

## What is that?

### INTRODUCTION

Nearly half of the primate cortex is dedicated to visual processing (Felleman and Van Essen, 1991). This demonstrates how reliant animals, especially primates, are on the sense of sight for daily life. Think about the way an animal goes about finding food, guiding interactions with other living creatures or selecting the tools that are needed to solve an immediate problem. These tasks are primarily conducted using information that is visual in nature.

One important function of vision is to parse the important features of a visual stimulus which can be used as a guide for behavior (i.e. not eating the bright red, probably poisonous, ladybug or choosing a pointed object in order to make an incision). One level of abstraction beyond identifying properties of a visual object, primates have the ability to synthesize these features to recognize a conceptual meaning or utility of the object in question. Not unlike recognition of lower-level visual percepts, object recognition is a fundamental function of the human brain and lacking it would impede evolutionary fitness. Humans can distinguish between tens of thousands of possible objects and we do so in fractions of a second (DiCarlo et al., 2012) and do so independent of particular transformations that do not alter the conceptual meaning of the object (Isik et al., 2014).

As made evident by the work being done in computer vision, object recognition is a very computationally demanding task. Machine algorithms have been shown to outperform humans when the problem space is rather restricted, like in number or fingerprint recognition or even distinguishing between cars and not cars (Borji and Itti, 2014). When given the task of core object recognition or scene analysis, the models do not recreate behaviors typical of humans given the same task. Research being done from a wide range of fields have provided evidence elucidating the possible mechanism for quick and reliable classification of visual objects.

In broad strokes, the cortical ventral stream has been implicated as the anatomical substrate of object recognition. The neural system supporting this function has been characterized as a hierarchical neurobiological network of increasingly complex representations as you progress through the ventral stream.

Moving forward, our operational definition of object recognition is as follows: the ability to perceive an object's physical properties (such as shape, color and texture) and apply semantic attributes to the object, which includes the understanding of its use, previous experience with the object and how it relates to others (Enns, 2004). This review is organized in a way that mimics the visual pathways under consideration.

Here, I review the steps in the visual processing hierarchy that are relevant to the computational task of object recognition. We begin where the visual process begins, in connections between the retina and the primary visual cortex. Next, I review the anatomical underpinnings of the cortical visual stream and then expand upon the functional roles that each cortical area plays. Next, I examine how these networks may solve the issue of invariance and how meaning results from isolated features. Last, I consider the recurrence that is evident in anatomical studies that has, until recently, been ignored by computational models.

Defining the mechanisms of object recognition in the mammalian nervous system would be invaluable to the development of computer vision technologies. Potential applications include implementing "vision" sensors in autonomous robots, detecting presence of particular object for

visual surveillance and categorizing medical imaging results among many others. The economic value of these applications are a major motivator for research in the neural mechanisms of object recognition.

## **BEFORE THE CORTEX**

Between the retina and the cortex, the visual pathway goes through a series of steps. An important property of these early steps is the segregation of different kinds of information. This differentiation occurs initially at the retina. One level of distinction occurs because of the two types of photoreceptors. Rods are more sensitive to photons. Cones are sensitive to color and grant a higher degree of visual acuity. Also, there is an array of retinal ganglion cells that have a maximal response to particular aspects of a stimulus. For instance, P cells are sensitive to changes in contrast and M cells are more sensitive to movement.

Retinal ganglion cells then course through the optic nerve and tract to terminate in the lateral geniculate nucleus (LGN) of the thalamus. Projections from the different types of retinal ganglion cells terminate in different layers of the nuclei structure. The layers' response properties follow that of their input cells. Neurons in the LGN are monocular and have concentric or center-surround receptive fields (Callaway, 2005). Neurons in the optic tract terminate on more structures than the LGN, but those pathways are involved in other visual functions other than perception and do not fall in the scope of this paper.

Axons of neurons in the LGN terminate in different layers of the primary visual cortex. The M-cell pathway ends in layer  $IVC\alpha$  and the P-cell pathway ends in layer  $IVC\beta$ . Further, the koniocellular neurons that carry information about color terminate in the V1 blobs (Callaway, 2005).

The segregation of different kinds of information in the early visual pathway is important to consider in object recognition. Not only are the kinds of information separated, but the parallel tracks hold for much smaller units, such as by neuron. This maintains the retinotopy observed along this entire pathway. Also, very little processing of the visual signal has occurred by this step in the chain. It is easy to treat the early visual pathway as a feature extraction mechanism. The hierarchical model of categorization relies on synthesizing this low-level, "feature" information to represent the next-most complex level of visual perception.

At the level of the primary visual cortex, there is a retinotopic mapping of the current visual world of an individual, with separate tracks for shape, movement and color. Later stages in the overall visual pathway, housed in the ventral stream between V1 and the apex of the inferior temporal lobe, are responsible for the hierarchical processing of the (so-far) parallel features. The ventral stream mainly receives input from the parvocellular or P-cell pathway and the koniocellular pathway. Information from the M channel is also represented, but to a lesser degree.

## **ANATOMICAL UNDERPINNINGS**

Lesion studies and temporary inactivation/activation studies point to the ventral visual stream as the localization of object recognition. The ventral pathway starts in the primary visual cortex and courses through the occipitotemporal cortex (including visual area two, V2, and four, V4) to the anterior part of the inferior temporal (IT) cortex with a likely extension into the ventrolateral prefrontal cortex and perirhinal cortex among other structures that aid in recognition of visual stimuli (Kravitz et al., 2013). Lower level features are coded early in the

pathway and representations get more complex moving anteriorly from V1 to the apex of the inferior temporal lobe.

A fundamental study in the anatomy of the extended visual system was conducted by Daniel Felleman and David Van Essen in 1991. After pooling data from tens of anatomical studies of cortical visual areas, an undirected-graph representation of the areas involved in vision was mapped. Twenty-five neocortical areas were predominantly visual in function and seven areas associated with vision and other senses. The group found 305 connections among the thirty-two areas considered. Among these connections, about 40% were determined to be reciprocal (Felleman and Van Essen, 1991). Their version of the visual hierarchy contained ten levels of cortical processing. The hierarchy would have fourteen if the retina and lateral geniculate nucleus on the bottom and two memory-associated structures at the top are also included.

Beyond the anatomical connections between areas in the visual system, their analysis revealed some underlying principles of the organization of the ventral visual stream. First, there is a large number of visual areas. This resonates with primate's reliance on visual sensation over other modalities. Second, there is highly distributed connectivity among these areas. As can be observed in Figure 1, there is extensive contact between the articulating areas. Third, they highlighted the reciprocity of connections. This confirmed that the problem of object recognition is not a feed-forward scheme, ignoring higher-level information such as context, attention or expectation. In fact, the 40% reciprocity observed suggests there is a high degree of top-down modulation in visual processing. Fourth, the ventral stream areas are organized in a hierarchical fashion. Anatomically, the hierarchies are much less neat than is often depicted, but the principle still holds. The last principle outlined was the distinct, but intertwined, processing streams. In the visual cortex, each processing stream maintains a distinct profile of receptive field characteristics (Logothetis and Sheinberg, 1996). This extends the property of feature isolation evident in the earlier visual pathway, but the features represented in later visual cortex (V4 and IT) are much more complex due to the computations done in the early visual cortex (Lee, 2003).

Later studies confirmed the functional qualities that these earlier anatomical results implied. Intrinsic signal imaging of the IT cortex revealed that visually presented objects activated patches in a distributed manner. Stripping the same objects of a few of its visual features and re-presenting them to the monkey under study resulted in the activation of a subset of the patches excited by the original stimulus. See Figure 2 for differential activation of certain regions by objects and their simplified counterparts. The group concluded that an object is represented by a combination of cortical columns in the IT, each of which responsible for a particular feature (Tsunoda et al., 2001). Both the activation and inactivation are important in coding the complex representation of the object. These results give a physiological confirmation of the hierarchical structure suggested by Felleman and Van Essen. The demonstration of "feature" columns corresponds to the multiple processing streams. The lack of multiple processing streams would exhibit object-specific responses of neurons in the late stages of the ventral, rather than feature-responses made evident in the intrinsic signal imaging study. The cross-talk displayed anatomically between these individual streams is suggested.

A more understandable metaphor for the distinct, yet intertwined, streams would be various groupings found in civilized societies. In general, there is some form of hierarchical organization of authority, with different ranks having different roles. In an average system, the pecking order is not precisely defined, may be fluid depending on context and individuals may play a role in multiple hierarchies while serving a similar function in both. Important properties

of this complex system are: bidirectional information flow, the ability to bypass intermediate levels and parallelism of functionally distinct channels.

## **HIERARCHICAL PROCESSING THROUGH THE VISUAL PATHWAY**

### **Early stages of visual computation**

Hubel and Weisel's discovery of simple, complex and hypercomplex cells in the visual cortex was some of the first evidence toward the hierarchical nature of visual processing. These cell types were found in cats (Hubel and Wiesel, 1962) and later in two kinds of monkeys (Hubel and Wiesel, 1968).

Simple cells were at the lowest level of the hierarchy and are sensitive to the orientation of a visual stimulus. The convergence of multiple, aligned, center-surround receptive fields grants the simple cell with its characteristic receptive field. Simple cells are found mainly in layer IV of the primary visual cortex. The convergence of simple cell input to the complex cells creates the orientation- and direction-selectivity of the complex cells as well as its evident phase independence. Complex cells are found in primary and secondary visual cortices (V1 and V2). See Figure 3 for a graphical representation of response properties as well as the circuitry by which these receptive fields are built.

Hypercomplex cells are one level intricacy beyond the complex cells and they are found in The additional sensitivity they exhibit is length of the stimulus. Essentially, for a stimulus that matches the hypercomplex cells preferred orientation and direction, increasing illumination in a particular region elicits stronger responses but when that illumination extends beyond a critical length, the response becomes progressively weaker (Hubel and Wiesel, 1968). The term used to describe this feature is "end-stopping". These properties can be theoretically modelled as excitatory input from a complex cell within an activation region and inhibitory input from another complex cell in the antagonistic region.

These findings highlight two principles of visual processing that seem to have carried a lot of future object recognition work. The first is the feed-forward hierarchical scheme that increases in complexity of representation from layer to layer. As pointed out in the Anatomical Underpinnings section, the hierarchy is not unidirectional and totally convergent as Hubel and Wiesel's original model suggests. The role of top-down input and recurrence will be addressed later in this review, but the basic organization still stands. The second principle that still holds is that higher complexity representations are localized more anteriorly in the ventral stream.

### **Late stages of visual computation**

By the time that visual information reaches visual area V4 and IT cortex, a great degree of processing has occurred. If the first stages of the ventral stream are responsible for the extraction of local orientation and spatial frequency information, then later stages are responsible for encoding second- and third-order derivatives of orientation and contrast. Neurons in V4 have been observed to respond to sinusoidally-modulated gratings built from the polar and hyperbolic domain (Gallant et al., 1993). Further, neurons in the posterior IT display complex, distributed response patterns conveying information about multidimensional shapes, filtered through tuning functions for contour (Brincat and Connor, 2004). Representations in V4 and IT are able to discriminate shape information by integrating over object parts and their positional and connectional relationships.

Additionally, the representations in late stages of visual computation have increased tolerance or invariance to identity-preserving transformations. These ideas are hard to systematically test and has not been quantitatively examined until recently. By comparing population activity patterns in the V4 and IT, the increase of selectivity and invariance exhibited by successive stages in the ventral processing stream can be verified. Population codes were represented by vectors and used to create divisions in the response state space discriminating between images of a set. First, the group tested conjunction sensitivity in both areas. Conjunction sensitivity can be thought of as the universality of the features encoded by a population of neurons. Conjunction sensitivity was examined by recording population responses to a set of natural images and to scrambled versions of those images. In V4, moderate reductions in discriminability performance were observed for scrambled images in comparison with natural images (Figure 6 b, d). A higher degree of decrement was observed in the IT when comparing responses to scrambled and natural images (Figure 6 b, d) (Rust and DiCarlo, 2010). These results indicate support the hypothesis that neurons at lower levels of the visual hierarchy encode more local structure while neurons in higher levels of the hierarchy are more sensitive to the conjunction of these local features.

In a second test, population responses were recorded with a set of stimuli images designed to examine the V4 and IT cortex tolerance to identity-preserving transformations. This stimulus set had a reference image of a natural object and of that object in different positions, of varying sizes and within different contexts. Responses were used to quantify the ability of the neural population to generalize over the different transformations. Results demonstrated that the ability of IT population responses to encode object identity is more tolerant to size, position and context transformations than population responses of the V4 (Figure 7 c, d). This indicates that neurons in higher levels of the visual processing hierarchy and more anterior in the inferior temporal cortex have a greater generalization capacity than neurons in the lower levels in the hierarchy at more posterior cortical areas (Rust and DiCarlo, 2010).

Beyond increases in selectivity and generalizability as information flows through the visual stream, there is another widely agreed-upon tenet of late visual processing. It is generally accepted that readout or population decoding of IT responses is sufficient to support core object recognition (DiCarlo et al., 2012). Predictive performance of IT readout exceeds that of many of artificial intelligence's attempts at the computer vision problem (Hung et al. 2005).

## **THE INVARIANCE PROBLEM**

In the natural world, each time we encounter an object is entirely unique. Despite this, we are still able to recognize an object. The transformations drastically change the neural response at the level of the retina. But, the image transformations do not alter the core identity of the object. The invariance problem is the crux of why object recognition is computationally difficult. The first implementations of a hierarchical model disregarded these translations and performed very poorly because it would require the encoding of potentially infinite states of different viewing conditions for each object (DiCarlo and Maunsell, 2000).

Knowing the features that are independent in later stages of the visual hierarchy helps clarify how the visual processing hierarchy solves the problem of invariance. In a physiology study, recordings were collected from individual neurons of the inferior temporal cortex of a monkey. Monkeys were, first, presented with a paper-clip-like objects of unvarying size and position, but with a range of rotational aspects. Tracking the activity of a particular neuron to all the rotational aspects and to distractor stimuli shows that the neurons the group was recording

from had view-specific response properties (Logothetis et al., 1995). The neurons exhibited a broad tuning curve of about 60 degrees, centered around the preferred view (Figure 4). Next, monkeys were trained to recognize a particular paper-clip-like object using a single position and size. They were then tested on discrimination between the trained object and similar, non-target, objects using different positions and sizes. The cellular recordings from trials presenting the target object (over all the sizes and positions) were separated to analyze the effect of size and position on the response of these view-tuned units. As shown in parts C and D of Figure 5, responses to the best-tuned view remained relatively constant over a range of sizes and locations.

A 1999 study conducted by the same lab set out to build a biologically feasible recognition model that mimics the behavior of these view-tuned units. The goal was to determine how the computations in the ventral stream integrate the extracted features to determine object identity independent of size and position. As stated before, the “weighted-sum” approach is one way to integrate input that would result in responsiveness to higher-level features. The weighted-sum approach, though, fails when stimuli are scaled or translated. By replacing certain integration steps in the hierarchy with a maximum-picker function, the model mimicked monkey physiological data and achieved invariance (Reisenhuber and Poggio, 1999). Other neurophysiological data of IT neurons show that individual neurons display this kind of integrative technique. When presented with multiple stimuli, the response of the integrator neuron is dominated by the most effective stimulus (when presented individually) (Sato, 1989). Thus, a model including a “choose-max” functionality at some stations of integration is biologically feasible.

## **MAPPING COMPLEX FEATURES TO MEANING**

The last steps in visual processing is localized to the apex of the ventral stream in an area within the medial temporal lobe called the perirhinal cortex (Tyler et al., 2013). It is generally agreed upon that this region performs the highest level feature integrations required to discriminate between two very similar objects. The neurobiological correlates of conceptual analysis (assigning a meaning or identity to the visual stimulus object) are more controversial. Researchers have been particularly interested in the fusiform gyrus and the anteromedial temporal lobe as regions potentially housing the function of semantic categorization.

Functional imaging studies of the fusiform gyrus demonstrate its ability to differentiate between categories. Different portions of the fusiform area preferentially respond to different object categories such as tools and animals (Tyler et al., 2013). Patients with lesions to the anteromedial temporal cortex exhibit category-specific deficits for living things. These patients have difficulty differentiating between highly similar objects, more dramatically for animate objects. Both the fusiform gyrus and anteromedial temporal lobe have some role in assigning an identity to a visually presented object, though they contribute to different aspects of an object’s meaning.

Two key aspects of concept analysis are category and object-specific information. The former captures broader properties of a visual stimulus such as living versus non-living. The latter serves to differentiate between the identities of two similar objects. For instance, differentiating between a puma and a tiger would require more object-specific information. Feature statistics can be used to reflect these aspects. Sharedness quantifies the degree to which features in a concept are shared with other concepts. A high sharedness score denotes an object with more shared features. Objects with relatively higher sharedness value with greater BOLD activity in the bilateral lateral fusiform gyri (Figure 8). Further analysis showed that objects with

lower sharedness values had significant BOLD activity in the medial fusiform gyri (Tyler et al., 2013). This corresponds with living objects represented in the lateral fusiform gyrus and non-living objects in the medial fusiform gyrus. This results loosely supports the hypothesis that broad categorization occurs in the fusiform gyrus.

Correlation x Distinctiveness quantifies the relative correlational strength of shared versus distinctive features within a concept. A low CxD score reflects relatively weak influences of the most distinctive features compared to shared features and would require more complex conceptual integration to enable basic level identification. Results show that presentation of objects with low CxD scores elicited stronger activity in the anteromedial temporal lobe. High CxD scored stimuli did not evoke a patterned response. The group concluded that anteromedial temporal lobe may contribute to distinguishing between objects that are highly similar (Tyler et al., 2013).

An extension of the perceptual hierarchy of the earlier visual processing stages into a conceptual hierarchy based on semantic feature statistics is an appealing and intuitive hypothesis. The results of the Tyler et al., 2013 study do fit into the framework of this hypothesis. In fact, a recent model designed from the largely feedforward architecture of the ventral stream exhibits performance and relative response latencies similar to human in an animal classification task (Serre et al., 2006). The authors commented that their model did not implement the anatomically proven bidirectional information flow. Granting that visual processing is driven largely by feedforward architecture, their results indicate that the hierarchy extends to the behavioral task of assigning meaning to a stimulus. Not only does their model recreate the hit and false-alarm rates of humans, but it also had poor performance on individual images that humans also struggled with.

Where the decisions about category boundaries take place within the cortex remains unclear. The theoretical framework for how these computations solve the semantic recognition problem, though, has been formed.

## **THE ROLE OF RECURRENCE**

Results demonstrating the efficacy and humanity of an entirely feedforward architecture of the visual processing stream rely on unobscured object stimuli. Here, a very small proportion of the information utilized is sourced from feedback connections between or within areas. This leaves the anatomical substrates for feedback connections functionally useless in theories of object recognition. The recurrent networks become necessary in imperfect viewing conditions such as when there limited viewing angles, poor luminosity or occlusion.

In a recent study using intracranial field potential recording in epilepsy patients, the spatiotemporal dynamics of object completion were examined. Object selectivity was retained despite presenting partial information, meaning that neurons that showed strong, selective responses to the whole object would also have a strong response to the partial version of that object. The difference between the responses to the whole and partial condition, though, was in the timing. The responses to partial objects were delayed compared to responses of the corresponding whole objects. Response latencies depended on the sets of features obscured in each trial. On average, the partial object responses required about 100 msec additional processing time as compared to whole object responses (Tang et al., 2014). The authors ascribed the delayed physiological responses to the recurrent computations required to utilize prior knowledge in categorizing or deciding the identity of a visually presented object. This theory

would give the top-down and horizontal projections an important role in the visual processing pathway.

The Leabra Vision computational model developed at University of Colorado, Boulder, has the recurrent architecture and demonstrates the contributions of bidirectional information flow. The group tracked the similarity of activity in each layer of the network compared to the final activation state of each such layer for the completely unblocked trial. This revealed the ability of the network to reconstruct the encoding of an unblocked response when only presented with a fraction of that visual information available. Recurrent networks are able to generate a complete internal representation when inputs are highly ambiguous, but only after considerably more cycles of processing. Without the efferent connections, the model was unable to code the core identity of the occluded object (O'Reilly et al., 2013). Interestingly, the V2/V4 level within the model does not reach equal activity to the reference state but the IT or semantic level does. Further tests of the Leabra Vision model suggest that the inhibitory feedback is the driving force in the completion effect.

Another possible contribution of the recurrent networks in the visual processing hierarchy is to refine the representation of the stimulus at each level of abstraction. The concept of abstraction is very advantageous to human-engineered systems. It simplifies the mechanisms at each level in the hierarchy such that only the details relevant to that stage of processing are available to be accessed. It also allows the overall process to be done in very discrete steps, with a low degree of coordination over the multiple steps. Recurrent networks at each level in the visual hierarchy would be able to specify a role for each layer. Little neural evidence has been collected to confirm or deny the role of recurrence to grant levels of abstraction in the ventral stream. Despite this, vision researchers do entertain its theoretical value (DiCarlo et al., 2012).

## **CONCLUSION, OPEN QUESTIONS AND FUTURE DIRECTIONS**

The current theory of visual object recognition has been outlined in a theoretical way. It is described as a complex system with multiple hierarchical and parallel streams that display a high degree of cross-talk. Earlier stages of the stream are much better understood than later stages. Most studies assume an extension of lower-level principles into the higher levels of the hierarchy. Sufficient physiological, computational and behavioral evidence indicate these overarching principles along the pathway, but we are far from a totally complete characterization of the entire ventral, visual stream.

A major problem in research on object recognition is that there is no widely-accepted operational, systematic definition of “success” for proposed models. A successful definition of success must be able to account for behavioral, physiological and computational properties in all stages of the processing hierarchy. Beyond this, we lack the psychophysical data required to characterize human object recognition. Studies of the visual processing stream will be much more informative and directed once natural object recognition is quantitatively characterized and qualities of success are defined.

The mass of the ventral stream and the great number of computations needed to categorize a simple object poses a first major issue in the theory of object recognition. The intermediate levels of abstraction in the visual pathway are not formally defined. First, the fundamental units of processing at each level have not been completely elucidated. Second, the mechanisms required to synthesize lower-complexity information into higher-level forms have been probed but not confirmed. Potential algorithmic processes of intermediate stages of the



visual hierarchy will need to be tested against mammalian input-output functions in order to clarify the progressive abstractions of object recognition.

In conclusion, the fields of computer vision and neural visual processing have come a long way in describing the conversion of electromagnetic waves into an identified concept. Many open questions still remain. Working theories of object recognition in primates, namely the functional operations of a hierarchical organization, has been implemented into computer vision technologies and have already revolutionized the way we can process signals technologically.

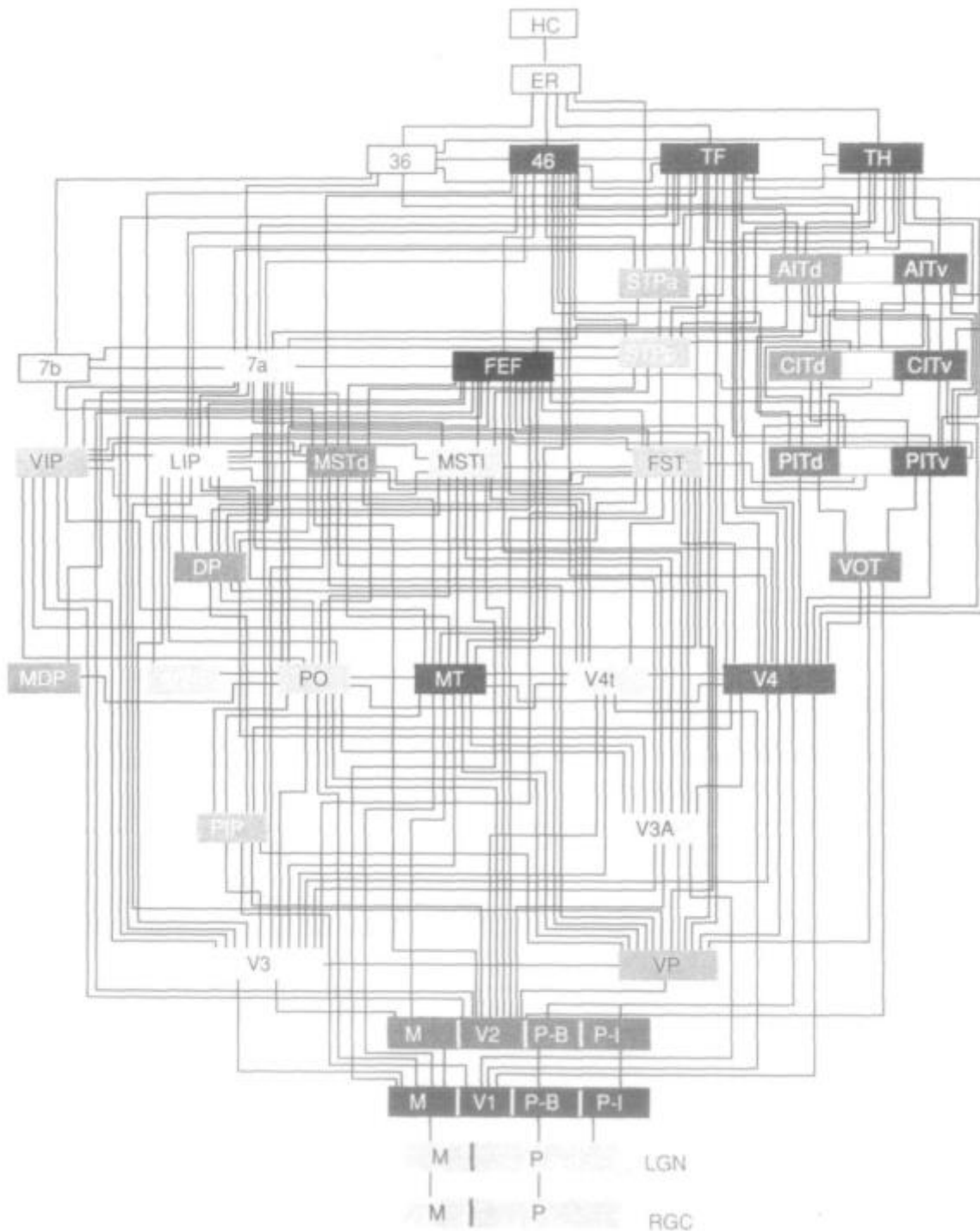


Figure 1  
Felleman and Van Essen's graphical representation of the distributed, hierarchical connections between 32 cortical visual areas.

(Felleman and Van Essen, 1991)

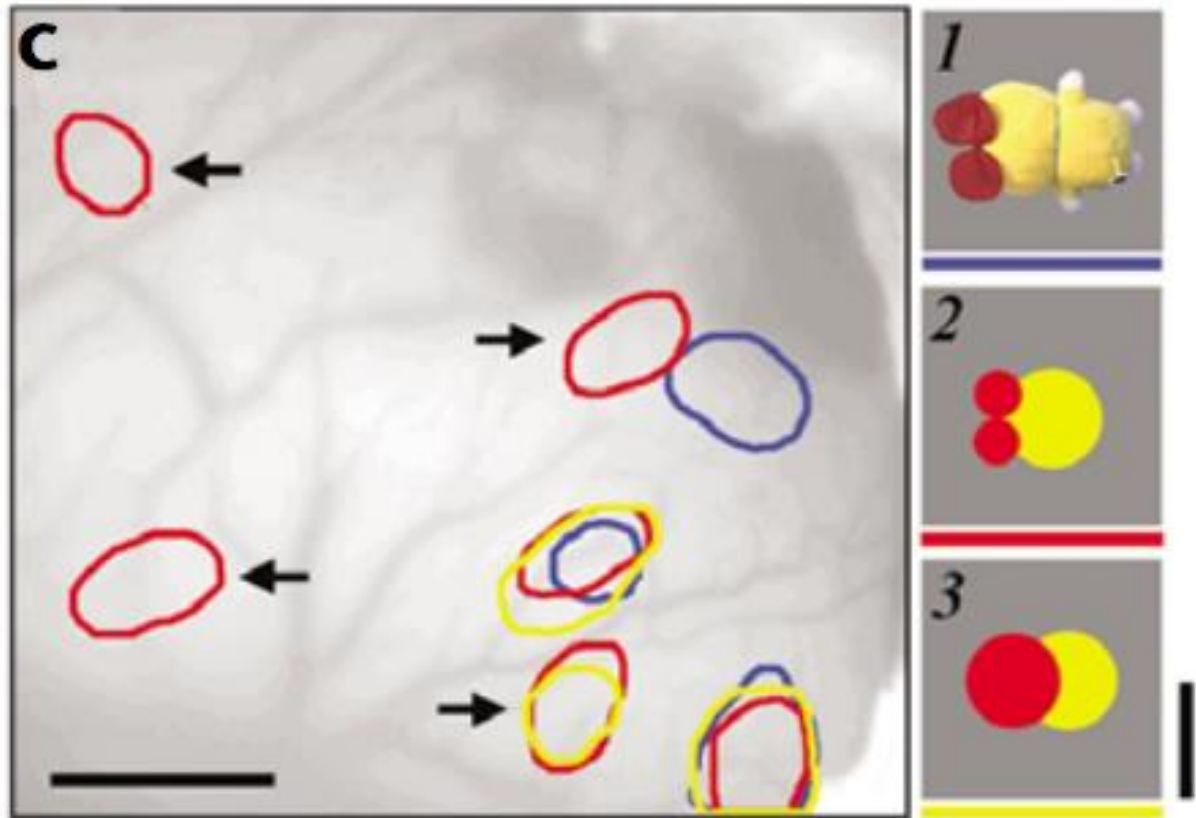


Figure 2  
Representation of complex object images and simplifications of them in the inferior temporal cortex. Outlines form the boundaries of activation with respective stimuli.

(Tsunoda et al., 2001).

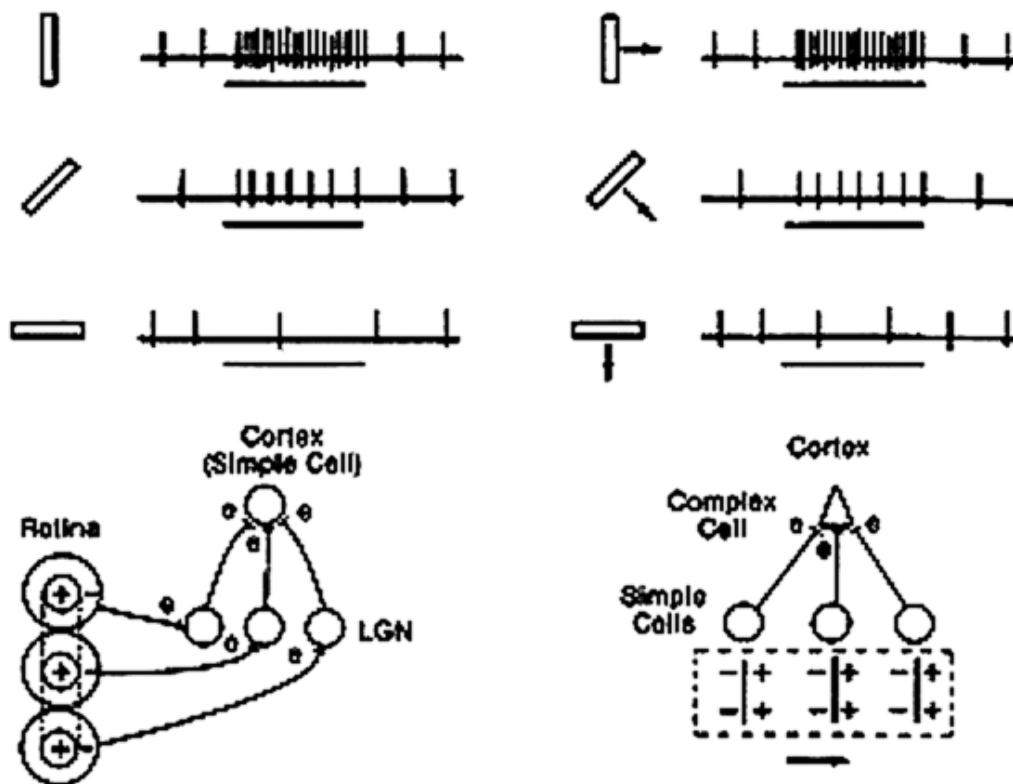


Figure 3  
Graphical representation of response properties and potential circuitry of simple and complex cells as originally proposed by Hubel and Wiesel.

(Hubel and Wiesel, 1968)

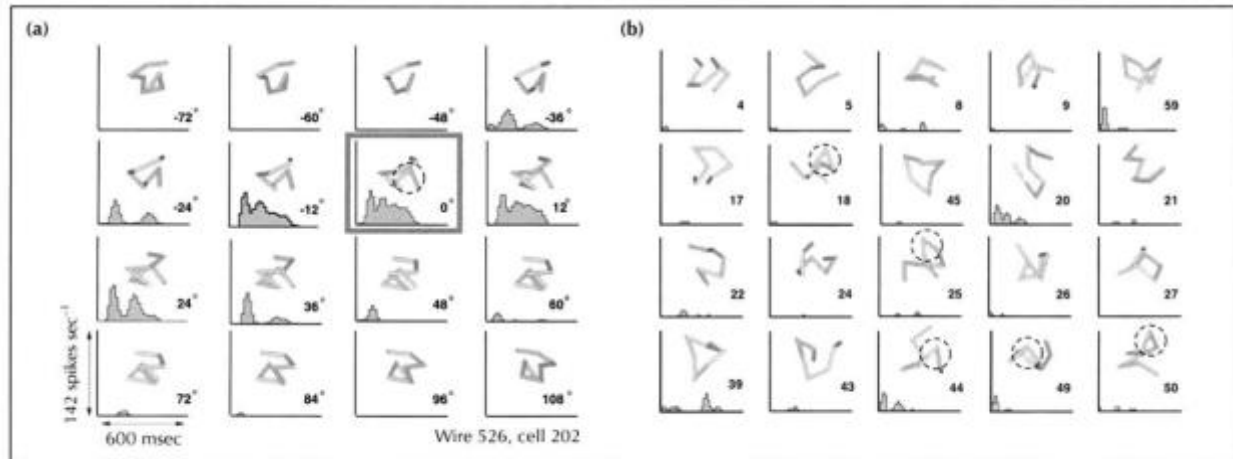


Figure 4  
View-selective response of an IT neuron to a wire-like object. Peristimulus histograms (PSTHs) show the activity of a view-selective neuron when (a) the target or (b) distractors were presented.

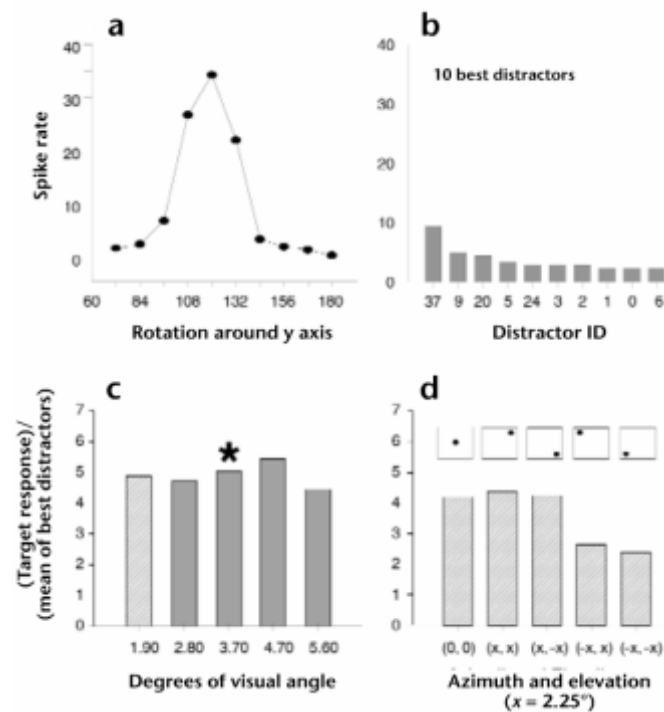


Figure 5  
A view-selective neuron responding invariantly to changes in size and position. (A) Tuning curve showing activity of for a limited number of views of the object. (B) Responses of the cell to the 10 best distractor stimuli. The cell's response change indices to size (C) and position (D).

(Logotheitis, Pauls, Poggio 1995)

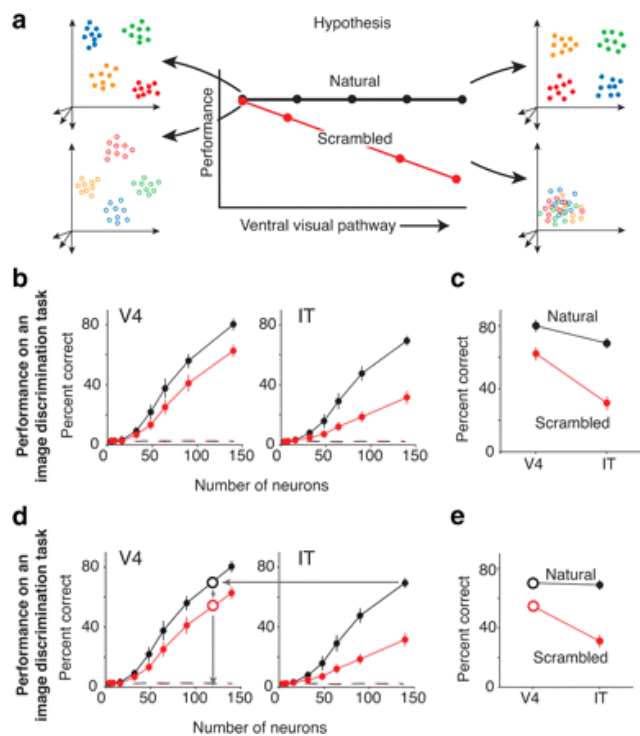


Figure 6  
Results from experiment probing conjunction sensitivities of the V4 and IT.

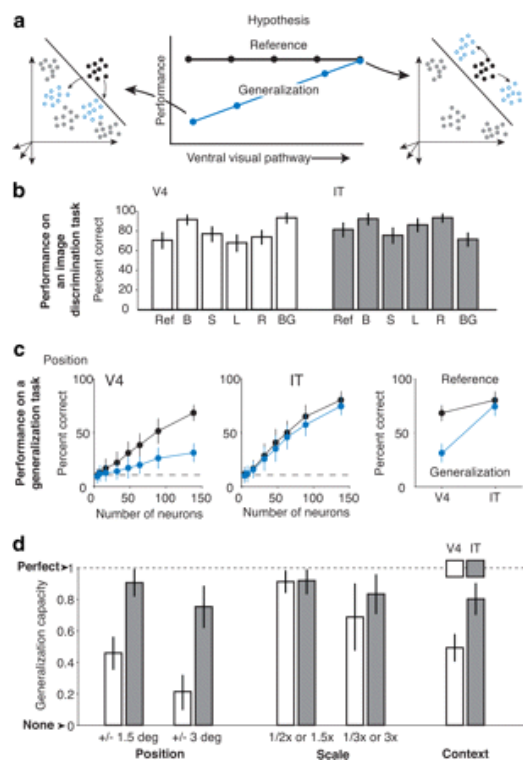


Figure 7  
Results from experiment probing for generalization capacity of the V4 and IT.

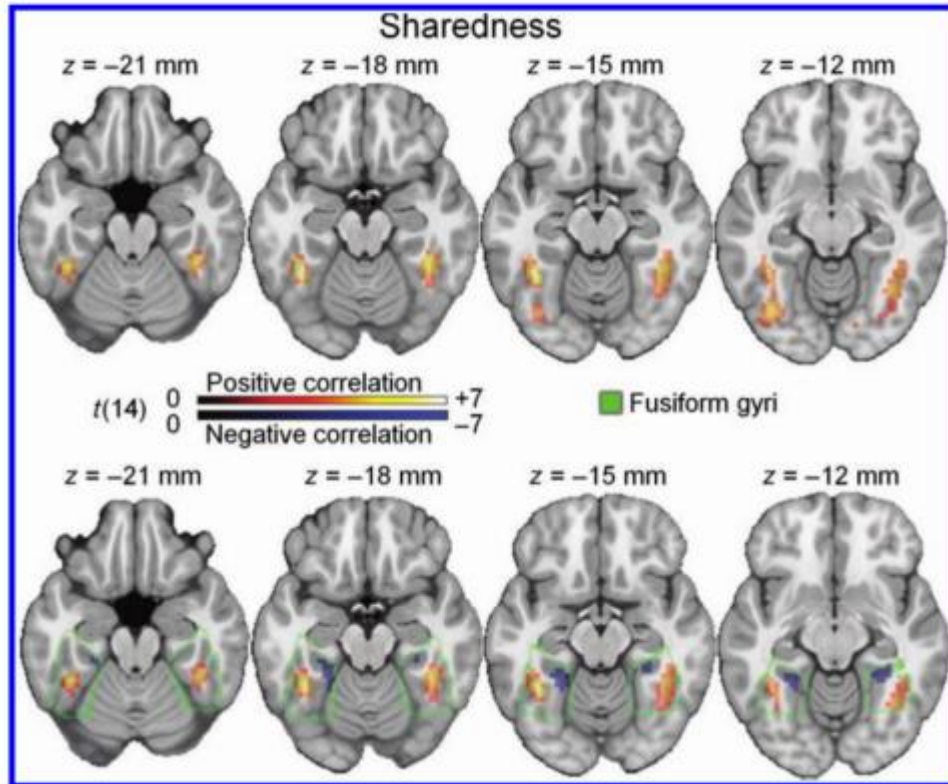


Figure 8  
Activation of the fusiform gyrus during presentation of object with high “Sharedness” statistics.  
(Tyler et al., 2013)

# Work Cited

- Borji, A. and Itti, L., "Human vs. Computer in Scene and Object Recognition," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 113-120.
- Brincat, Scott L., and Charles E. Connor. "Underlying Principles of Visual Shape Selectivity in Posterior Inferotemporal Cortex." *Nature Neuroscience Nat Neurosci* 7.8 (2004): 880-86. Web.
- Gallant, J., J. Braun, and D. Van Essen. "Selectivity for Polar, Hyperbolic, and Cartesian Gratings in Macaque Visual Cortex." *Science* 259.5091 (1993): 100-03. Web.
- Callaway, E. M. Structure and function of parallel pathways in the primate early visual system. *The Journal of physiology* **566**, 13–19 (2005).
- Cichy, R., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nature Neuroscience* **17**, 455–462 (2014).
- Connor, C., Brincat, S. & Pasupathy, A. Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* **17**, 140–147 (2007).
- DiCarlo, JJ & Maunsell, J. Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature neuroscience* (2000). doi:10.1038/77722
- DiCarlo, J., Zoccolan, D. & Rust, N. How does the brain solve visual object recognition? *Neuron* **73**, 415–34 (2012).
- Enns, James T. *The Thinking Eye, the Seeing Brain: Explorations in Visual Cognition*. New York: W.W. Norton, 2004. Print.
- Felleman, D. & Essen, D. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb Cortex* **1**, 1–47 (1991).
- Gallant, J., J. Braun, and D. Van Essen. "Selectivity for Polar, Hyperbolic, and Cartesian Gratings in Macaque Visual Cortex." *Science* 259.5091 (1993): 100-03. Web.
- Hung, C. P., Kreiman, G., Poggio, T. & DiCarlo, J. J. Fast readout of object identity from macaque inferior temporal cortex. *Science* **310**, 863–6 (2005).
- Isik, L, Meyers, EM & Leibo, JZ. The dynamics of invariant object recognition in the human visual system. *Journal of ...*(2014). doi:10.1152/jn.00394.2013
- Kravitz, DJ, Saleem, KS & Baker, CI. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences* (2013).
- Lee, T. Computations in the early visual cortex. *J Physiology-paris***97**, 121–139 (2003).
- Logothetis, N. & Sheinberg, D. Visual Object Recognition. *Annu Rev Neurosci* **19**, 577–621 (1996).
- O'Reilly, Randall C. et al. "Recurrent Processing during Object Recognition." *Frontiers in Psychology* 4 (2013): 124. *PMC*. Web. 27 May 2016.
- Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nat Neurosci* **2**, 1019–1025 (1999).
- Rust, N. C. & DiCarlo, J. J. Selectivity and tolerance (invariance) both increase as visual information propagates from cortical area V4 to IT. *The Journal of Neuroscience* **30**, 12978–12995 (2010).
- Sato, T. "Interactions of Visual Stimuli in the Receptive Fields of Inferior Temporal Neurons in Awake Macaques." *Exp Brain Res Experimental Brain Research* 77.1 (1989): n. pag. Web.
- Serre, T., Oliva, A. & Poggio, T. A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci* **104**, 6424–6429 (2007).



- Tang, H, Buia, C, Madhavan, R, Crone, NE & Madsen, JR. Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron* (2014).
- Tsunoda, Kazushige, Yukako Yamane, Makoto Nishizaki, and Manuba Tanifuji. "Complex Objects Are Represented in Macaque Inferotemporal Cortex by the Combination of Feature Columns." *Nature* 4.8 (2001): n. pag. *Nature Neuroscience*. Web. 27 May 2016.
- Tyler, LK, Chiu, S, Zhuang, J & Randall, B. Objects and categories: feature statistics and object processing in the ventral stream. *Journal of Cognitive Neuroscience* (2013).