: gyri. Black lines delineate retinotopic regions. Blue: Object-selective regions (objects > scrambled objects) including LO and pFus/OTS ventrally as well as dorsal foci along the intraparietal sulcus (IPS). Red: face-selective regions (faces > non face objects), including the FFA, a region in lateral occipital cortex and a region in posterior STS. Magenta: overlap between face and object selective regions. Green: place-selective regions (places > objects) including the PPA and a dorsal region lateral to the IPS. Dark green: overlap between place and object selective. All maps thresholded at $p < 0.001$, voxel level. (b) LO and pFus (but not V1) responses are correlated with recognition performance (Ungerleider et al., 1983; Grill-Spector et al., 2000). To superimpose recognition performance and fMRI signals on the same plot all values were normalized relative to the maximum response for the 500ms duration stimulus. For fMRI signals (blue, red and orange) = $\frac{\% \text{ signal}(\text{condition})}{\% \text{ signal}(500\text{ms})}$. For recognition performance (black) = $\frac{\% \text{ correct}(\text{condition})}{\% \text{ correct}(500\text{ms})}$.

The first set of fMRI studies on object and face recognition in humans was devoted to identifying the regions in the brain that are object and face selective. Electrophysiology research in monkeys suggested that neurons in higher level regions respond to shapes and objects more than simple stimuli such as lines, edges and patterns (Desimone et al., 1984; Fujita et al., 1992; Logothetis et al., 1995). Based on these

findings fMRI studies measured brain activation when people viewed pictures of objects compared to when people viewed scrambled objects (have the same local information and statistics, but do not contain an object) or texture patterns (e.g. checkerboards, which are robust visual stimuli, but do not elicit a percept of a global form). These studies found a constellation of regions in the lateral occipital cortex (termed the lateral occipital complex, LOC) beginning around the lateral occipital sulcus, posterior to MT, extending ventrally into the occipito-temporal sulcus (OTS) and the fusiform gyrus (Fus) that respond more to objects than controls. The LOC is located lateral and anterior to early visual areas (V1-V4), (Grill-Spector et al., 1998a; Grill-Spector et al., 1998b) and is typically divided into two subregions, LO: a region in lateral occipital cortex, adjacent and posterior to MT and pFus/OTS: a ventral region overlapping the OTS and posterior fusiform gyrus (Fig. 1).

LOC responds similarly to many kinds of objects and object categories (including novel objects) and is thought to be in the intermediate or high-level stages of the visual hierarchy. Importantly, LOC activations are correlated with subjects' object recognition performance. High LOC responses correlate with successful object recognition (hits), and low LOC responses correlate with trials in which objects are present, but are not recognized (misses) see Fig. 1b. There are also object selective regions in the dorsal stream (Grill-Spector, 2003; Grill-Spector and Malach, 2004), but these regions do not correlate with object recognition performance (Fang and He, 2005) and may be involved in computations related to visually guided actions towards objects (Culham et al., 2003). However, a comprehensive discussion of the dorsal stream's role in object perception is beyond the scope of this chapter.

In addition to the LOC, researchers found several additional ventral regions that show preferential responses to specific object categories. Searching for regions with categorical preference was motivated by reports that suggested that lesions to the ventral stream can produce very specific deficits - such as the inability to recognize faces, or the inability to read words, while other visual (and recognition) faculties are preserved. By contrasting activations to different kinds of objects researchers found ventral regions that show higher responses to specific object categories such as lateral fusiform regions that respond more to animals than tools, (and medial fusiform regions that respond to tools more than animals, (Martin et al., 1996; Chao et al., 1999)), a region in the left OTS that responds more strongly to letters than textures, (the “visual word form area”, VWFA, (Cohen et al., 2000)), several foci that respond more strongly to faces than other objects (Kanwisher et al., 1997; Haxby et al., 2000; Hoffman and Haxby, 2000; Grill-Spector et al., 2004) including the well known “fusiform face area”, FFA (Kanwisher et al., 1997), regions that respond more strongly to houses and places than faces and objects (including a region in the parahippocampal gyrus, the “parahippocampal place area”, PPA (Epstein and Kanwisher, 1998)) and regions that respond more strongly to body parts than faces and objects (including a region near MT called the “extrastriate body area”, EBA (Downing et al., 2001) and a region in the fusiform gyrus, “the fusiform body area”, FBA (Schwarzlose et al., 2005)). Nevertheless, many of these object-selective and category-selective regions respond to more than one object category and also respond strongly to object fragments (Grill-Spector et al., 1998b; Lerner et al., 2001; Lerner et al., 2008). This suggests caution when interpreting the nature of the selective responses. It is

possible that the underlying representation is perhaps of object parts, features and/or fragments and not of whole objects or object categories.

Findings of category selective regions in the human brain initiated a fierce debate about the principles of functional organization in the ventral stream. Are there regions in the cortex that are specialized for any object category? How abstract is the information represented in these regions? E.g., is category information represented in these regions, or low-level visual features that are associated with categories? I will address these questions in detail in the section *Debates About the Nature of Functional Organization in the Human Ventral Stream*.

Cue-Invariant Responses in the LOC

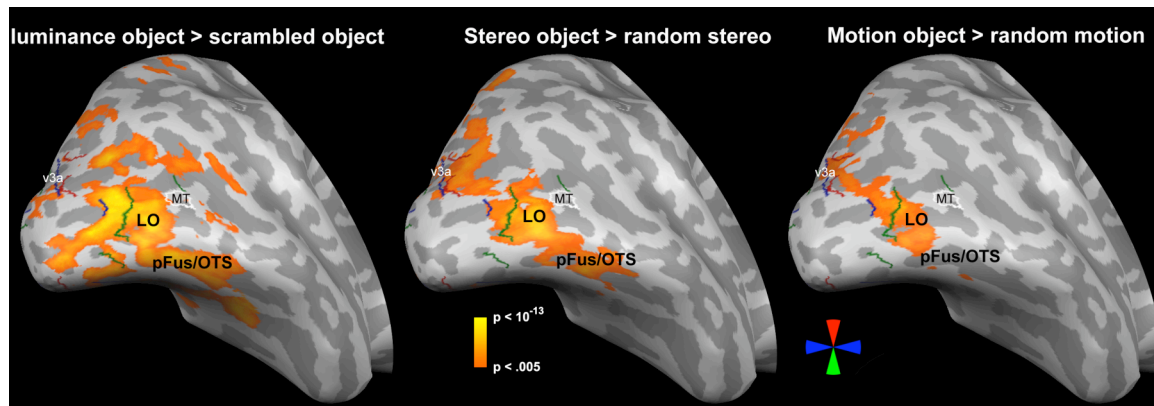


Figure 2: Selective responses to objects across multiple visual cues across the LOC. Statistical maps of selective response to object from luminance, stereo and motion information in a representative subject. All maps were thresholded at $P < 0.005$, voxel level, and are shown on the inflated right hemisphere of a representative subject. **(a)** luminance objects > scrambled objects. **(b)** Objects generated from random dot stereograms vs. structureless random dot stereograms (perceived as a cloud of dots) and **(c)** Objects generated from dot motion vs. the same dots moving randomly. Visual meridians are represented by the red, blue and green lines: *red*: upper; *blue*: horizontal; *green*: lower. White contour: motion selective region, MT.

While findings of object-selective responses in the human brain were suggestive of the involvement of these region in processing objects, there are many differences

between objects and scrambled objects (or objects and texture patterns). Objects have a shape, surfaces, contours, they are associated with a meaning and semantic information and they are generally more interesting than texture patterns. Each of these factors may affect the higher fMRI response to objects than controls. Further, differences in low-level visual properties across objects and controls may be driving some of these effects.

Converging evidence from several studies revealed an important aspect of coding in the LOC: it responds to object shape, not low-level visual features. Several studies showed that all LOC subregions (LO and pFus/OTS) are activated more strongly when subjects view objects independent of the type of visual information that defines the object form (Grill-Spector et al., 1998a; Kastner et al., 2000; Kourtzi and Kanwisher, 2000, 2001; Gilaie-Dotan et al., 2002; Vinberg and Grill-Spector, 2008), see Fig. 2. LOC responds more strongly to (1) objects defined by luminance compared to luminance textures, (2) objects generated from random dot stereograms than structureless random dot stereograms, (3) objects generated from structure from motion relative to random (structureless) motion and (4) objects generated from textures than texture patterns. LOC's response to objects is also similar across object format* (gray-scale, line drawings, silhouettes) and it responds to objects delineated by both real and illusory contours (Mendola et al., 1999; Stanley and Rubin, 2003). Kourtzi and Kanwisher (Kourtzi and Kanwisher, 2001) also showed that when objects have the same shape but different contours there is fMRI-adaptation (fMRI-A, indicating a common neural substrate), but there is no fMRI-A when the shared contours were identical but the perceived shape was different, suggesting that LOC responds to global shape rather than local contours, see

* Selective responses to faces and houses across stimulus format (photographs, line drawings, and silhouettes) have also been shown for the FFA and PPA, respectively.

also (Lerner et al., 2002; Kourtzi et al., 2003). Overall, these studies provided fundamental knowledge by showing that LOC's activation is driven by shape rather than low-level visual information that generates form.

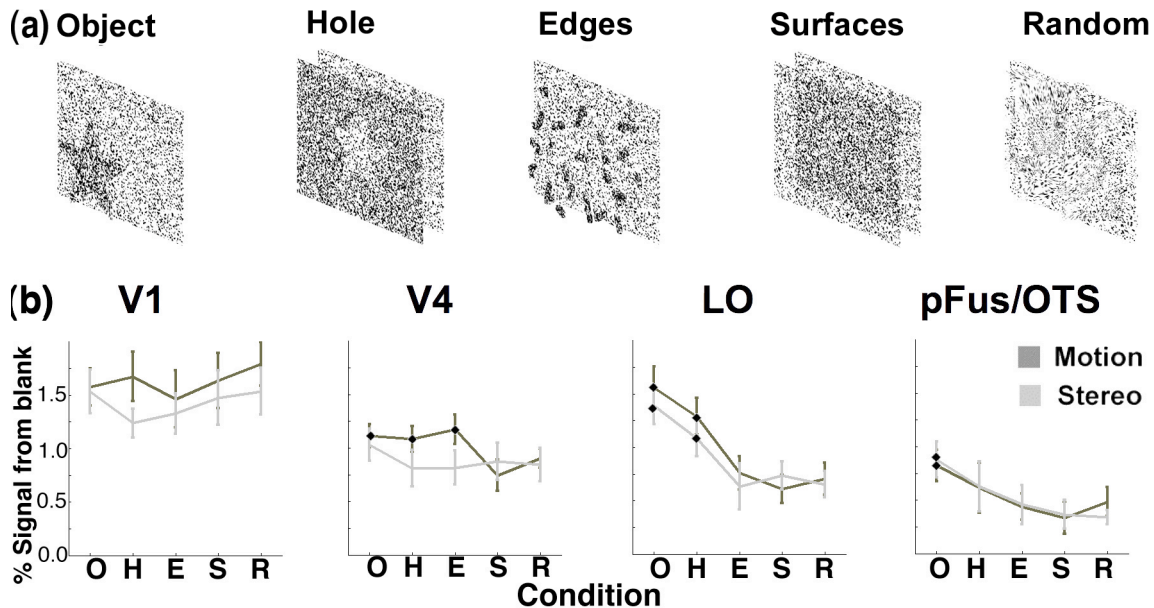


Figure 3: Responses to shape, edges and surfaces across the ventral stream. Adapted from (Vinberg and Grill-Spector, 2008). (a) Schematic illustration of experimental conditions. Stimuli were generated from either motion or stereo information alone and had no luminance edges or surfaces (except for the screen border which was present during the entire experiment including blank baseline blocks). For illustration purposes darker regions indicate front surfaces. From left to right: *Object* on the front surface in front of a flat background plane. *Hole* on the front surface in front of a flat background. (c) *Disconnected edges* in front of a flat background. Edges were generated by scrambling the shape contours. *Surfaces*: Two semitransparent flat surfaces at different depths. *Random* stimuli with no coherent structure, edges, global surfaces or global shape. Random stimuli had the same relative disparity or depth range as other conditions. See examples of stimuli: <http://www-psych.stanford.edu/~kalanit/jnpstim/> (b) Responses to objects, holes, edges and global surfaces across the visual ventral processing hierarchy. Responses: mean \pm SEM across 8 subjects. *O*: object; *H*: hole; *S*: surfaces; *E*: edges; *R*: random. *Diamonds*: significantly different than random at $P < 0.05$.

In a recent study we further examined whether LOC response to objects is driven by their global shape or surface information and whether LOC subregions are sensitive to border ownership. One open question in object recognition is whether the region in the image that belongs to the object is first segmented from the rest of the image (figure-ground segmentation) and then recognized, or whether knowing the shape of an object

aids its segmentation (Peterson and Gibson, 1994b, 1994a; Nakayama et al., 1995). To address these questions, we scanned subjects when they viewed stimuli that were matched for their low-level information and generated different percepts: (1) a percept of an object in front of a background object, (2) a shaped hole (same shape as the object) in front of a background, (3) two flat surfaces without shapes, (4) local edges (created by scrambling the object contour) in front of a background, or (5) random dot stimuli with no structure (Fig. 3a, and (Vinberg and Grill-Spector, 2008)). We repeated the experiment twice, once with random dots that were presented stereoscopically and once with random dots that moved. We found that LOC responses (both LO and pFUS/OTS) were higher for objects and shaped holes than either surfaces, local edges or random stimuli (Fig. 3b). These results we observed for both motion and stereo cues. In contrast, LOC responses were not higher for surfaces than random stimuli and were not higher for local edges than random stimuli. Thus, adding either local edge information or global surface information, does not increase LOC response. However, adding a global shape produces a significant increase in LOC response. These results provide clear evidence that cue invariant responses in the LOC are driven by object shape, rather than global surface information or local edge information.

Interestingly, recent evidence shows that the LOC is sensitive to border ownership/figure-ground segmentation (Appelbaum et al., 2006; Vinberg and Grill-Spector, 2008). We found that LO and pFus/OTS responses were higher for objects (shapes presented in the foreground) than the same shapes when they defined holes in the foreground. The difference between the objects and the holes was the assignment of the figure region (or border ownership of the contour defining the shape). This higher

response to objects than holes was a unique characteristic of LOC subregions and did not occur in other visual regions (Fig. 3). This result suggests that LOC prefers shapes (and contours) when they define the figure region. One implication of this result is that the same brain machinery may be involved in both object recognition and in determining which region in the visual input contains the figure region. Thus, one consideration for computer scientists is that an effective object recognition algorithm should be able to determine both what is the object in the scene as well as determine which region in the scene corresponds to the object.

Neural Bases of Invariant Object Recognition

The literature reviewed so far provides accumulating evidence that LOC is involved in recognition and processing of object form. The next question that one may ask, given LOC's role in object perception, how does it deal with the variability in objects' appearance? There are many factors that can affect the appearance of objects. Changes in objects appearance can occur as a result of the object being at different locations relative to the observer, which will affect the retinal projection of objects in terms of their size and position. Also, the 2-dimensional projection of a 3-dimensional object on the retina varies considerably due to changes in its rotation and viewpoint relative to the observer. Other changes in appearance occur because of differential illumination conditions, which affect objects' color, contrast and shadowing. Nevertheless, humans are able to recognize objects across large changes in their appearance, which is referred to as *invariant object recognition*.

A central topic of research in the study of object recognition, is understanding how invariant recognition is accomplished. One view suggests that invariant object recognition is accomplished because the underlying neural representations are invariant to the appearance of objects. Thus, there will be similar neural responses even when the appearance of an object changes considerably. One means by which this can be achieved is by extracting from the visual input features or fundamental elements (such as geons (Biederman, 1987) that are relatively insensitive to changes in objects' appearance. According to one influential model (Recognition by components, RBC, Biederman, 1987) objects are represented by a library of geons (that are easy to detect in many viewing conditions) and their spatial relations. Other theories suggest that invariance may be generated through a sequence of computations across a hierarchically organized processing stream in which the level of invariance increases from one level of the processing to the next. For example, at the lowest level neurons code local features and in higher levels of the processing stream neurons respond to more complex shapes and are less sensitive to changes in position and size (Riesenhuber and Poggio, 1999).

Neuroimaging studies of invariant object recognition found differential sensitivity across the ventral stream to object transformations such as size, position, illumination and viewpoint. Intermediate regions such as LO show higher sensitivity to image transformations than higher-level regions such as pFus/OTS. Notably, accumulating evidence from many studies suggests that at no point in the ventral stream neural representations are *entirely* invariant to object transformations. These results support an account in which invariant recognition is supported by a pooled response across neural populations that are sensitive to object transformations. One way in which this can be

accomplished is by a neural code that contains independent sensitivity to object information and object transformation (DiCarlo and Cox, 2007). For example, neurons may be sensitive to both object category and object position. As long as the categorical preference is retained across object transformations invariant object information can be extracted.

Object and Position Information in the LOC

One variation that the object recognition system needs to deal with is variation in the size and position of objects. Size and position invariance are thought to be accomplished in part by an increase in the size of neural receptive fields along the visual hierarchy. That is, as one ascends the visual hierarchy, neurons respond to stimuli across a larger part of the visual field. At the same time a more complex visual stimulus is necessary to elicit significant responses in neurons (e.g. shapes instead of oriented lines). Findings from electrophysiology suggest that even at the highest stages of the visual hierarchy neurons retain some sensitivity to object location and size (though electrophysiology reports vary significantly about the degree of position sensitivity of IT neurons (Op De Beeck and Vogels, 2000; Rolls, 2000; DiCarlo and Maunsell, 2003)). A related issue is whether position sensitivity of neurons in higher visual areas manifests as an orderly, topographic representation of the visual field. Researchers have examined position and size sensitivity in the LOC and nearby cortex (such as PPA and FFA) using measurements of the mean response across a region of interest (ROI), fMRI-A, in which they measured sensitivity to changes in object size or position, and examination of the distributed response across the ventral stream to the same object or object category across

sizes and positions.

Several studies documented sensitivity to both eccentricity and polar angle in distinct ventral stream regions. Both object-selective and category-selective regions in the ventral stream respond to objects presented at multiple positions and sizes. However, the amplitude of response to object varies across different retinal positions. LO, pFUS/OTS as well as category-selective regions (e.g. FFA, PPA) respond more strongly to objects presented in the contralateral vs. ipsilateral visual field (Grill-Spector et al., 1998b; Hemond et al., 2007; McKyton and Zohary, 2007). Some regions (LO and EBA) also respond more strongly to objects presented in the lower visual field (Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008). Responses also vary with eccentricity: LO, FFA and the VWFA respond more strongly to centrally presented stimuli, and the PPA responds more strongly to peripherally presented stimuli (Levy et al., 2001; Hasson et al., 2002; Hasson et al., 2003; Sayres and Grill-Spector, 2008).

Using fMRI-A, my colleagues and I have shown that pFus/OTS, but not LO exhibits some degree of invariance to objects' size and position (Grill-Spector et al., 1999). fMRI-A is a method that allows characterizing the sensitivity of neural representations to stimulus transformations at a sub-voxel resolution. fMRI-A is based on findings from single unit electrophysiology showing that when objects repeat, there is a stimulus-specific decrease in IT cells' response to the repeated image, but not to other object images (Miller et al., 1991; Sawamura et al., 2006). Similarly, fMRI signals in higher visual regions show a stimulus-specific reduction (fMRI-A) in response to repetition of identical object images (Grill-Spector et al., 1999; Grill-Spector and Malach, 2001; Grill-Spector et al., 2006a). We showed that fMRI-A can be used to test the

sensitivity of neural responses to object transformation by adapting cortex with a repeated presentation of an identical stimulus and examining adaptation effects when the stimulus is changed along an object transformation (e.g. changing its position). If the response remains adapted, it indicates that neurons are insensitive to the change. However, if responses return to the initial level (recover from adaptation) it indicates sensitivity to the change (Grill-Spector and Malach, 2001).

Using fMRI-A we found that repeated presentation of the same face or object at the same position and size produces reduced fMRI activation. This is thought to reflect stimulus-specific neural adaptation. Presenting the same face or object in different positions in the visual field or at different sizes also produces fMRI-A in pFUS/OTS and FFA, indicating insensitivity to object size and position (Grill-Spector et al., 1999), see also (Vuilleumier et al., 2002). This result is consistent with electrophysiology findings that showed that IT neurons that respond similarly to stimuli at different positions in the visual field also show adaptation when the same object is shown in different positions (Lueschow et al., 1994). In contrast, LO recovered from fMRI-A to images of the same face or object when presented at different sizes or positions. This indicates that LO is sensitive to object position and size.

Recently, several groups examined the sensitivity of the distributed response across the visual stream to object category and object position (Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008) and also object identity and object position (Eger et al., 2008). These used pattern analyses (PA) and classifier methods developed in machine learning to examine what information is present in the distributed responses across voxels in a cortical region. The distributed response can carry different information from the

mean response of an ROI when there is variation across voxels' responses.

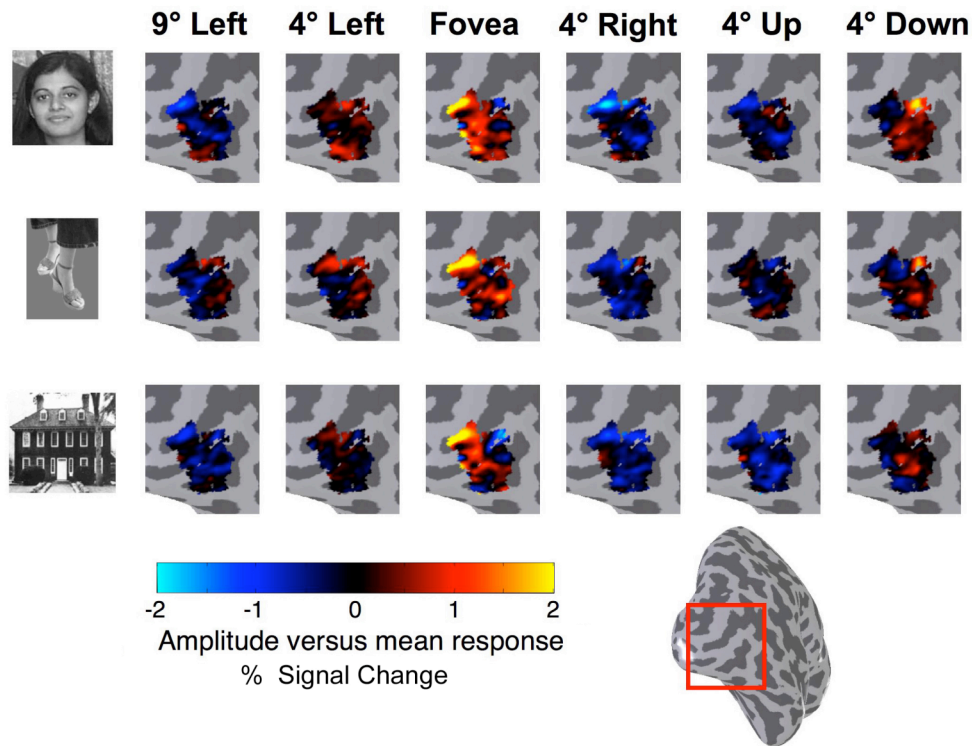


Figure 4: LO distributed response patterns to different object categories and stimulus positions. Data are shown on the lateral aspect of the right hemisphere cortical surface for a representative subject. Adapted from (Sayres and Grill-Spector, 2008). Each panel shows the distributed LO fMRI amplitudes after subtracting from each voxel its mean response. *Red and yellow*: responses that are higher than the voxel's mean response. *Blue and cyan*: responses that are lower than the voxel's mean response. *Inset*: Inflated right cortical hemisphere, with red square indicating the zoomed region. Note that pattern change significantly across columns (positions) and to a lesser extent rows (categories).

To examine sensitivity to position information several studies examined whether distributed response patterns to same object category (or object exemplar) is the same (or different) when the same stimulus is presented in a different position in the visual field. In PA researchers typically split the data into two independent sets and examine the cross-correlation between the distributed responses to the same (or different) stimulus in the same (or different) position across the two datasets. This gives a measure of the sensitivity of distributed responses to object information and position. If responses are

position invariant, there will be a high correlation between the distributed responses to the same object category (or exemplar) at different positions. If responses are sensitive to position, there will be a low correlation between responses to the same object category (or exemplar) at different positions.

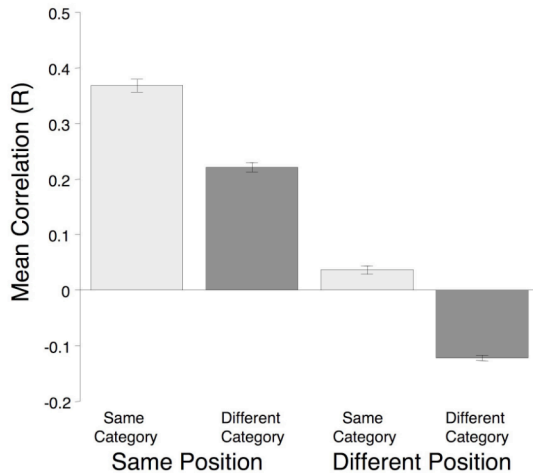


Figure 5: Mean cross correlations between LO distributed responses across two independent halves of the data for the same or different category at the same or different position in the visual field. *Position effects:* LO response patterns to the same category were substantially more similar if they were presented at the same position versus different positions (first and third bars, $p < 10^{-7}$). *Category effects:* the mean correlation was higher for same-category response patterns than for different-category response patterns when presented in the same retinotopic position (first two bars; $p < 10^{-4}$). Error bars indicate SEM across subjects. Adapted from (Sayres and Grill-Spector, 2008)

Fig. 4 illustrates distributed LO responses to three categories: houses, limbs and faces at six visual locations (fovea; 4° up, right, down, or left from the fovea; 9° left of fovea). Activation patterns for the same object category presented at different positions vary considerably (compare the response patterns in the same row across the different columns in Fig. 4). There is also variation (but to a lesser extent) across different object categories when presented at the same retinal position (same column, different rows in Fig. 4). Surprisingly, position effects in LO were larger than category effects. That is, showing objects from the same category, but at a different position significantly reduced the correlation between activation patterns (Fig. 5, first vs. third bars) more than changing the object category in the same position (Fig. 5, first vs. second bar). Importantly, position and category effects were independent, as there were no significant interactions

between position and category (all F values < 1.02 , all P values > 0.31). Thus changing both object category and position produced maximal decorrelation between distributed responses (Fig. 5, fourth bar).

We also examined whether position sensitivity in LO is manifested as an orderly topographic map (similar to retinotopic organization in lower visual areas), by measuring retinotopic maps in LO using standard traveling wave paradigms (Wandell, 1999; Sayres and Grill-Spector, 2008). We found a continuous mapping of the visual field in LO both in terms of eccentricity and polar angle. This topographic map contained an over-representation of the contra-lateral and lower visual field (more voxels preferred these visual field positions than ipsilateral and upper visual fields). While we did not consistently find a single visual field map (a single hemifield or quarterfield representation) in LO, this analysis suggests that there is preserved retinotopic information in LO which may underlie the position effects observed in analyses of distributed LO responses.

Overall, our data show that different object categories produce relatively small changes to both the mean and distributed response across LO (categorical effects are larger in the fusiform and parahippocampal gyri). In comparison, a modest 4° change in an object's position produces signal changes in LO which are as large or larger than this modulation. This 4° displacement is well within the range for which humans can categorize and detect objects (Thorpe et al., 2001). This indicates a difference between the position sensitivity of recognition behavior, and that of neural populations in LO. Though it is possible that performance on more difficult recognition tasks, such within-category recognition is more sensitive to object displacement.

A related recent study examined position sensitivity using PA more broadly across the ventral stream and provides additional evidence for a hierarchical organization across the ventral stream (Schwarzlose et al., 2008). Schwarzlose and colleagues found that distributed responses to a particular object category (faces, body parts or scenes) was similar across positions in ventro-temporal regions (e.g., pFUS/OTS and FBA) but changed across positions in occipital regions (e.g., EBA and LO). Thus, accumulating evidence from both fMRI-A and PA studies suggests a hierarchy of representations in the human ventral stream through which representations become less sensitive to object position as one ascends the visual hierarchy.

Implications for Theories of Object Recognition

It is important to relate imaging results to the concept of position-invariant representations of objects and object categories. What exactly is implied by the term "invariance" depends on the scientific context. In some instances, this term is taken to reflect a neural representation that is abstracted so as to be independent of viewing conditions. A fully invariant representation, in this meaning of the term, is expected to be completely independent of retinal position information (Biederman and Cooper, 1991). However, in the context of studies of visual cortex, the term is more often considered to be a graded phenomenon, in which neural populations are expected to retain some degree of sensitivity to visual transformations (like position changes) but in which stimulus *selectivity* is preserved across these transformations (Kobatake and Tanaka, 1994; Rolls and Milward, 2000; DiCarlo and Cox, 2007). In support of this view, a growing literature suggests that maintaining local position information within a distributed neural representation may actually aid invariant recognition in several ways (Dill and Edelman,

2001; DiCarlo and Cox, 2007; Sayres and Grill-Spector, 2008). First, maintaining separable information about position and category may also allow maintaining information about the structural relationships between object parts (Edelman and Intrator, 2000). Indeed some experiments suggest that LO may contain both an object-based (McKyton and Zohary, 2007) and a retinal-based reference frame (Sayres and Grill-Spector, 2008). The former reference frame may provide a basis for structural encoding. Second, separable position and object information may provide a robust way for generating position invariance by using a population code. Under this model objects are represented as manifolds in a high dimensional space spanned by a population of neurons. The separability of position and object information may allow for fast decisions based on linear computations (e.g. linear discriminant functions) to determine the object identity (or category) across positions see (DiCarlo and Cox, 2007). Finally, separable object and position information may allow concurrent localization and recognition of objects. That is, recognizing what the object is and also determining where it is.

Evidence for Viewpoint Sensitivity Across the LOC

Another source of change in object appearance that merits separate consideration is changes across rotation in depth. In contrast to position or size changes, where invariance may be achieved by a linear transformation, the shape of objects changes with depth rotation. This is because the visual system receives 2D retinal projections of 3D objects. Some theories suggest that view-invariant recognition across object rotations and/or changes in the observer viewing angle is accomplished by largely view-invariant representations of objects (generalized cylinders', (Marr, 1980); Recognition by

Components', RBC', (Biederman, 1987). That is, the underlying neural representations respond similarly to an object across its views. However, other theories suggest that object representations are view-dependent, that is, they consist of several 2D views of an object (Ullman, 1989; Poggio and Edelman, 1990; Bulthoff and Edelman, 1992; Edelman and Bulthoff, 1992; Bulthoff et al., 1995; Tarr and Bulthoff, 1995). Invariant object recognition is accomplished by interpolation across these views (Ullman, 1989; Poggio and Edelman, 1990; Logothetis et al., 1995) or by a distributed neural code across view-tuned neurons (Perrett et al., 1998).

Single unit electrophysiology studies in primates indicate that the majority of neurons in monkey inferotemporal cortex are view-dependent (Desimone et al., 1984; Logothetis et al., 1995; Perrett, 1996; Wang et al., 1996; Vogels and Biederman, 2002) with a small minority (5-10%) of neurons showing view-invariant responses across object rotations (Logothetis et al., 1995; Booth and Rolls, 1998).

In humans, results vary considerably. Short-lagged fMRI-A experiments, in which the test stimulus is presented immediately after the adapting stimulus (Grill-Spector et al., 2006a), suggest that object representations in the lateral occipital complex are view-dependent (Grill-Spector et al., 1999; Gauthier et al., 2002; Fang et al., 2007), but see (Valyear et al., 2006). However, long-lagged fMRI-A experiments in which many intervening stimuli occur between the test and adapting stimulus (Grill-Spector et al., 2006a), have provided some evidence for view-invariant representations in ventral LOC, especially in the left hemisphere (James et al., 2002; Vuilleumier et al., 2002) and the PPA, (Epstein et al., 2008). Also, a recent study showed that the distributed LOC responses to objects remained stable across 60° rotations (Eger et al., 2008). Presently,

there is no consensus across experimental findings in the degree to which ventral stream representations are view-dependent or view-invariant. These variable results may reflect differences in the neural representations depending on object category and cortical region, and/or methodological differences across studies (e.g. level of object rotation and fMRI-A paradigm used).

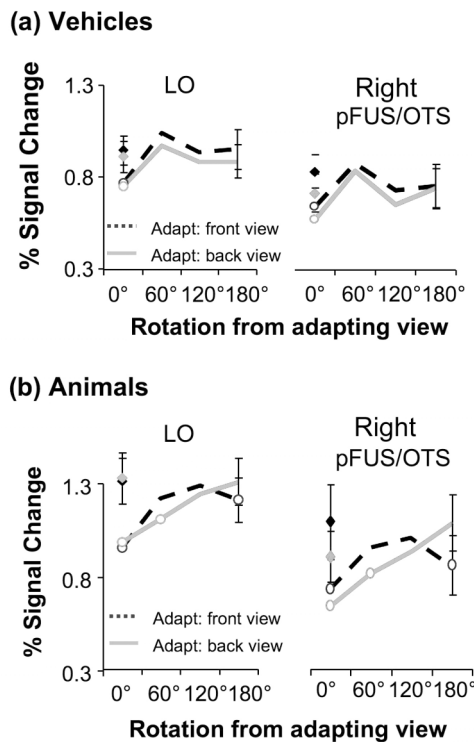


Figure 6: LO responses during fMRI-A experiments of rotation sensitivity

Each line represents response after adapting with a front (dashed black) or back view (solid gray) of an object. The nonadapted response is indicated by the diamonds (black for front view and gray for back view). The open circles indicate significant adaptation, lower than nonadapted, $P < 0.05$, paired t-test across subjects; (a) shows vehicle data and (b) animal data. Responses are plotted relative to a blank fixation baseline. Error bars indicate SEM across 8 subjects.

To address these differential findings, in a recent study we used a parametric approach to investigating sensitivity to object rotation and used a computational model to link between putative neural tuning and resultant fMRI measurements (Andresen et al., 2008, In press). The parametric approach allows a richer characterization of rotation sensitivity because it measures the degree of sensitivity to rotations rather than characterizing representations as one of two possible alternatives: “invariant” or “not invariant”. We used fMRI-A to measure viewpoint sensitivity as a function of the rotation

level for two object categories: animals and vehicles. Overall, we found sensitivity to object rotation in the LOC. However, there were differences across categories and regions. First, there was higher sensitivity to vehicle rotation than animal rotation. Rotations of 60° produced a complete recovery from adaptation for vehicles, but rotations of 120° were necessary to produce recovery from adaptation for animals (Fig. 6). Second, we found evidence for over-representation of the front view of animals in the right pFus/OTS: its responses to animals were higher for the front view than the back view (compare black and gray circles in Fig. 6b-right). In addition fMRI-A effects across rotation varied according to the adapting view (Fig. 6b-right). When adapting with the back view of animals we found recovery from adaptation for rotations of 120° or larger, but when adapting with the front view of animals there was no significant recovery from adaptation across rotations. One interpretation is that there is less sensitivity to rotation when adapting with front views than back views of animals. However, subjects' behavioral performance in a discrimination task across object rotations showed that they are equally sensitive to rotations (performance decreases with rotation level) whether rotations are relative to the front or back of an animal (Andresen et al., 2008, In press), suggesting that this interpretation is unlikely. Alternatively, the apparent rotation cross-adaptation may be due to lower responses for back views of animals. That is the apparent adaptation across rotation from the front view to the back view is driven by lower responses to the back view rather than adaptation across 180° rotations.

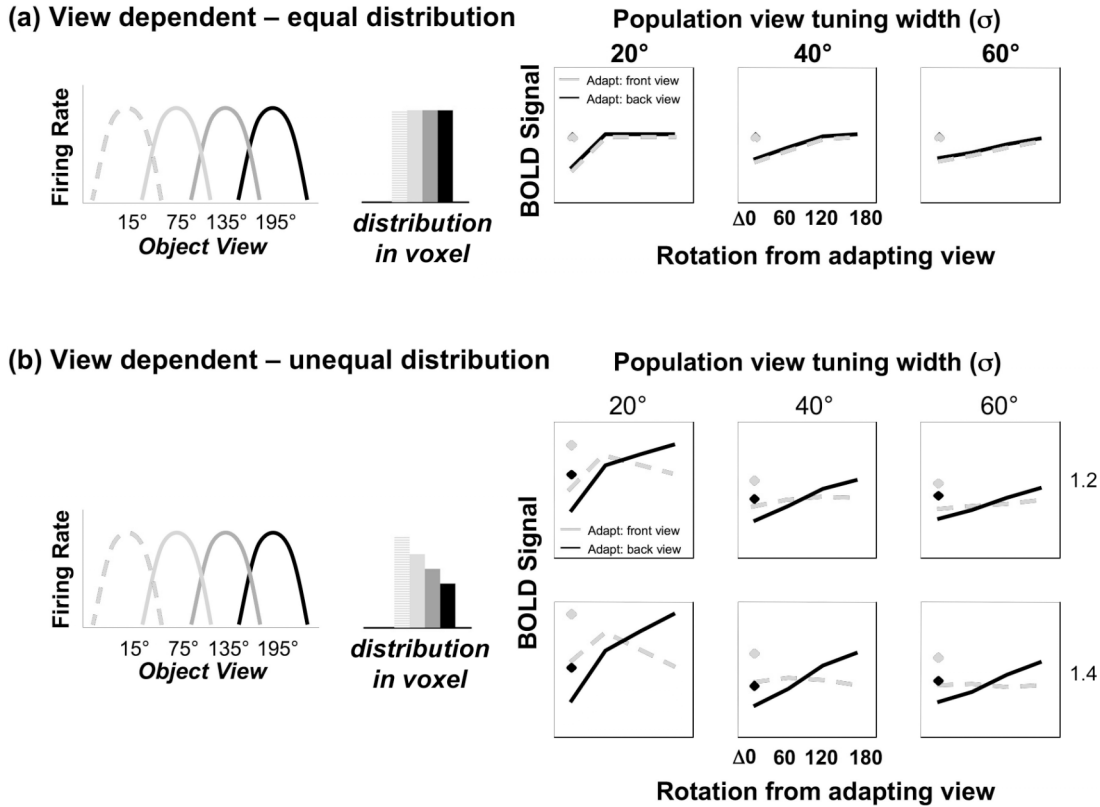


Figure 7: Simulations predicting fMRI responses of putative voxels containing a mixture of view-dependent neural populations. *Left:* Schematic illustration of the view tuning and distribution of neural populations tuned to different views in a voxel. *Right:* result of model simulations illustrating the predicted fMRI-A data. In all panels the model includes 6 Gaussians tuned to specific views around the viewing circle, separated 60° apart. Across columns the view tuning width varies, across rows the distribution of neural populations preferring specific views varies. *Diamonds:* responses without adaptation, black: back view; gray: front view; *Lines:* response after adaptation with a front view (dashed gray line) or back view (solid black line). **(a)** Mixture of view-dependent neural populations that are equally distributed in a voxel. Narrower tuning (left) shows recovery from fMRI-A for smaller rotations than wider view tuning (right). This model predicts the same pattern of recovery from adaptation when adapting with the front or back view. **(b)** Mixture of view-dependent neural populations in a voxel with a higher proportion of neurons that prefer the front view. The number on the right indicates the ratio between the percentages neurons tuned to the front vs. back view. Top row: ratio=1.2; bottom: ratio=1.4. Because there are more neurons tuned to the front view in this model, it predicts higher BOLD responses to frontal views without adaptation (gray vs. black diamonds) and a flatter profile of fMRI-A across rotations when adapting with the front view.

To better characterize the underlying representations and examine which representations may lead to our observed results, we simulated putative neural responses in a voxel and predicted the resultant BOLD responses. In the model, each voxel contains

a mixture of neural populations, each tuned to a different object view (Fig. 7 and (Andresen et al., 2008, In press)). BOLD responses were modeled to be proportional to the sum of responses across all neural populations. We simulated the BOLD responses in fMRI-A experiments for a set of putative voxels that varied in the view tuning width of neural populations, the preferred view of different neural populations, the number of different neural populations, and the distribution of populations tuned to different views within a voxel. Results of the simulations indicate that two main parameters affected the pattern of fMRI data: (1) the view tuning width of the neural population and (2) the proportion of neurons in a voxel that prefer a specific object view.

Fig. 7a shows the response characteristics of a model of a putative voxel containing a mixture of view-dependent neural populations tuned to different object views, in which the distribution of neurons tuned to different views is uniform. In this model, Narrower tuning (left) shows recovery from fMRI-A for smaller rotations than wider view tuning (right). Responses to front and back views are identical when there is no adaptation (Fig. 7a, diamonds), and the pattern of adaptation as a function of rotation is similar when adapting with the front or back views (Fig. 7a). Such a model provides an account of responses to vehicles across object-selective cortex (as measured with fMRI), and for animals in LO. Thus, this model suggests that the difference between the representation of animals and vehicles in LO is likely due to a smaller population view tuning for vehicles than animals (a tuning width of $\sigma < 40^\circ$ produces complete recovery from adaptation for rotations larger than 60° , as observed for vehicles). Fig. 7b shows that the simulations predict that when there are more neurons in a voxel that are tuned to the front view, there will be higher BOLD responses to frontal views without adaptation (gray vs.

black diamonds) and a flatter profile of fMRI-A across rotations when adapting with the front view.

Implications for Theories of Object Recognition

Overall, our results provide empirical evidence for view-dependent object representation across human object-selective cortex that is evident both with standard fMRI and fMRI-A measurements. These data provide important empirical constraints for theories of object recognition, and highlight the importance of parametric manipulations for capturing neural selectivity to any type of stimulus transformation.

Given the evidence for neural sensitivity to object view, how is view invariant object recognition accomplished? One appealing model for view invariant object recognition is that objects are represented by a population code in which single neurons may be selective to a particular view, but the distributed representation across the entire neural population is robust to changes in object view (Perrett et al., 1998).

Does the view-specific approach necessitate a downstream view invariant neuron? One possibility that supports this account suggests that perceptual decisions may be performed by neurons outside visual cortex and these neurons are indeed view invariant. Examples of such view invariant neurons have been found in the hippocampus, perirhinal cortex and prefrontal cortex (Freedman et al., 2001, 2003; Quiroga et al., 2005; Quiroga et al., 2008). Alternatively, operations based on the population code (or a distributed code) across view-tuned neurons may be sufficient for view invariant decisions based on view-sensitive neural representations.

Debates About the Nature of Functional Organization in the Human Ventral Stream

So far we have considered general computational principles that are required by any object recognition system. Nevertheless, it is possible that some object classes or domains require specialized computations. The rest of this chapter examines functional specialization in the ventral stream that may be linked to these putative “domain-specific” computations.

As illustrated in Fig. 1 several regions in the ventral stream exhibit higher responses to particular object categories such as places, faces and body parts compared to other object categories. Findings of category selective regions initiated a fierce debate about the principles of functional organization in the ventral stream. Are there regions in the cortex that are specialized for any object category? Is there something special about computations relevant to specific categories that generate specialized cortical regions for these computations? That is, perhaps some general processing is applied to all objects, but some computations may be specific to certain domains and may require additional brain resources. A related question is how do these category selective regions come about? That is, are they innate or do they require experience to develop?

In explaining the pattern of functional selectivity in the ventral stream, four prominent views have emerged. The main debate centers on the question of whether regions that elicit maximal response for a category should be treated as a module for the representation of that category, or whether they are part of a more general object recognition system.

Limited category-specific modules and a general area for all other objects

Kanwisher and co-workers (Kanwisher, 2000; Op de Beeck et al., 2008) suggested that ventral temporal cortex contains a limited number of modules specialized for the recognition of special object categories such as faces (in the FFA), places (in the PPA), and body parts (in the EBA and FBA). The remaining object-selective cortex (LOC), which shows little selectivity for particular object categories, is a general-purpose mechanism for perceiving any kind of visually presented object or shape. The underlying hypothesis is that there are few “domain-specific modules” that perform computations that are specific to these classes of stimuli beyond what would be required from a general object recognition system. For example, faces, like other objects need to be recognized across variations in their appearance (a domain-general process). However, given the importance of face processing for social interactions, there are aspects of face processing that are unique. Specialized face processing may include identifying faces at the individual level (e.g. John vs. Harry), extracting gender information, gaze, expression etc. These unique face-related computations may be implemented in face-selective regions.

Process maps

Tarr and Gauthier proposed that object representations are clustered according to the type of processing that is required, rather than according to their visual attributes (Tarr and Gauthier, 2000). It is possible that different levels of processing may require dedicated computations that are performed in localized cortical regions. For example, faces are usually recognized at the individual level (e.g. “That is Bob Jacobs”), but many objects are typically recognized at the category level (e.g. “That is a horse”). Following this reasoning, and evidence that objects of expertise activate the FFA more than other objects

(Gauthier et al., 1999; Gauthier et al., 2000) Gauthier, Tarr and their colleagues have suggested that the FFA is not a region for face recognition, but rather a region for subordinate identification of any object category that is automated by expertise (Gauthier et al., 1999; Gauthier et al., 2000; Tarr and Gauthier, 2000).

Distributed object-form topography

Haxby *et al.* (2001) posited an ‘object form topography’ in which occipito-temporal cortex contains a topographically organized representation of shape attributes. The representation of an object is reflected by a distinct pattern of response across all ventral cortex, and this distributed activation produces the visual perception. Haxby *et al.* showed that the activation patterns for eight object categories were replicable, and that the response to a given category could be determined by the distributed pattern of activation across all ventral-temporal cortex. Further, they showed that it is possible to predict what object category subjects viewed even when regions that show maximal activation to a particular category (e.g., the FFA) were excluded (Haxby et al., 2001). Thus, this model suggests that ventral temporal cortex represents object category information in an overlapping and distributed fashion.

One of the reasons that this view is appealing is that a distributed code is a combinatorial code that allows representation of an large number of object categories. Given Biederman’s rough estimate that humans can recognize about 30,000 categories (Biederman, 1987), this provides a neural substrate that has a capacity to represent such a large number of categories. Second, this model posited a provocative view that when considering information in the ventral stream, one needs to consider the weak signals as much as the strong signals, as both convey useful information.

Topographic representation

Malach et al. (2002) suggested that eccentricity biases underlie the organization of ventral and dorsal stream object-selective regions because they found a correlation between category preference (higher response to one category over others) and eccentricity bias (higher response to a specific eccentricity than other eccentricities, (Levy et al., 2001; Hasson et al., 2002; Hasson et al., 2003)). Regions that prefer houses to other objects, also respond more strongly to peripheral than foveal stimulation. In contrast, regions that prefer faces or letters respond more strongly to foveal than peripheral stimulation. Malach and colleagues (Malach et al., 2002) proposed that the correlation between category selectivity and eccentricity bias is driven by spatial resolution needs. Thus, objects whose recognition depends on analysis of fine details are associated with foveal representations, and objects whose recognition requires large-scale integration are associated with peripheral representations. However, to date there is no clear evidence that eccentricity biases in the FFA are also coupled with better representation of high spatial frequency or smaller receptive fields (or conversely, that the PPA prefers low spatial frequencies or contains neurons with larger receptive fields).

Presently, there is no consensus in the field about which account best explains ventral stream functional organization. Much of the debate centers on the degree to which object processing is constrained to discrete modules or involves distributed computations across large stretches of the ventral stream (Op de Beeck et al., 2008). The debate is both about the spatial scale on which computations for object recognition occur and about the fundamental principles that underlie specialization in the ventral stream.

On the one hand, domain-specific theories need to address findings of multiple

foci that show selectivity. For example, there are multiple foci in the ventral stream that respond more strongly to faces versus objects. Thus, a strong modular account of a single “face module” for face recognition is unlikely. Second, the spatial extent of these putative modules is undetermined and it is unclear whether each of these category-selective regions corresponds to a visual area. Further, high-resolution fMRI, (1-2mm on a side) shows that the spatial extent of category selective regions is lower than that estimated with standard fMRI (3-4mm on a side) and that these regions appear more patchy (Schwarzlose et al., 2005; Grill-Spector et al., 2006b).

On the other hand, a very distributed and overlapping account of object representation in the ventral stream suffers from the potential problem that in order to resolve category information the brain may need to read out information present across the entire ventral stream (which is inefficient). Further, the fact that there is information in the distributed response does not mean that the brain uses the information in the same way that an independent classifier does. It is possible that activation in localized regions is more informative for perceptual decisions than the information available across the entire ventral stream (Grill-Spector et al., 2004; Williams et al., 2007). For example, FFA responses predict when subjects recognize faces and birds, but do not predict when subjects recognize houses, guitars or flowers (Grill-Spector et al., 2004).

Differential Development of Category Selectivity from Childhood to Adulthood

One research direction that can shed light on these debates is an examination of the development of ventral stream functional organization. What is the role of experience in shaping category selectivity in the ventral stream?

fMRI measurements of the development of the ventral stream

Development of functional selectivity within the ventral stream may manifest in several ways. First, the spatial extent of functionally selective regions may change with development. Spatial changes may manifest as reductions in the extent of activations (e.g. broader extent of face-selective activations in children than adults, due to a spatially spread-out and nondifferentiated network in children, see (Berl et al., 2006)) Alternatively, spatial changes may involve an increase in the size of functionally selective regions (e.g. the size of face-selective regions may be larger in adults than children) which may be associated with the development of functional specialization (Gauthier and Nelson, 2001; Johnson, 2001; Nelson, 2001). Second, development may change the magnitude of response. Third, development may be associated with changes in connectivity and/or interaction between regions as suggested by the interactive specialization model (Johnson, 2001). These developmental changes are not mutually exclusive as development may be associated with both increased cortical specialization and changes in connectivity.

To address these questions our lab (Golarai et al. 2007) identified face, place and object selective regions within individual children (7-11 year olds), adolescents (12-14 year olds) and adult subjects (18-35 year olds) while subjects fixated and reported infrequent events when two consecutive images were identical (1-back task). We found a prolonged development of the right FFA (rFFA) and left PPA (lPPA) that manifested as an expansion of the spatial extent of these regions across development from age 7 to adulthood (Fig. 8). The rFFA and lPPA were significantly larger in adults than in children, with an intermediate size of these regions in adolescents. Notably, children's

rFFA was about a third of the adult size, but still evident in 85% of children. These developmental changes could not be explained by smaller anatomical cortical volumes of the fusiform gyrus or parahippocampal gyrus (Golarai et al., 2007), which were similar across children and adults, or higher BOLD-related confounds in children, i.e. larger BOLD-related noise or larger subject motion, see (Golarai et al., 2007; Grill-Spector et al., 2008) because results remained the same for a subset of subjects that were matched for BOLD-related confounds across ages. These developmental changes were specific to the rFFA and IPPA as we found no differences across ages in the size of the LOC or the size of the pSTS face-selective region. Finally, within the functionally defined FFA, PPA and LOC, there were no differences in the level of response amplitudes to faces, objects and places across ages.

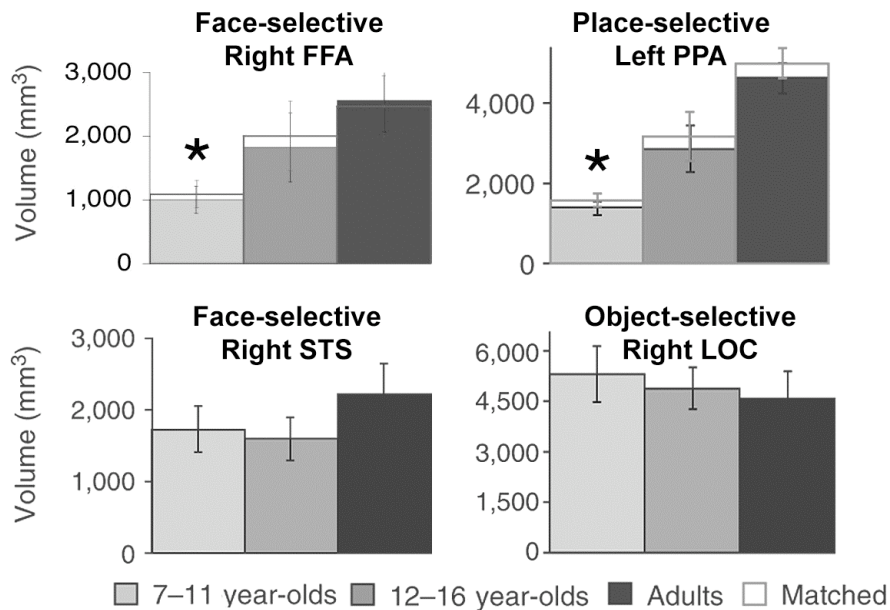


Figure 8: Volume of the rFFA, IPPA, STS and LOC across children, adolescents and adults. Filled bars indicate average volume across all subjects, which include 20 children, 10 adolescents and 15 adults. Open bars indicate the average volumes for the subset of subjects that were matched for BOLD-related confounds and include 10 children, 9 adolescents and 13 adults. Error bars indicate SEM across subjects. Asterisks indicate significantly different than adult, $P < 0.05$. Adapted from Golarai et al. 2007. Note that different panels have different scales on the y-axis.

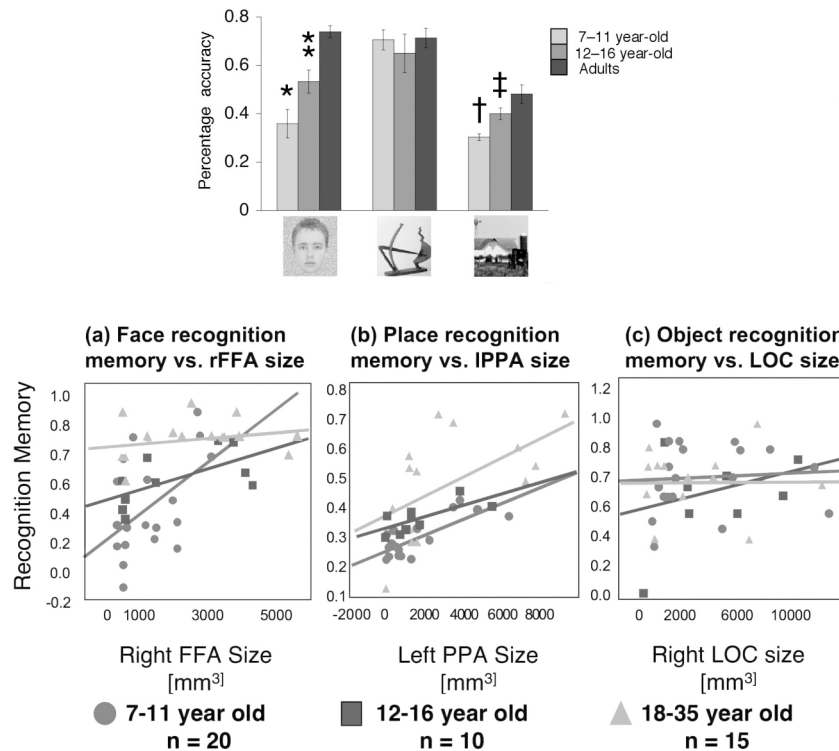


Figure 9: Recognition memory vs. size of FFA, PPA, and LOC. (a) Recognition memory accuracy for different categories across age groups. Recognition memory for faces was significantly better in adults than children ($*P < 0.0001$) or adolescents ($**P < 0.03$). Adolescents' memory for faces was better than children's ($**P < 0.03$). Recognition memory for places was better in adults than in children ($†P < 0.0001$). Adolescents' memory for places was better than children's ($‡P < 0.01$). Recognition accuracy for objects was not different across age groups. Error bars indicate SEM. (b) Recognition memory for faces vs. FFA size; Correlations are significant within children and adolescents ($r > 0.49$, $P < 0.03$), but not adults. (c) Recognition memory of places vs. PPA size. Correlations are significant within each age group ($r > 0.59$, $P < 0.03$). (d) Recognition memory for objects vs. LOC size. No correlations were significant ($P > 0.4$). Adapted from Golarai et al. 2007.

We also measured recognition memory outside the scanner and found that that face and place recognition memory increased from childhood to adulthood (Golarai et al. 2007). Further, face recognition memory was significantly correlated with rFFA size (but not the size of other regions) in children and adolescents (but not adults) and place-recognition memory was significantly correlated with IPPA size (but not the size of other regions) in each of the age groups (Fig. 9). These data suggest that improvements in face

and place recognition memory during childhood and adolescence are correlated with increases in the size of the rFFA and IPPA, respectively.

In another recent study (Scherf et al., 2007) examined the development of the ventral stream in children (5-8 year olds), adolescents (11-14 year olds) and adults using movie clips containing faces, objects, buildings and navigation. Using group analysis methods they reported the absence of face-selective activations (vs. objects, buildings and navigation) in 5-8 year olds in both the fusiform gyrus and pSTS. The lack of FFA in young children in the group analysis may be due to smaller and more variable FFA location in children that would affect its detection in a group analysis. Indeed when Scherf and colleagues performed individual subject analysis they found face-selective activations in 80% of their child subjects, but the extent of activations were smaller and more variable in their location compared to adults. Similar to Golarai et al. they found no difference in the spatial extent or level of response amplitudes to objects in the LOC. However, different from Golarai et al, they reported no developmental changes in the PPA. The variant results may be due to differences in stimuli (pictures vs. movies), task (one-back task vs. passive viewing) and analysis methods (single subject vs. group analysis) across the two studies. For example, Golarai and colleagues instructed subjects to fixate and perform a 1-back task. In this task children had the same accuracy as adults, but were overall slower in their responses, with no differences across categories. In the Scherf study, subjects watched movies, and there was no measurement of performance during scan. It is possible, that in their study children had differential different eye movements or attention compared to adults and this affected their findings.

Implications of Differential Development of Visual Cortex

Overall, fMRI findings suggest differential developmental trajectories across the human ventral visual stream. Surprisingly, these data suggest that more than a decade is necessary for the development of an adult-like rFFA and IPPA. This suggests that experience over a prolonged time may be necessary for the normal development of these regions. This result is surprising especially given that there is evidence for preferential viewing of face-like in newborn babies and evidence for face-selective ERPs within the first 6-12 months (for a review see (Johnson, 2001). One possibility suggested by Johnson and colleagues is that face processing has an innate component that may be implemented in subcortical pathways (e.g. through the superior colliculus) which biases newborns to look at faces. However, cortical processing of faces may require extensive experience (Gauthier and Nelson, 2001; Nelson, 2001) and may develop later.

The reasons for differential development across ventral stream regions are unknown. Importantly, it is difficult to disentangle maturational components (genetically programmed developmental changes) from experience-related components as both are likely to play a role during development. One possibility is that the type of representations and computations in the rFFA and IPPA may require more time and experience to mature than those in the LOC. Second, different cortical regions may mature at different rates due to genetic factors. Third, the FFA may retain more plasticity (even in adulthood) than LOC as suggested by studies that show that FFA responses are modulated by expertise (Gauthier et al., 1999; Tarr and Gauthier, 2000). Fourth, the neural mechanisms underlying experience-dependent changes may differ among LOC, FFA, and PPA.

Open Questions and Future Directions

In sum, neuroimaging research in the past decade has advanced our understanding of object representations in the human brain. These studies have identified the functional organization of the human ventral stream, showed the involvement of ventral stream regions in object recognition, and have laid fundamental stepping stones in understanding the neural mechanisms underlying invariant object recognition.

However, many questions remain. First, what is the relationship between neural sensitivity to object transformations and behavioral sensitivity to object transformations? Do biases in neural representations produce biases in performance? For example, empirical evidence shows over-representation of the lower visual field in LO. Does this lead to better recognition in the lower than upper visual field? A second open question is related to the development of the ventral stream: To what extent is experience (vs. genes) necessary for shaping functional selectivity in the ventral stream? Third, do object representations remain plastic in adulthood? What is the temporal scale of plasticity and are experience-induced changes long lasting? Fourth, what computations are implemented in distinct cortical regions involved in object recognition? Does the "aha" moment in recognition involve a specific response in a particular brain region or does it involved a distributed response across a large cortical expanse? Combining experimental methods such as fMRI and MEG will provide both high spatial and temporal resolution, which is critical to addressing this question. Fifth, what is the pattern of connectivity between ventral stream visual regions? While the connectivity in monkey visual cortex has been extensively explored (Van Essen et al., 1990; Moeller et al., 2008) there is little knowledge about connectivity between cortical visual areas in the human ventral stream.

This knowledge is necessary for building a model of hierarchical processing in humans and any neural network model of object recognition. Future directions that combine methodologies, such as psychophysics with fMRI, MEG with fMRI, or DTI with fMRI will be instrumental in addressing these fundamental questions.

REFERENCES

- Andresen DR, Vinberg J, Grill-Spector K (In press) The representation of object viewpoint in the human visual cortex. *Neuroimage*.
- Appelbaum LG, Wade AR, Vildavski VY, Pettet MW, Norcia AM (2006) Cue-invariant networks for figure and background processing in human visual cortex. *J Neurosci* 26:11695-11708.
- Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001) Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29:529-535.
- Berl MM, Vaidya CJ, Gaillard WD (2006) Functional imaging of developmental and adaptive changes in neurocognition. *Neuroimage* 30:679-691.
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115-147.
- Biederman I, Cooper EE (1991) Evidence for complete translational and reflectional invariance in visual object priming. *Perception* 20:585-593.
- Booth MC, Rolls ET (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* 8:510-523.
- Bulthoff HH, Edelman S (1992) Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci U S A* 89:60-64.
- Bulthoff HH, Edelman SY, Tarr MJ (1995) How are three-dimensional objects represented in the brain? *Cereb Cortex* 5:247-260.
- Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913-919.
- Cohen L, Dehaene S, Naccache L, Lehericy S, Dehaene-Lambertz G, Henaff MA, Michel F (2000) 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 (Pt 2):291-307.
- Culham JC, Danckert SL, DeSouza JF, Gati JS, Menon RS, Goodale MA (2003) Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp Brain Res* 153:180-189.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051-2062.
- DiCarlo JJ, Maunsell JH (2003) Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89:3264-3278.
- DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. *Trends Cogn Sci* 11:333-341.
- Dill M, Edelman S (2001) Imperfect invariance to object translation in the discrimination of complex shapes. *Perception* 30:707-724.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470-2473.
- Edelman S, Bulthoff HH (1992) Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Res* 32:2385-2400.
- Edelman S, Intrator N (2000) (Coarse coding of shape fragments) + (retinotopy) approximately = representation of structure. *Spat Vis* 13:255-264.

- Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G (2008) fMRI activity patterns in human LOC carry information about object exemplars within category. *J Cogn Neurosci* 20:356-370.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598-601.
- Epstein RA, Parker WE, Feiler AM (2008) Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *J Neurophysiol*.
- Fang F, He S (2005) Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nat Neurosci* 8:1380-1385.
- Fang F, Murray SO, He S (2007) Duration-dependent fMRI adaptation and distributed viewer-centered face representation in human visual cortex. *Cereb Cortex* 17:1402-1411.
- Farah MJ (1995) *Visual Agnosia*: MIT press.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291:312-316.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci* 23:5235-5246.
- Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343-346.
- Gauthier I, Nelson CA (2001) The development of face expertise. *Curr Opin Neurobiol* 11:219-224.
- Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 3:191-197.
- Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci* 2:568-573.
- Gauthier I, Hayward WG, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (2002) BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron* 34:161-171.
- Gilaie-Dotan S, Ullman S, Kushnir T, Malach R (2002) Shape-selective stereo processing in human object-related visual areas. *Hum Brain Mapp* 15:67-79.
- Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JD, Grill-Spector K (2007) Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat Neurosci* 10:512-522.
- Grill-Spector K (2003) The neural basis of object perception. *Curr Opin Neurobiol* 13:159-166.
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107:293-321.
- Grill-Spector K, Malach R (2004) The human visual cortex. *Annu Rev Neurosci* 27:649-677.
- Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci* 7:555-562.
- Grill-Spector K, Henson R, Martin A (2006a) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14-23.

- Grill-Spector K, Sayres R, Ress D (2006b) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat Neurosci* 9:1177-1185.
- Grill-Spector K, Golarai G, Gabrieli J (2008) Developmental neuroimaging of the human ventral visual cortex. *Trends Cogn Sci* 12:152-162.
- Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3:837-843.
- Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R (1998a) Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron* 21:191-202.
- Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, Malach R (1998b) A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* 6:316-328.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187-203.
- Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37:1027-1041.
- Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34:479-490.
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223-233.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425-2430.
- Hemond CC, Kanwisher NG, Op de Beeck HP (2007) A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS ONE* 2:e574.
- Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci* 3:80-84.
- James TW, Humphrey GK, Gati JS, Menon RS, Goodale MA (2002) Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron* 35:793-801.
- Johnson MH (2001) Functional brain development in humans. *Nat Rev Neurosci* 2:475-483.
- Kanwisher N (2000) Domain specificity in face perception. *Nat Neurosci* 3:759-763.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302-4311.
- Kastner S, De Weerd P, Ungerleider LG (2000) Texture segregation in the human visual cortex: A functional MRI study. *J Neurophysiol* 83:2453-2457.
- Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856-867.
- Kourtzi Z, Kanwisher N (2000) Cortical regions involved in perceiving object shape. *J Neurosci* 20:3310-3318.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506-1509.

- Kourtzi Z, Tolias AS, Altmann CF, Augath M, Logothetis NK (2003) Integration of local features into global shapes: monkey and human fMRI studies. *Neuron* 37:333-346.
- Lerner Y, Hendler T, Malach R (2002) Object-completion effects in the human lateral occipital complex. *Cereb Cortex* 12:163-177.
- Lerner Y, Epshtein B, Ullman S, Malach R (2008) Class information predicts activation by object fragments in human object areas. *J Cogn Neurosci* 20:1189-1206.
- Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R (2001) A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex* 11:287-297.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4:533-539.
- Logothetis NK, Pauls J, Poggio T (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5:552-563.
- Lueschow A, Miller EK, Desimone R (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 4:523-531.
- Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6:176-184.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135-8139.
- Marr D (1980) Visual information processing: the structure and creation of visual representations. *Philos Trans R Soc Lond B Biol Sci* 290:199-218.
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996) Neural correlates of category-specific knowledge. *Nature* 379:649-652.
- McKyton A, Zohary E (2007) Beyond retinotopic mapping: the spatial representation of objects in the human lateral occipital complex. *Cereb Cortex* 17:1164-1172.
- Mendola JD, Dale AM, Fischl B, Liu AK, Tootell RB (1999) The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *J Neurosci* 19:8560-8572.
- Miller EK, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254:1377-1379.
- Moeller S, Freiwald WA, Tsao DY (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320:1355-1359.
- Nakayama K, He ZJ, Shimojo S (1995) Visual surface representation: A critical link between low-level and high-level vision. In: *An invitation to cognitive sciences: visual cognition* (Kosslyn SM, Osherson DN, eds). Cambridge: MIT Press.
- Nelson CA (2001) The development and neural bases of face recognition. *Infant and Child Development* 10:3-18.
- Op De Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426:505-518.
- Op de Beeck HP, Haushofer J, Kanwisher NG (2008) Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci* 9:123-135.
- Perrett DI (1996) View-dependent coding in the ventral stream and its consequence for recognition. In: *Vision and movement mechanisms in the cerebral cortex* (Camaniti R, Hoffmann KP, Lacquaniti AJ, eds), pp 142-151. Strasbourg: HFSP.

- Perrett DI, Oram MW, Ashbridge E (1998) Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* 67:111-145.
- Peterson MA, Gibson BS (1994a) Must shape recognition follow figure-ground organization? An assumption in peril. *Psychol Sci* 5:253-259.
- Peterson MA, Gibson BS (1994b) Object recognition contributions to figure-ground organization: operations on outlines and subjective contours. *Percept Psychophys* 56:551-564.
- Poggio T, Edelman S (1990) A network that learns to recognize three-dimensional objects. *Nature* 343:263-266.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102-1107.
- Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I (2008) Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci U S A* 105:3599-3604.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019-1025.
- Rolls ET (2000) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron* 27:205-218.
- Rolls ET, Milward T (2000) A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Comput* 12:2547-2572.
- Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fMRI adaptation paradigm. *Neuron* 49:307-318.
- Sayres R, Grill-Spector K (2008) Relating retinotopic and object-selective responses in human lateral occipital cortex. *J Neurophysiol* In press.
- Scherf KS, Behrmann M, Humphreys K, Luna B (2007) Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev Sci* 10:F15-30.
- Schwarzlose RF, Baker CI, Kanwisher NK (2005) Separate Face and Body Selectivity on the Fusiform Gyrus. *J Neurosci* 25:11055-11059.
- Schwarzlose RF, Swisher JD, Dang S, Kanwisher N (2008) The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci U S A* 105:4447-4452.
- Stanley DA, Rubin N (2003) fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron* 37:323-331.
- Tarr MJ, Bulthoff HH (1995) Is human object recognition better described by geometric structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *J Exp Psychol Hum Percept Perform* 21:1494-1505.
- Tarr MJ, Gauthier I (2000) FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci* 3:764-769.
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520-522.
- Thorpe SJ, Gegenfurtner KR, Fabre-Thorpe M, Bulthoff HH (2001) Detection of animals in natural images using far peripheral vision. *Eur J Neurosci* 14:869-876.

- Ullman S (1989) Aligning pictorial descriptions: an approach to object recognition. *Cognition* 32:193-254.
- Ungerleider LG, Mishkin M, Macko KA (1983) Object vision and spatial vision: Two cortical pathways. *Trends Neurosci* 6: 414-417.
- Valyear KF, Culham JC, Sharif N, Westwood D, Goodale MA (2006) A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. *Neuropsychologia* 44:218-228.
- Van Essen DC, Felleman DJ, DeYoe EA, Olavarria J, Knierim J (1990) Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey. *Cold Spring Harb Symp Quant Biol* 55:679-696.
- Vinberg J, Grill-Spector K (2008) Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J Neurophysiol* 99:1380-1393.
- Vogels R, Biederman I (2002) Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex. *Cereb Cortex* 12:756-766.
- Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci* 5:491-499.
- Wandell BA (1999) Computational neuroimaging of human visual cortex. *Annu Rev Neurosci* 22:145-173.
- Wang G, Tanaka K, Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272:1665-1668.
- Williams MA, Dang S, Kanwisher NG (2007) Only some spatial patterns of fMRI response are read out in task performance. *Nat Neurosci* 10:685-686.

ACKNOWLEDGEMENTS

I thank David Andresen, Golijeh Golarai, Rory Sayres and Joakim Vinberg for their contributions to the research summarized in this chapter. David Andresen, for his contributions to understanding viewpoint-sensitive representations in the ventral stream. Golijeh Golarai, for spearheading the research on the development of the ventral study. Rory Sayres, for his contributions to research on position invariance, and distributed analyses of the ventral stream. Joakim, for his dedication to researching cue-invariant and figure-ground processing in the ventral stream. I thank David Remus, Kevin Weiner, Nathan Witthoft and two anonymous reviewers for comments on this manuscript. This work was supported by National Eye Institute Grants 5 R21 EY-016199-02 and and Whitehall Foundation Grant 2005-05-111-RES.