Oxford Research Encyclopedia of Neuroscience

Summary and Keywords

Humans and other primates rely on vision. Our visual system endows us with the ability to perceive, recognize, and manipulate objects, to avoid obstacles and dangers, to choose foods appropriate for consumption, to read text, and to interpret facial expressions in social interactions. To support these visual functions, the primate brain captures a high-resolution image of the world in the retina and, through a series of intricate operations in the cerebral cortex, transforms this representation into a percept that reflects the physical characteristics of objects and surfaces in the environment. To construct a reliable and informative percept, the visual system discounts the influence of extraneous factors such as illumination, occlusions, and viewing conditions. This perceptual "invariance" can be thought of as the brain's solution to an inverse inference problem in which the physical factors that gave rise to the retinal image are estimated. While the processes of perception and recognition seem fast and effortless, it is a challenging computational problem that involves a substantial proportion of the primate brain.

Keywords: ventral pathway, macaque visual cortex, invariant object recognition, temporal pathway, "what", pathway, rhesus monkey, neurophysiology, shape discrimination, cortical visual processing, partial occlusion

Signal Transformations Along the Object-Processing Pathway

When we look at a complex visual scene, its image is encoded in the activity patterns of retinal cells as a fine-grained representation of local contrast. This representation is highly dependent on viewing conditions (e.g., the position and pose of objects in the scene, the viewing distance, the plane of focus) and on extraneous factors, including occlusions and illumination conditions. The retinal image is transformed via successive stages of cortical processing along the object-processing pathway (also ventral, temporal, or "what" pathway). In the primate brain, this pathway runs from V1 through areas V2 and V4, terminating in subregions of the inferotemporalcortex (IT) (Figure 1) (Felleman & Van Essen, 1991). The transformed visual representations along this pathway are thought to facilitate the parsing of visual scenes into component objects and regions and to

mediate our perception of scenes by signaling the identity of these components and their spatial relationships (Felleman & Van Essen, 1991; Logothetis & Sheinberg, 1996; Ungerleider & Mishkin, 1982). These transformed representations also underlie our ability to recognize objects in the scene irrespective of size or position in the retinal image, 3D pose, or clutter and occlusions due to nearby objects.
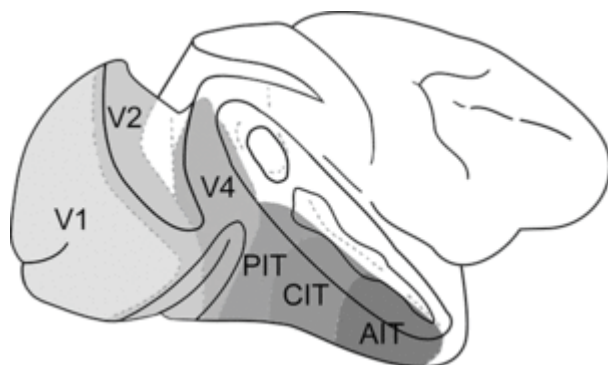
Figure 1. The Object-Processing Pathway. Lateral view of the macaque brain showing the different cortical areas involved. This pathway, known variously as the "what," ventral or temporal pathway, is important for visual form processing and object identification.

(Adapted with permission from Parker, 2007.)

To understand the algorithmic operations that mediate shape perception and recognition in the primate brain, one experimental strategy has been to investigate the representational basis in each cortical processing stage, i.e., to identify the visual features encoded in areas V1, V2, V4, and IT. Once the features encoded are understood, we could attempt to deduce the algorithms that support the transformations from one level to the next. In area V1, neurons encode visual stimuli in terms of local orientation and spatial frequency (Albrecht, De Valois, & Thorell, 1980; Hubel & Wiesel, 1959, 1968; Movshon, Thompson, & Tolhurst, 1978A, 1978B). Specifically, each small patch of the retinal image is represented by a subpopulation of V1 neurons that signal the orientation and scale of image features at a particular position in visual space, and the full retinal image is represented by thousands of such V1 subpopulations that tile the visual field. In area V2, some neurons encode line conjunctions and orientation combinations (Anzai, Peng, & Van Essen, 2007; Hegde & Van Essen, 2000; Ito & Komatsu, 2004), and others are sensitive to spatial structure in texture patches (Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013). Whereas the representations at the level of V1 and V2 are consistent with encoding "stuff," i.e., the surface characteristics of images (Adelson & Bergen, 1991), an explicit representation of object boundaries, an intermediate encoding of "things," begins to emerge in area V4.

Many V4 neurons encode stimuli in terms of the curvature of the boundary at specific positions relative to object center (Pasupathy & Connor, 2001), and their responses cannot be explained by a spectral receptive field model (Oleskiw, Pasupathy, & Bair, 2014). For example, one neuron may respond preferentially to stimuli that contain a sharp convexity to the upper right (Figure 2A), and another to stimuli that contain a concavity to the top (Figure 2B). The responses of neurons that encode boundary form can be described quantitatively as a two-dimensional Gaussian function in object shape space, defined by curvature and the angular position relative to object center. Together, a

population of such neurons can provide a complete representation of isolated objects in terms of the boundary features (Figure 2C; see also Pasupathy & Connor, 2002).

Curvature-based object representations have also been demonstrated in subregions of IT (Brincat & Connor, 2004; Ponce, Hartmann, & Livingstone, 2017; Schwartz, Desimone, Albright, & Gross, 1983), and fMRI imaging studies have confirmed the existence of a curvature-processing network in visual cortex (Yue, Pourladian, Tootell, & Ungerleider, 2014). These neurophysiological findings of a curvature-based code for object representation are consistent with shape theories and psychophysical studies, which have long argued for the importance of curvature as a representational basis (Asada & Brady, 1984; Attneave, 1954; Besl & Jain, 1985; Marimont, 1984; Verri & Yuille, 1986; Watt & Andrews, 1982; Wilson, Wilkinson, & Asaad, 1997). Additionally, these findings support the hypothesis for a structural representation of objects in terms of object parts and their positional relationships (Biederman, 1987; Connor, Brincat, & Pasupathy, 2007). Structural representations can be compact—because few parts can define an object—and versatile—since the same dictionary of parts can be used to represent a plethora of objects (Connor et al., 2007).
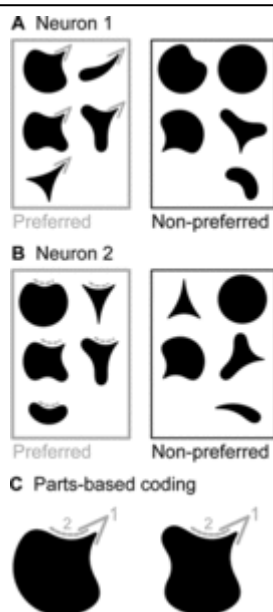
Figure 2. Shape Representation in Cortical Area V4. **A**. The stimulus preferences of a hypothetical neuron are shown. This neuron responds strongly to some shapes (see gray rectangle) but not others (see black rectangle). These responses can be explained in terms of selectivity for a sharp convex projection to the top right, relative to object center. **B**. The stimulus preferences of a second hypothetical neuron. This neuron's preferences for shape stimuli can be explained in terms of selectivity for a broad concave indentation to the top, relative to object center. **C**. Because both shapes shown contain a sharp convexity to the top right (1) and a concavity to the top (2), these stimuli will evoke a strong response from hypothetical neurons 1 and 2. Different V4 neurons respond to the different contour features that constitute the shape. In this way, any arbitrary shape can be encoded in terms of its component contour features in area V4.

(Illustration based on previously published data; Pasupathy & Connor, 2001, 2002.)

Beyond boundary curvature, there is evidence that IT neurons represent objects in terms of their skeletal shape, another structural basis (Hung, Carlson, & Connor, 2012). IT neuronal responses are also more strongly modulated by non-accidental properties that are invariant to rotations in depth (e.g., straight versus curved boundaries) than by equivalent metric variations that are depth-dependent, consistent with the proposal that IT shape representations highlight perceptually salient features relevant for essential categorizations (Kayaert, Biederman, & Vogels, 2003, 2005A; Kayaert, Biederman, Op de Beeck, & Vogels, 2005B).

IT neurons provide a sparse representation of objects (see, e.g., Gross & De Schonen, 1992; Tamura & Tanaka, 2001), with clustered groups of neurons having similar preferences (Desimone, Albright, Gross, & Bruce, 1984; Downing, Jiang, Shuman, & Kanwisher, 2001; Fujita, Tanaka, Ito, & Cheng, 1992; Perrett, Rolls, & Caa, 1982; Tanaka, 1996; Tsao, Freiwald, Tootell, & Livingstone, 2006). For example, neuronal clusters selective for faces and body parts have been demonstrated in several subregions of IT. While this evidence supports the proposal of a "gnostic field," a region of cortex subserving a particular class of perception (Konorski, 1967), it does not support the idea of a "grandmother cell"—a hypothetical neuron that responds to one specific visual stimulus and no other. Instead, the current working model proposes a sparse distributed representation of objects mediated by groups of neurons that encode specific structural parts and modulated by semantic category (Kiani, Esteky, Mirpour, & Tanaka, 2007). Semantic-based representations may rely on learning,

attention, and behavioral relevance (Cukur, Nishimoto, Huth, & Gallant, 2013). This sparse but non–grandmother-cell-like encoding has also been described in the medial temporal lobe, a region known to be important for memory and recognition (Quiroga, Kreiman, Koch, & Fried, 2008). Finally, recent evidence suggests that IT neurons may represent the weight of objects in addition to their form (Gallivan, Cant, Goodale, & Flanagan, 2014). To summarize, visual signals are transformed across stages of the object-processing pathway to give rise to a representation that reflects not only the physical features of objects in a scene, but also their cognitive and behavioral relevance.

Invariant Representations
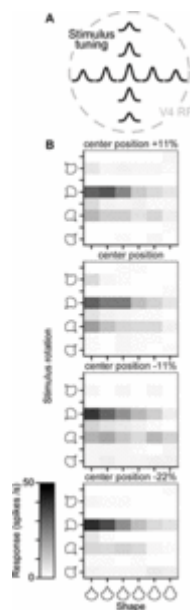
Figure 3. Position-Invariant Tuning for Stimulus Form. **A**. Each bell-shaped curve represents a hypothetical neuron's tuning for stimulus shape at various positions within the receptive field (dashed circle). The tuning curve scales with position, but the position of the peak and the shape of the tuning curve do not change as a function of stimulus position. The influence of stimulus position can be expressed in terms of a position-dependent multiplicative gain that modulates the shape tuning curve. **B**. Example responses of a V4 neuron to 48 stimuli presented at 4 positions within the RF. Each panel shows the neuronal responses at one position within the RF to 6 shapes (abscissa), each presented at 8 rotations (ordinate). The magnitude of positional shifts is expressed as center ± % of RF diameter. This neuron was strongly tuned for stimulus shape and rotation, responding best to stimuli with a convex projection to the lower left. Tuning preferences were similar at all positions tested. The coefficient of correlation between responses at center and other positions was >0.9, implying similar shape preferences at all positions.

(Illustration based on previously published data; El-Shamayleh & Pasupathy, 2016.)

The visual image cast on the retina depends on the relationship between the viewer and the object. Thus, the retinal image of an object can be dramatically different depending on the precise position and pose of the object, gaze angle, viewing distance, and the observer's plane of focus. These factors pose a major challenge to the visual system, because successful recognition requires mapping different retinal images to a single object identity. The implementation of this many-to-one mapping from retinal image to object identity has captivated biological and computer vision scientists over the years. Hubel and Wiesel (1962) originally observed that, whereas simple and complex V1 cells are both selective for line orientation, the responses of simple cells are highly dependent on stimulus position within the receptive field (RF), but the responses of complex cells are not. They speculated that the "generalization" of orientation tuning in complex cells may engender generalization of form selectivity across position within the RF of cells selective for higher order form. Indeed, position-invariant neuronal shape coding has been demonstrated by several groups in V4 (El-Shamayleh & Pasupathy, 2016; Gallant, Braun, & Van Essen, 1993) and IT (Desimone et al., 1984; Ito, Tamura, Fujita, & Tanaka, 1995; Rust & DiCarlo, 2010; Sáry et al., 1993; Tanaka, 1996). Within the RF of a given neuron, response magnitude may change across stimulus

position, but the neuron's stimulus preference is typically maintained (Figure 3A). V4 neurons are tuned to boundary curvature in an object-centered reference frame (see Figure 2), and positional shifts simply result in a translation of object center but no associated change in the position of boundary features relative to object center. Thus, curvature tuning in V4 is expected to be independent of stimulus position, and the responses to a shape stimulus *s* at position *p* can be described by the following equation:

$$r(s,p)=f(s)\times g(p)$$

where *f* denotes the tuning for stimulus form and *g* denotes the modulatory influence of the absolute position of the stimulus within the RF. The independence of stimulus tuning implies that neuronal preferences for shape stimuli translated spatially will be strongly correlated; this is indeed what we have observed in area V4 (Figure 3B) (El-Shamayleh & Pasupathy, 2016). It is important to note that the invariance of shape tuning does not imply the loss of position information; rather, the responses of most V4 and IT neurons are modulated by stimulus position, i.e., *g* is a function of *p*. Therefore, neurons in V4 and IT encode information about stimulus identity and position in a separable manner, allowing for both attributes to be decoded from neuronal populations.
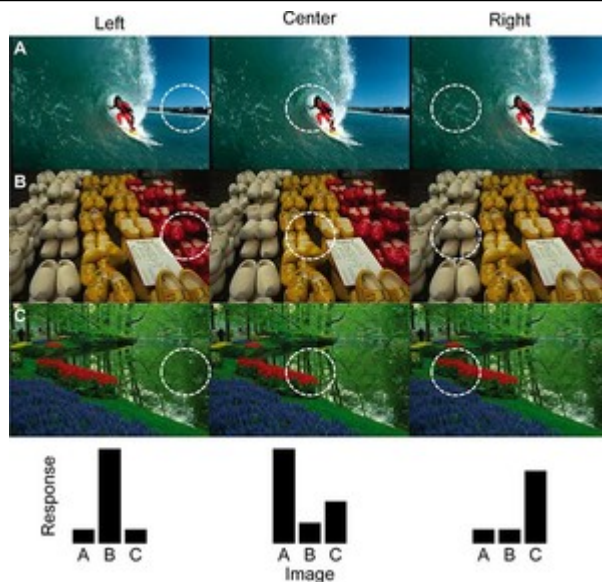
Figure 4. Pitfalls in Assessing Position-Invariant Coding. A hypothetical experimental paradigm (top) and results (bottom) are shown. Three arbitrary natural images (**A-C**) are presented at three positions relative to a neuron's RF (white circle): shifted left, centered, shifted right relative to RF. Stimuli extend well beyond the RF of the neuron. If the neuron were only selective for the color red, it would respond best when a red patch in the image falls on its RF, i.e., for stimulus **B** at the left position, stimulus **A** at the center, and stimulus **C** at the right position. Such an observation could be interpreted as position-dependent coding, but the same neuronal responses could be fully accounted for by biases in the particular images and stimulus manipulations used. To avoid this pitfall, studies of invariant coding should take into account the feature selectivity and spatial RF of the neurons studied.

Some studies have reported limited position-invariant shape tuning in V4 and IT, but this finding may be due to how invariance is tested. A neuron with position-invariant shape tuning may appear to be highly position-dependent in two cases: (1) if invariance is tested beyond the confines of the RF (where measured responses will be dominated by neuronal noise) or (2) if we do not know the critical stimulus attribute that drives neuronal responses. For example, when probed with a set of natural stimuli, a neuron that prefers a red stimulus in its RF may appear to exhibit position-dependent stimulus preference if its color preference is not known (Figure 4). Thus, tuning invariance cannot be adequately examined by presenting an arbitrary set of stimuli at multiple positions within the visual field. To disambiguate response patterns consistent with multiple encoding strategies, it is important to characterize the spatial RF and then probe position invariance using stimuli in which a relevant stimulus feature is varied parametrically within the confines of the RF. With this controlled experimental approach, neurons in V4 and IT consistently show invariance in their form tuning across position as illustrated (Figure 3B; see also El-Shamayleh & Pasupathy, 2016).

As with position, size-invariant tuning for object shape has also been demonstrated in V4 and IT (Brincat & Connor, 2004; El-Shamayleh & Pasupathy, 2016; Hikosaka, 1999; Ito et al., 1995; Logothetis & Sheinberg, 1996; Rust & DiCarlo, 2010; Sáry et al., 1993; Schwartz et al., 1983; Tanaka, 1996). In V4, ~70% of shape-selective neurons encode boundary form in a size-invariant manner; their responses are consistent with encoding the curvature of boundary segments, defined in Cartesian coordinates, and normalized by stimulus size (El-Shamayleh & Pasupathy, 2016), which is equivalent to encoding

curvature in polar coordinates. In V4 and IT, the responses of neurons that exhibit size-invariant tuning for object shape are nevertheless modulated by stimulus size (e.g., see El-Shamayleh & Pasupathy, 2016). Thus, information about stimulus form as well as size can be decoded from a population of these neurons.
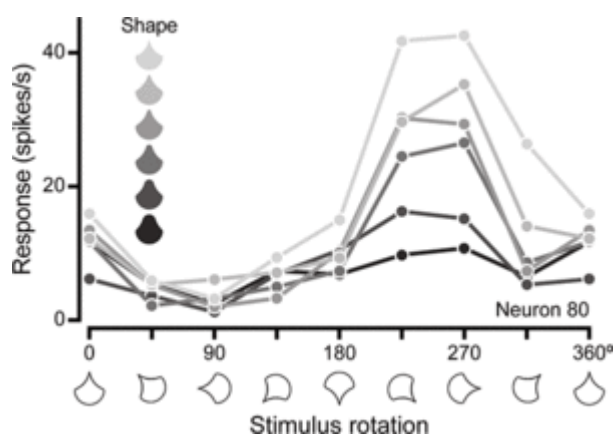
Figure 5. Tuning for Stimulus Rotation. Responses of an example V4 neuron with strong tuning for stimulus rotation. Each line shows the responses of a neuron to 8 stimulus rotations (abscissa). Responses were strongest for stimuli at 225° and 270° and weaker at other rotations. This tuning for stimulus rotation was consistent across all 6 shapes shown (see color-coded stimuli in inset).

(Illustration based on previously published data; El-Shamayleh & Pasupathy, 2016.)

Unlike position and size, there appears to be limited invariance for object rotation in visual cortex (Figure 5; e.g., see El-Shamayleh & Pasupathy, 2016; Logothetis & Sheinberg, 1996). The observation of weak rotation invariance is consistent with a structural, parts-based code for objects in V4 and IT. When stimuli are rotated in the fronto-parallel plane, the position of boundary features will change relative to object center. Additionally, when stimuli are rotated in depth, some features may become occluded. Under these conditions, a V4 neuron tuned to a specific boundary feature relative to object center will not be expected to maintain its preference when stimuli are rotated. Thus, whereas size- and position-invariant object recognition in the visual system could be mediated by the responses of neurons in V4 and IT that exhibit size- and position-invariant form tuning, rotation-invariant recognition would require storing many templates in memory for each object, and matching the stimuli viewed against any of these templates would trigger successful recognition (Logothetis & Sheinberg, 1996; Riesenhuber & Poggio, 2000).

Occlusions and Scene Context

Another major challenge for the visual system is the fact that the retinal image of an object also depends on contextual factors, such as occlusions caused by other objects, illumination conditions, and shadows. This means that the same retinal image could be the product of different object/context combinations. For example, an image (Figure 6A, left) could be interpreted either as a collection of arbitrary dark gray shapes, or as a letter B that is partially occluded by light gray rectangles. The latter interpretation is more salient when the occluding rectangles are not in the same color as the background (Figure 6A, right). Thus, developing a visual percept from a retinal image requires solving

an ill-posed, inverse problem that lacks a unique solution. In other words, this is a one-to-many mapping problem, unlike the invariance problem discussed earlier, which is a many-to-one mapping problem. Not much is known about how the primate brain accomplishes this feat, but neurophysiological experiments are beginning to reveal some of the underlying principles.
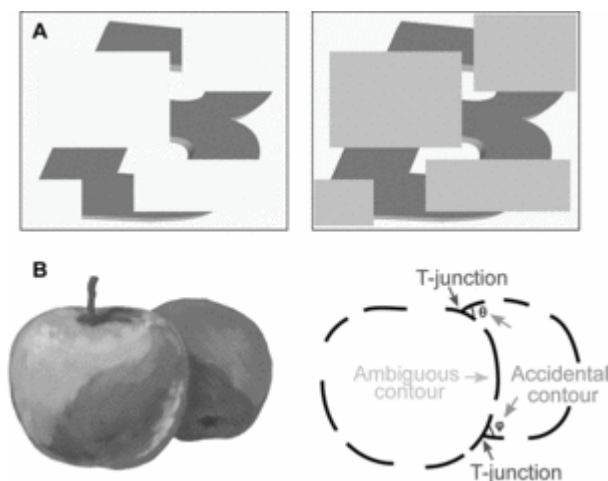

Figure 6. Partial Occlusions. **A**. The perceptual interpretation of the image in the two panels is drastically different depending on whether the occluding rectangles are the same color as the background (left) or different (right). Other regions of the image are identical in both panels. **B**. A partially occluded apple and its component contours in the retinal image. The T-junctions, accidental contours, and the ambiguous contour are labeled.

Let us consider the problem of partial occlusion. When one apple partially occludes another (Figure 6B, left), the retinal image includes several contour features at the junction of the occluding and occluded object surfaces. First, a pair of T-junctions (Figure 6B, right) is formed at the intersection of the occluded and occluding boundaries. Second, accidental contours defined by a curvature discontinuity (angles θ and φ) are evident at the T-junctions. Third, an ambiguous contour, which could be interpreted as convex or concave, intercedes two object surfaces. The set of boundary contours illustrated (Figure 6B) is equally consistent with the apples pictured and with a crescent shape on the right adjoining a circular object on the left, with no partial occlusion. However, perceptually, we discount the accidental contours and assign the ambiguous contour to the occluding object, such that we perceive an apple on the right that is partially occluded by another on the left. Shape theorists and psychophysicists have long postulated that the perception of partial occlusion begins with the detection of T-junctions (Clowes, 1971; Elder & Zucker, 1998; Guzman, 1968; Helmhotz, 1910; Huffman, 1971; Rubin, 2001; Waltz, 1975), and models of image segmentation have typically invoked the explicit or implicit encoding of T-junctions, followed by the instantiation of rules to identify the direction of the occluding boundary at T-junctions (Craft, Schutze, Niebur, & von der Heydt, 2007; Sajda & Finkel, 1995; Zhaoping, 2005).
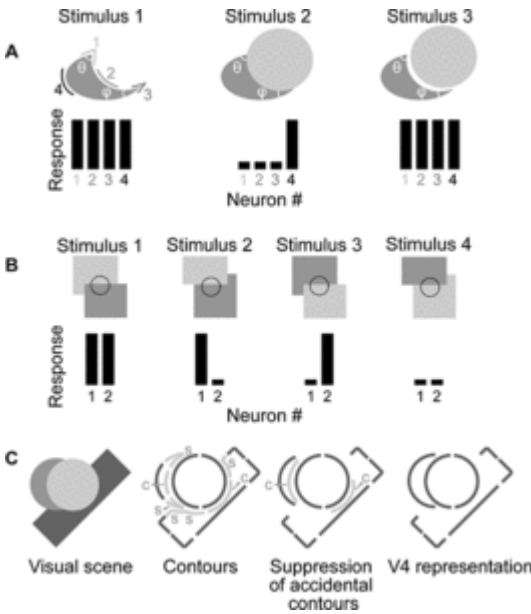
Figure 7. Suppression of Accidental Contours and Border Ownership Signals in Visual Cortex. **A**. Suppression of accidental contours, as observed in V4. Responses of 4 hypothetical neurons (Neuron 1–4) to three stimulus conditions (Stimuli 1–3): a crescent shape presented in isolation, adjoined by a circle implying partial occlusion, and alongside a circle with a small gap in between. Neurons 1–4 have the following feature preferences: (1) sharp convexity to the top, (2) concavity to the upper right, (3) sharp convexity to the right, (4) broad convexity to the bottom left. All four neurons will respond strongly to the crescent in isolation (Stimulus 1) and to the crescent and circle with a gap in between (Stimulus 3). Only Neuron 4, which prefers the broad convexity, will respond to the partial occlusion condition (Stimulus 2). (Illustration based on results presented in Bushnell et al., 2011.) **B**. Border ownership signals, as observed in V2. Responses of two hypothetical V2 neurons to four stimulus conditions are shown. Black circle indicates the neuron's RF. Stimulus characteristics within the RF are identical for Stimuli 1–2 and Stimuli 3–4. Responses of Neuron 1 illustrate tuning for contrast polarity: similar responses to Stimuli 1 and 2 and to 3 and 4. Responses of Neuron 2 illustrate border ownership signals. Responses are modulated by the position of the occluding object, i.e., the object to which the edge belongs. In this case, responses are strongest when the edge within the RF belongs to the object below the RF regardless of edge contrast polarity. (Illustration based on results previously presented in Zhou et al., 2000.) **C**. A schematic of how a simple scene may be encoded by a population of V4 neurons (modified from Bushnell et al., 2011). An example image (first panel) and its component contours (second panel) are shown. Accidental sharp convexities at T-junctions are labeled s. Concavities formed at the junction between the occluding and occluded surfaces are labeled c. The suppression of accidental sharp convexities produces a fragmented contour map (third panel). Collinear and co-circular facilitation mechanisms lead to the suppression of the ambiguous concavities (c) and the development of border ownership signals (fourth panel).

Neurophysiological studies in area V4 suggest that partial occlusion context provided by neighboring objects strongly modulates neuronal responses (Figure 7A; also see Bushnell, Harding, Kosai, & Pasupathy, 2011). V4 neurons that preferentially encode a sharp convexity to the top, or to the right, or a concavity to the top-right (Figure 7A; see neurons 1–3) will respond strongly to a crescent shape in isolation (stimulus 1), but not when the crescent is adjoined by a circle (compare responses of neurons 1–3 to stimulus 1 and 2). In the latter case, the angles θ and φ are accidental contour features, and the boundary between the two objects may be convex or concave, depending on whether it is assigned to the circle or to the crescent shape, respectively. When a small gap is introduced between the crescent and the circle (stimulus 3), the partial occlusion context no longer exists: θ, φ, and the concavity are all real contour features. In this case, neuronal responses are again comparable to the crescent in isolation. Many V2 neurons also exhibit similar context-dependent response modulations. Von der Heydt and colleagues (Zhou, Friedman, & Von der Heydt, 2000) have studied V2 neuronal responses to a variety of partially occluded stimuli in which the local contrast polarity and position of the occluding object were systematically varied (Figure 7B). While many V2 neurons (e.g., see neuron 1) simply signal the contrast polarity of the edge running through the RF (black dotted circle), others have a preference for the position of the object to which the edge belongs, i.e., the position of the occluding object. For example, neuron 2 prefers an edge when it belongs to an object

below the RF; its responses to stimulus 1 are stronger than to stimulus 2, even though both provide identical stimulation within the RF.

Thus, signals related to partial occlusion context provided by neighboring objects modulates neuronal responses in the intermediate stages of visual cortex to produce a representation that faithfully encodes real contours, that suppresses accidental contour features, and that resolves the ambiguity of contours that lie between the occluding and occluded surfaces. These contextual modulations are fast: suppression of accidental contours in V4 is observed, on average, ~60 ms after stimulus onset (Bushnell et al., 2011), while border ownership signals in V2 are observed 60–70 ms after stimulus onset (Zhou et al., 2000). These contextual signals are therefore likely based on neuronal processing within V4 and do not reflect feedback from higher cortices such as IT or prefrontal cortex, where response latencies are longer. Results from additional control experiments are consistent with the hypothesis that local competition between two contours—a sharp convexity at the T-junction and a smooth continuous contour (e.g., contour labeled *s* and its neighboring convex contour in Figure 7C)—at the same location in the visual scene could produce the observed suppression of accidental contours (Bushnell et al., 2011). Local competition between contours could be achieved, for instance, if neurons tuned to a broad convexity inhibit neurons tuned to a sharp convexity, and where these two neuronal groups have a common set of V1 or V2 inputs. Local competition augmented by collinear facilitation (Craft et al., 2007; Sajda & Finkel, 1995; Zhaoping, 2005) could facilitate correct boundary assignment at the interface between the occluding and the occluded object boundaries (as diagrammed in Figure 7C). Thus, contextual modulations that produce suppression of accidental contours and accurate border assignment could represent the bottom-up instantiation of the Gestalt prior for continuity.
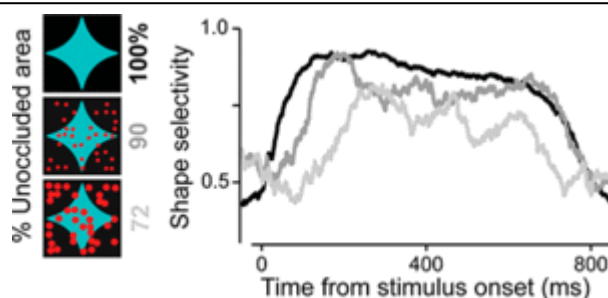
Figure 8. Shape Selectivity Under Partial Occlusion. Shape selectivity of an example V4 neuron, measured as the area under the receiver-operating characteristic curve, constructed from the neuron's responses to two shape stimuli. Line color represents the level of occlusion provided by a set of random dots in a contrasting color of varying diameter (see inset shapes). Occlusion level was quantified as % unoccluded area: 100% level (black line) corresponds to unoccluded stimuli, and lower numbers correspond to stimuli with greater occlusion. For unoccluded stimuli, shape selectivity emerges rapidly and peaks early; with increasing occlusion, selectivity increases more slowly and peaks later.

(Illustration based on data published in Kosai et al., 2014.)

In addition to bottom-up processes, feedback from higher cortical areas is also likely to facilitate perception under occlusion (Rust & Stocker, 2010). When the occluded object is highly familiar, e.g., a tiger in the bushes, recognition could be triggered rapidly at the highest stages of object processing on the basis of a single diagnostic feature (e.g., a tiger's stripes). Feedback to intermediate stages of cortical processing could then clarify the representations that underlie perception. Neurophysiological studies in area V4 support this hypothesis. When monkeys are required to report whether two shape stimuli presented in sequence are the same or different, shape selectivity in V4 emerges earlier for unoccluded stimuli and later for occluded stimuli (Figure 8; see also Kosai, El-Shamayleh, Fyall, & Pasupathy, 2014). The delayed selectivity for occluded stimuli is observed only when the animal is engaged in a perceptual shape discrimination task and not under passive fixation conditions (our unpublished observations). In IT and the ventrolateral prefrontal cortex, two areas hypothesized to be engaged in object recognition and memory, many neurons respond more vigorously and more selectively to occluded objects compared to unoccluded objects (Fyall, El-Shamayleh, Choi, Shea-Brown, & Pasupathy, 2017; Namima & Pasupathy, 2016). The stronger responses to occluded stimuli in higher cortices, when fed back to V4, could produce the delayed and augmented selectivity observed under occlusion and could facilitate the perception of occluded stimuli.

To summarize, the visual system relies on both bottom-up neural signals relaying spatial form cues as well as top-down neural signals carrying influences of familiarity and experience. Together, these signals underlie the generation of stable and consistent object percepts in the face of ambiguous representations due to occlusion.

Contour Grouping and Segmentation

Scene segmentation—the process of parsing a scene into a meaningful arrangement of regions and objects—is another computational problem faced by the visual system. Not much is known about how this is solved by the primate brain, but an emerging view suggests that curvilinear contour grouping, combined with contextual modulations and top-down signals, may facilitate the preferential encoding of object boundaries and thus contribute to scene segmentation.

In computer vision, algorithms have traditionally focused on region-based segmentation, where the image is partitioned into pixel sets with coherent image properties such as brightness, color, and texture (Leung & Malik, 1998). Region-based algorithms can be fast, but they can fail when objects contain steep gradients in color, texture, or luminance. Recent algorithms therefore combine region-based segmentation with curvilinear contour grouping (Leung & Malik, 1998), the main strategy proposed by psychophysicists. Here, contours are grouped based on Gestalt rules (Wertheimer, 1938) of similarity, proximity, continuity, common fate, symmetry, and convexity. In pathfinder displays with Gabor patch elements (Figure 9A), human subjects can detect a continuous path easily if the elements lie parallel to the path, but not if they lie orthogonal to it. Field, Hayes, and Hess (1993) hypothesized that this asymmetry in detection was due to curvilinear facilitation: the enhanced encoding of contour elements that lie along a curve.
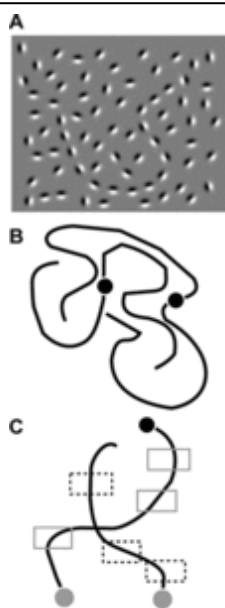
Figure 9. Parallel and Serial Contour Grouping. **A**. Example pathfinder display. The string of elements on a curve is salient, and human subjects can detect it despite the distractor elements (based on Field et al., 1993). This process is thought to depend on collinear facilitation. **B**. Curve tracing task. Subjects are required to report if the two black dots belong to the same curve (based on Jolicoeur et al., 1986). In this task, behavioral reaction time depends on the distance between the dots. **C**. Another variant of the curve tracing task. Monkeys are required to make a saccade from the fixation spot (black dot) to the target dot that is connected to it. V1 neurons with RFs on the target path (solid rectangles) but not those with RFs on the distractor path (dotted rectangles) show enhanced activity during the sustained portion of the neuronal response (Roelfsema et al., 1998).

Neurophysiological studies in V1 have demonstrated that neuronal responses to stimuli within the RF can be facilitated by collinear elements outside the RF, especially in the presence of randomly oriented distractors (Bauer & Heinze, 2002; Kapadia, Ito, Gilbert, & Westheimer, 1995; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998). Long-range horizontal connections in V1 could underlie this collinear facilitation (Schmidt, Goebel, Lowel, & Singer, 1997), as could feedback signals from higher cortices: for example, neurons in V4 that are sensitive to the boundary of elongated curves could modulate the responses in early visual cortex (see Roelfsema, 2006, for a more c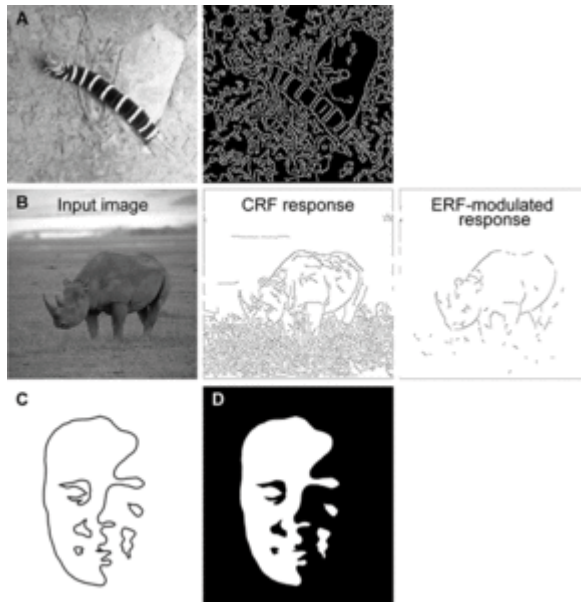omplete discussion). When visual stimuli include three-dimensional surface configurations, collinear facilitation processes in V2 appear to integrate depth information such that neuronal responses are consistent with encoding the amodal completion of occluded contours and the segmentation of surfaces (Bakin, Nakayama, & Gilbert, 2000).

In addition to collinear facilitation, which could operate in parallel and segregate collinear elements from background elements, psychophysical studies suggest that a serial process may be required for segregating one curve from another. In curve-tracing tasks in which subjects are asked to report whether two points in a visual display are connected (Figure 9B), behavioral reaction time increases with increasing distance between the queried points (Jolicoeur, Ullman, & McKay, 1986). Neurophysiological recordings in monkeys trained to make a saccade to a dot on a target curve in the presence of a distractor curve (Figure 9C) show neuronal response facilitation in V1 ~100 ms or more after stimulus onset during the sustained portion of the neuronal response

(Roelfsema, Lamme, & Spekreijse, 1998). This finding supports the existence of an incremental contour grouping process. Under this hypothesis, enhanced activity would gradually spread across the representation of an object in visual cortex; such a process corresponds to the labeling of image elements with object-based attention (Roelfsema & Houtkamp, 2011). To summarize, contour grouping in visual cortex likely includes a fast collinear facilitation process and a slower serial grouping process that is related to object-based attention.

In natural scenes, segmentation on the basis of contour grouping alone can be challenging. The challenge arises because even a very simple natural scene devoid of clutter is associated with an extremely complex edge map (Figure 10A), and teasing apart which contours belong together can be difficult. A novel idea, proposed by Gheorghiu and colleagues (2014), is that contextual modulations in visual cortex serve to de-texturize images, enhancing contour representations at the expense of textures and thus facilitating the signaling of object boundaries. Many neurons in visual cortex exhibit tuned surround

Figure 10. Component Edges in Natural Images and the Role of Contextual Modulations in De-texturizing Scenes. **A**. A natural scene (left) and the component edges in the scene, as computed by a Canny edge detector (right). **B**. The role of surround suppression in de-texturizing an image. An input image (left), the component edges based on a Canny detector (middle) and the representation of the image by a complex cell model with strong surround suppression (right). (From Gheorgiu et al., 2014.) **C**. An example two-tone Mooney face. **D**. Luminance-based segmentation of the image in **C**, ignoring shadow boundaries, makes correct recognition difficult.

(Inspired by Cavanagh, 1991.)

suppression, where responses are suppressed when the RF center and surround are activated by stimuli with similar characteristics (e.g., see Blakemore & Tobin, 1972; Cavanaugh, Bair, & Movshon, 2002; Kapadia et al., 1995; Knierim & Van Essen, 1992; Levitt & Lund, 1997; Nelson & Frost, 1985; Nothdurft, Gallant, & Van Essen, 1999). Responses to textures could therefore be suppressed due to inherent spatial correlations in texture regions. This idea is illustrated in the output of a model that tiles complex-like cells exhibiting strong iso-orientation suppression across the image (Figure 10B, right). Responses of this model are weak for the grassy regions of the image because neighboring regions are highly correlated.

To illustrate their point, Gheorghiu and colleagues implemented high levels of suppression in the model, which completely eliminated the representation of texture regions, whereas more realistic levels of surround modulations produced partial suppression of the texture regions. A scene, when de-texturized in this way, results in the preferential encoding of object boundaries thereby facilitating segmentation. Psychophysical results from shape-frequency and shape-amplitude adaptation after-effects also support the hypothesis that contextual modulations serve to de-texturize images. In these experiments, subjects were adapted to a single sinusoidal contour and then tested on the perceived frequency and amplitude of another contour (Gheorghiu & Kingdom, 2006, 2007; Kingdom & Prins, 2005). These studies found that while adaptation effects were strong when the adaptor and test were both single contours, effects were much weaker when the adapting contour was flanked by parallel contours yielded a stimulus that resembled a texture. These findings support the hypothesis that neurons encoding single contours in visual cortex may be suppressed when those contours are part of a texture. Results from fMRI experiments also support the hypothesis that texture and contour processing are segregated in the brain. Dumoulin, Dakin, and Hess (2008) evaluated how contrast-energy contained in contours and textures within natural images affected BOLD responses in visual cortex. Whereas BOLD responses in V1 were consistent with representing natural images based on local oriented filters, responses in extrastriate areas were stronger for contours compared to textures, thereby amplifying sparse contour (as opposed to texture) information within natural images. Preliminary neurophysiological results from our group are consistent with this observation: V4 neurons respond more strongly to isolated shapes than to luminance-matched texture patches, and this preference persists even when shapes are overlaid on texture patches or when a texture is shown through a shape aperture (Kim, Bair, & Pasupathy, 2017). Nevertheless, these results do not imply that textures are poorly represented in this part of the visual cortex, and recent neurophysiological studies have documented selective responses to texture patches in V2 that reflect the encoding of higher-order image statistics (Freeman et al., 2013). Taken together, these results support the idea that different neuronal subsets in the intermediate stages of cortex may be specialized to encode contours (in V4) and textured regions (in V2) by taking advantage of differences in the statistical characteristics of these visual cues.

Finally, for successful segmentation to occur, the visual system must distinguish veridical object boundaries from internal contours and borders cast by shadows. This requirement has been elegantly demonstrated with Mooney images (Figure 10C) where the segmentation and interpretation of an image are dramatically different when information from shadows is withheld (Figure 10D; see also Cavanagh, 1991). In such cases, the segmentation process may need to be guided by object prototypes stored in memory that coarsely match subsets of contours in the image. As with occlusion context, image segmentation and scene understanding are likely to depend on feedback signals from higher cortices, triggered by successful recognition. Further experiments are needed to clarify how these processes unfold.

Subscriber: University of Washington; date: 11 March 2019

Models of Visual Processing

One benchmark for success in our pursuit of understanding visual perception is to build models of cortical processing that can accurately predict responses to novel stimuli and match the behavioral levels of object recognition in primates. Currently, we have good models that can capture the responses of simple and complex cells in V1. Beyond V1, however, our progress has been limited. To explain boundary form encoding in V4, Cadieu and colleagues (2007) proposed a contour template model, a specific instantiation of the hierarchical max model proposed by Poggio and colleagues (Riesenhuber & Poggio, 2000). Intuitively, the model achieves selectivity for a sharp convexity to the top of object center, for example, by pooling orientation signals from V1 neurons tuned to 45° and 135°. However, such a contour template model does not capture the object-centered nature of selectivity for boundary form in V4 and fails to achieve the level of position invariance observed in real neurons (Bair, Popovkina, De, & Pasupathy, 2015). In fact, no current model of V4 neurons can account for object-centered encoding. Discovering how object-centered encoding is built in V4 is a necessary and critical step to advancing current models of form processing.

In recent years, with the advent of efficient learning algorithms for deep neural networks, computer vision has made great strides with developing algorithmic solutions for object recognition: on some tasks, these networks have reached levels that are comparable to humans (Kriegeskorte, 2015). Briefly, these feedforward hierarchical models are typically four or more layers deep and are composed of features in each stage that are learned by training on large-scale, labeled image data. Encouraged by the improved performance of deep networks, many recent studies have focused on comparing the internal representations of these models to the response properties of visual cortical neurons (e.g., see Khaligh-Razavi & Kriegeskorte, 2014). This approach can be insightful because the emergence of similar encoding features in models and neurons could imply similar computational strategies. For example, model units in the first layer behave like oriented filters similar to the properties of V1 neurons while many neurons in later stages of the models show tuning to boundary curvature similar to the properties of V4 neurons (Pospisil, Pasupathy, & Bair, 2016). Dissecting the underlying architecture of model units could provide insights into how these response properties arise in the brain. Furthermore, the detailed study of model units could promote the development of more targeted experiments on neurons, given the practical constraints that limit experimental time. However, similarities between the responses of model units and neurons to a few hundred stimuli should not be overinterpreted because there are no guarantees of model uniqueness: very different architectures could produce similar responses when probed with a small set of stimuli (see Kriegeskorte, 2015, for a detailed review). For example, the responses of a hypothetical neuron described earlier (Figure 4) are consistent both with the responses of a neuron selective for the color red and the responses of a neuron with position-dependent form selectivity. These two possibilities cannot be differentiated on the basis of the nine stimuli shown (Figure 4), and detailed characterization of RF position dependence and other properties will be needed. Thus, the discovery of

representational similarity between models and neurons calls for more experiments with more diverse stimuli to ask how model units and real neurons may diverge. Such a divergence may help to constrain the number and types of candidate models.

Finally, most models of visual form processing, including deep neural networks, are envisioned as hierarchical feedforward pathways. This model design choice is partly to simplify the task of building a model, but it is also motivated by findings of behavioral studies and ERP recordings in humans which suggest that base category discrimination (e.g., animal versus non-animal) can be achieved ~120 ms after stimulus onset (Kirchner & Thorpe, 2006; Thorpe, Fize, & Marlot, 1996), using the earliest neural signals in IT. Nevertheless, given that perception relies heavily on contextual information and on learning and experience, recurrence and feedback signals play a critical role. A complete and accurate model of form processing must therefore include these influences.

Moving Forward

Decades of studies in human and non-human primates have revealed a great deal about the neural basis of visual shape and object perception. We now have concrete models for visual processing in the earliest stages, conceptual models for processing in the intermediate stages, and we are beginning to discover how the visual system tackles problems like occlusion. These discoveries have primarily come from rigorous experiments that correlate neuronal responses to stimuli presented within the RF. So, rather than abandoning this strategy, we need to augment it. We need to combine the careful characterization of RF position and basic RF tuning properties of visual neurons with the study of neuronal responses to tens of thousands of novel and diverse stimuli, ranging from parameterized artificial stimuli to isolated natural objects and entire visual scenes. We need to expand our stimulus repertoire to include naturalistic stimulus features, such as surface shading, blur, and 3D form. Given that movement cues are also important for object segmentation, it will be important to investigate how dynamic shape stimuli are encoded in the object-processing pathway. These data could then constrain working models. In return, model predictions will direct the design of new experiments and stimuli by constraining plausible hypotheses for underlying computations.

A major bottleneck to progress has been the constraint of experimental time, especially given that a typical neurophysiological recording session in the awake, behaving monkey lasts 4–6 hours, allowing an experimenter to probe responses to hundreds, but not thousands of visual stimuli. As such, a key technological advance would be the ability to study the same neurons over days, weeks, and months. This would provide experimentalists the unprecedented opportunity to study the responses of neurons to many stimuli, both during passive fixation and under a variety of behavioral conditions and contexts. This technology would also facilitate the longitudinal tracking of neuronal responses during the course of visual experience. These richer datasets will help us to reveal how the primate visual cortex achieves the encoding, segmentation, and

perception of visual objects and scenes. These data will also illuminate how experience shapes visual perception, another important frontier.

Acknowledgments

References

Adelson, E. H., & Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 3–20). Cambridge, MA: MIT Press.

Albrecht, D. G., De Valois, R. L., & Thorell, L. G. (1980). Visual cortical neurons: Are bars or gratings the optimal stimuli? *Science*, *207*(4426), 88–90.

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, *10*(10), 1313–1321.

Asada, H., & Brady, M. (1984, January). The curvature primal sketch. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, *8*(1), 2–14.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.

Bair, W., Popovkina, D., De, A., & Pasupathy, A. (2015). *Modeling shape representation in area V4*. Workshop at MODVIS, May 13–15; St. Pete Beach, FL. Purdue ePubs. **https:// docs.lib.purdue.edu/modvis/2015/session02/7/**.

Bakin, J. S., Nakayama, K., & Gilbert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *Journal of Neuroscience*, *20*(21), 8188–8198.

Bauer, R., & Heinze, S. (2002). Contour integration in striate cortex. Classic cell responses or cooperative selection? *Experimental Brain Research*, *147*(2), 145–152.

Besl, J., & Jain, R. (1985). Three-dimensional object recognition. *Computing Surveys*, *17*, 75–145.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Blakemore, C., & Tobin, E. A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, *15*(4), 439–440.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, 7(8), 880–886.

Bushnell, B. N., Harding, P. J., Kosai, Y., & Pasupathy, A. (2011). Partial occlusion modulates contour-based shape encoding in primate area V4. *Journal of Neuroscience*, *31*(11), 4012–4024.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology*, *98*(3), 1733–1750.

Cavanagh, P. (1991). What's up in top-down processing. In A. Gorei (Ed.), *Representations of vision* (pp. 295–304). Cambridge: Cambridge University Press.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2547–2556.

Clowes, M. B. (1971). On seeing things. *Artificial Intelligence*, *17*, 79–116.

Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, *17*(2), 140–147.

Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, *97*, 4310–4326.

Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*(8), 2051–2062.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Dumoulin, S. O., Dakin, S. C., & Hess, R. F. (2008). Sparsely distributed contours dominate extra-striate responses to complex scenes. *Neuroimage*, *42*(2), 890–901.

Elder, J. H., & Zucker, S. W. (1998). Evidence for boundary-specific grouping. *Vision Research*, *38*, 143–152.

El-Shamayleh, Y., & Pasupathy, A. (2016). Contour Curvature As an Invariant Code for Objects in Visual Area V4. *Journal of Neuroscience*, *36*(20), 5532–5543.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field." *Vision Research*, *33*(2), 173–193.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*(6402), 343–346.

Fyall, A., El-Shamayleh, Y., Choi, H., Shea-Brown, E. T., & Pasupathy, A. (2017). **Dynamic representation of partially occluded objects in primate prefrontal and visual cortex**. *Elife*, 6, pii, e25784.

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, *259*(5091), 100–103.

Gallivan, J. P., Cant, J. S., Goodale, M. A., & Flanagan, J. R. (2014). Representation of object weight in human ventral visual cortex. *Current Biology*, *24*(16), 1866–1873.

Gheorghiu, E., & Kingdom, F. A. (2006). Luminance-contrast properties of contour shape processing revealed through the shape-frequency after-effect. *Vision Research*, *46*(21), 3603–3615.

Gheorghiu, E., & Kingdom, F. A. (2007). The spatial feature underlying the shape frequency and shape-amplitude after-effects. *Vision Research*, *47*(6), 834–844.

Gheorghiu, E., Kingdom, F. A., & Petkov, N. (2014). Contextual modulation as de-texturizer. *Vision Research*, *104*, 12–23.

Gross, C. G., & De Schonen, S. (1992). Representation of visual stimuli in inferior temporal cortex. *Philosophical Transaction of The Royal Society of London B: Biological Sciences*, *335*, 3–10.

Guzmán, A. (1968). Decomposition of a visual scene into three-dimensional bodies. In *Proceedings of the December 9–11, 1968, fall joint computer conference, part I* (AFIPS '68 (Fall, part I)). ACM, New York, USA, 291–304.

Hegde, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience*, *20*(5), RC61.

Helmholtz, H. (1910). *Treatise on physiological optics*. New York: Dover.

Hikosaka, K. (1999). Tolerances of responses to visual patterns in neurons of the posterior inferotemporal cortex in the macaque against changing stimulus size and orientation, and deleting patterns. *Behavioral Brain Research*, *100*, 67–76.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*(1), 215–243.

Huffman, D. A. (1971). Impossible objects as nonsense sentences. *Machine Intelligence*, *5*, 295–323.

Hung, C. C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, *74*(6), 1099–1113.

Ito, M., & Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *Journal of Neuroscience*, *24*(13), 3313–3324.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, *73*(1), 218–226.

Jolicoeur, P., Ullman, S., & Mackay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, *14*(2), 129–140.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*(4), 843–856.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience*, *23*(7), 3016–3027.

Kayaert, G., Biederman, I., & Vogels, R. (2005a). Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cerebral Cortex*, *15*(9), 1308–1321.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005b). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, *22*(1), 212–224.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), e1003915.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309.

Kim, T., Bair, W., & Pasupathy, A. (2017). Neural responses to shape and texture stimuli in macaque area V4. *Journal of Vision*, *17*(10), 291.

Kingdom, F., & Prins, N. (2005). Different mechanisms encode the shapes of contours and contour-textures. *Journal of Vision*, *5*(8), 463.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776.

Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961–980.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*(3), 856–867.

Konorski, J. (1967). *Integrative activity of the brain*. Chicago: University of Chicago Press.

Kosai, Y., El-Shamayleh, Y., Fyall, A. M., & Pasupathy, A. (2014). The role of visual area V4 in the discrimination of partially occluded shapes. *Journal of Neuroscience*, *34*(25), 8570–8584.

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modelling biological vision and brain information processing. *Annual Review of Vision Science*, *1*, 417–446.

Leung, T., & Malik, J. (1998). Contour continuity in region based image segmentation. *Computer Vision—ECCV'98*, 544–559.

Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature*, *387*(6628), 73–76.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Marimont, D. H. (1984). A representation for image curves. *AAAI Proceedings*, *84*, 237–242.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology*, *283*, 101–120.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology*, *283*, 53–77.

Namima, T., & Pasupathy, A. (2016). *Neural responses in the inferior temporal cortex to partially occluded and occluding stimuli*. Society for Neuroscience Meeting Abstracts, San Diego, CA.

Nelson, J. I., & Frost, B. J. (1985). Intracortical facilitation among co-oriented, coaxially aligned simple cells in cat striate cortex. *Experimental Brain Research*, *61*(1), 54–61.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. *Visual Neuroscience*, *16*(1), 15–34.

Oleskiw, T. D., Pasupathy, A., & Bair, W. (2014). Spectral receptive fields do not explain tuning for boundary curvature in V4. *Journal of Neurophysiology*, *112*, 2114–2122.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, *8*(5), 379–391.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, *86*(5), 2505–2519.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*(12), 1332–1338.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3), 329–342.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, *391*(6667), 580–584.

Ponce, C. R., Hartmann, T. S., & Livingstone, M. S. (2017). End-stopping predicts curvature tuning along the ventral stream. *Journal of Neuroscience*, *37*(3), 648–659.

Pospisil, D., Pasupathy, A., & Bair, W. (2016). **Comparing the brain's representation of shape to that of a deep convolutional neural network**. In J. Suzuki, T. Nakano, & H. Hess (Eds.), *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)* (BICT'15) (pp. 516–523). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium.

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not "grandmother-cell" coding in the medial temporal lobe. *Trends in Cognitive Science*, *12*(3), 87–91.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, *3*(Suppl.), 1199–1204.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*, 203–227.

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics*, *73*(8), 2542–2572.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, *30*, 339–366.

Rust, N. C., & Dicarlo, J. J. (2010). Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *Journal of Neuroscience*, *30*(39), 12978–12995.

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current Opinion in Neurobiology*, *20*(3), 382–388.

Sajda, P., & Finkel, L. H. (1995). Intermediate-level visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience*, 7, 267–291.

Sáry, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, *260*, 995–997.

Sary, G., Vogels, R., Kovacs, G., & Orban, G. A. (1995). Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. *Journal of Neurophysiology*, *73*(4), 1341–1354.

Schmidt, K. E., Goebel, R., Lowel, S., & Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *European Journal of Neuroscience*, *9*(5), 1083–1089.

Schwartz, E. L., Desimone, R., Albright, T. D., & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(18), 5776–5778.

Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, *11*(5), 384–399.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670–674.

Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Verri, A., & Yuille, A. (1986). **Perspective projection invariants**. In *Artificial intelligence lab*. Cambridge, MA: MIT Press.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P.H. Winston (Ed.), *Psychology of Computer Vision* (pp. 19–91). New York: McGraw-Hill.

Watt, R. J., & Andrews, D. P. (1982). Contour curvature analysis: Hyperacuities in the discrimination of detailed shape. *Vision Research*, *22*(4), 449–460.

Wertheimer, G. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 71–88). London: Routledge & Kegan Paul.

Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, *37*(17), 2325–2330.

Yue, X., Pourladian, I. S., Tootell, R. B., & Ungerleider, L. G. (2014). Curvature-processing network in macaque visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), 3467–3475.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Science*, *10*(7), 301–308.

Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area V2. *Neuron*, *47*, 143–153.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*, 6594–6611.

Anitha Pasupathy

Department of Biological Structure, University of Washington

Yasmine El-Shamayleh

Department of Physiology and Biophysics, University of Washington

Dina V. Popovkina

Department of Neuroscience, University of Washington