Vision: Brain Mechanisms

ROBERT SHAPLEY and NAVA RUBIN
New York University

I. Introduction
II. V1–Vn and Retinotopy
III. Segmentation
IV. MT and Motion
V. Color in Ventral Occipital Brain Areas
VI. Object Recognition in LOC and Ventral Occipital–Temporal Cortex
VII. Conclusion

GLOSSARY

magnetic resonance imaging Imaging of the internal structure of body tissue by the observation of magnetic resonances induced in them.

occipital cortex Cerebral cortex belonging to the occiput or back part of the head.

retinotopic map A map that preserves the spatial relations of the sensory receptors of the retina.

segmentation The process of division into segments.

stereopsis The ability to perceive depth and relief by stereoscopic vision.

temporal cortex The lowest lobe of the brain lying below the Sylvian fissure.

Vision appears to be a simple matter, which does not require any mental effort: We open our eyes and see. However, the subjective ease of seeing masks a tremendous amount of brain processing required to allow us to perceive the world through the visual sense. The most direct evidence for this is the large portions of cortex that are devoted to visual processing. The entire occipital lobe, as well as parts of the temporal and parietal lobe, is concerned with interpreting the image impinging on the retina. The operations done by the numerous visual cortical areas are understood only partially. Part of the difficulty stems from the lack of good theoretical understanding of visual processing. For example, research on artificial vision has not been able to produce “seeing machines” that would perform visual tasks that humans perform effortlessly, such as navigating through a crowded room or recognizing a familiar face. Better theoretical understanding would have allowed to map the various brain processes onto known computational stages of visual processing. Instead, the study of the brain mechanisms of vision and research on the computational aspects of vision have progressed in parallel, each field drawing inspiration from findings in the other. In this article, we survey some of the major findings on the function of the various areas of the visual cortex.

I. INTRODUCTION

The visual cortex takes signals about the retinal images that are relayed to it through the lateral geniculate nuclei (LGN) and seeks to recognize objects and navigate in the world based on vision. Visual signals are processed in V1 (also known as the striate cortex) and V2, the primary and secondary visual cortical areas. From V1 and V2, signals are sent to multiple visual areas anterior and lateral to these early visual areas in the brain. Some of these extrastriate visual areas seem to be very specialized for certain aspects of the visual image, such as motion or color. How these different areas contribute to visual perception has become much clearer in recent years.

Knowledge about the visual system’s functional anatomy has expanded tremendously in the past decade through the use of brain imaging techniques.
Optical imaging of the upper layers of the cerebral cortex in nonhuman primates, using activity-dependent changes in reflectance, can be used to study structure on the scale of several square millimeters. Magnetic resonance imaging (MRI) both in nonhuman primates and in human subjects allows scientists to study the structure of the visual system noninvasively. Functional MRI (fMRI) enables researchers to study changes in the brain’s blood oxygenation that are caused by variations in neuronal activity, thus enabling noninvasive functional mapping of human brain activity. Most of the results we cite about human brain activity correlated with perception are derived from fMRI experiments. fMRI has lower spatial resolution than optical imaging, but it allows the experimenter to study brain volumes from 100 to 1000 ml and reflects activity not only from the cortex’s superficial layers but also through the depth of the cortex.

It appears that the brain must find meaningful patterns in order to compute visual attributes correctly as well as to be able to place these attributes in and around perceived objects in a scene organization that is consistent with the visual image. This view is based on experimental observations by, among others, Hans Wallach that visual attributes of perceived objects, such as color, shape, and motion, are linked, and they all need to be computed by the brain. Thus, we are compelled by the evidence from human perception, from fMRI, and from physiological experiments on animals to conceive of perception as an active process in which visual image information is combined with memory to compute “why things look as they do.” This is most easily understood when visual images are ambiguous—for example, when selected visual images produce figure/ground reversals or binocular rivalry. Then it becomes clear that there are multiple states of brain activity corresponding to multiple “interpretations” that are consistent with the visual image. What one sees in this situation changes with time and the state of the brain and must involve time variations in the computations that the visual system is doing based on the visual image.

The first neural computation we consider is the mapping of the visual world onto the cortex in V1 and adjacent visual cortex.

II. V1–Vn AND RETINOTOPY

In order to understand appearance, we must account for the location of perceived objects in the world. Perceived location depends on the mapping from location in the world to location on the cortex. A schematic diagram of human visual cortex is depicted in Fig. 1. This diagram emphasizes the visual field...
mapping onto different brain regions, V1–V3. Human primary visual cortex V1 is located deep in the calcarine fissure, and it extends slightly onto the cortex around the posterior pole. Most of the visual cortex located on the posterior surface of the human cerebral cortex is in fact extrastriate cortex; that is, V2, V3, V4, and up to Vn. V1 is continuous above and below the calcarine fissure and forms a continuous boundary with V2, which surrounds it completely. V2 is composed of two disjoint cortical regions, upper and lower V2, as is the case for V3 and V4. From lesion studies we know that the upper half of V1 receives LGN input that is looking only at the lower visual field, whereas lower V1 is looking at the upper visual field. This is true also of V2, V3, etc. It is well-known that the early, retinotopic areas V1–V3 represent the visual field contralateral to the cerebral hemisphere in which they are located. Thus, a lesion in the upper region of V2 will cause a visual field defect only in the lower contralateral quadrant of the visual field.

The map of the visual field in V1 was demonstrated elegantly in the monkey visual cortex by Roger Tootell and colleagues. They used the 2-deoxyglucose activity-labeling technique as illustrated in Fig. 2. The stimulus used was a “ring and ray” pattern presented for approximately 1 hr while the radioactively labeled 2-deoxyglucose circulated through the bloodstream of the monkey. Active cells picked up the label and appeared dark in the autoradiograph in Fig. 2. The center of the ring and ray pattern was placed where the foveae of the eyes were looking, and the pattern in the contralateral hemifield (Fig. 2) activated the neurons in the hemisphere that is imaged in the figure. The cortex was flattened physically to enable the flat two-dimensional (2D) image of the cortical map to be made. The mapping of the ring and ray pattern is smooth and relatively precise onto V1 cortex. The circular rings were spaced at distances that grew geometrically larger, and their map onto V1 is approximately onto lines that are approximately equidistant from each other. This means the mapping is approximately logarithmic, as has often been noted. The stimulus rays also map into approximately straight lines that are separated from each other by a fixed distance. The vertical meridian is mapped into the V1/V2 border, whereas the horizontal meridian divides V1 in half, with the upper field mapped to lower V1 and vice versa. Figure 2 is a succinct summary of the main facts about retinotopic mapping in primate V1 cortex.

Retinotopic mapping in the human V1 cortex, and in extrastriate visual areas as well, has been demonstrated with fMRI. The technique developed by Engel and colleagues is shown in Fig. 3. The basic idea is to use a periodic stimulus, such as an expanding ring or a rotating wedge. Then neurons that are functionally connected to the retinal region excited by the stimulus will be activated only during one phase of the stimulus cycle. Location will be encoded in the phase of the fMRI response with respect to the stimulus cycle. This technique is therefore called phase mapping. The phase of fMRI response with respect to the expanding ring’s cycle encodes radial distance from the point of expansion (usually the fovea). The phase of fMRI response with respect to the rotating wedge encodes the azimuth of the response location in the frontoparallel plane. From Fig. 2, one might expect that the active regions of V1 excited by a given azimuth should lie along straight lines in a flattened map of the cortex since this is what was observed in the 2-deoxyglucose maps. Also, one should expect that regions of a given retinal eccentricity in flattened-cortex phase maps should lie along straight lines that are approximately perpendicular to the cortical projection of the vertical meridian, as demonstrated by the monkey 2-deoxyglucose data.

Data on the retinotopic map in human visual cortex are given in Fig. 4. Here, the representation of the visual cortex as a flat sheet was not achieved by physical flattening as in the monkey experiments since these were in vivo, noninvasive measurements and the cortex was not available for physical flattening. Rather, the cortex was virtually flattened by computer image processing of the 3D image of the cortex obtained from MRI measurements, using an algorithm that preserved local distances. Separate eccentricity and azimuth maps were measured to create the mapping in the figure. As shown, only azimuth is coded by color. The V1 map is consistent with the monkey V1 map in terms of both eccentricity and azimuth. The map indicates that human V1 also represents only the contralateral hemifield, consistent with the effects of lesions on human visual cortex.

The results summarized in Fig. 4 also reveal important facts about the mapping of visual space in extrastriate cortex; for example, V2 cortex is adjacent to V1. Also, there is a retinotopic map in V2. The V1 and V2 maps are separated by the representation of the vertical meridian in each cortical area, so the vertical meridian projection serves as a marker for the V1/V2 boundary. This also means that V2, like V1, is only receiving visual input from the contralateral visual hemifield. Furthermore, the boundary between V2 and V3, a distinct visual cortex region more lateral than V2,
is the projection of the horizontal meridian. It is important that the projection to V2 (and also V3) of visual image regions just above and just below the horizontal meridian projects to cortical subregions that are widely separated. This means that local circuits that may compute spatial linking between local stimuli encounter a wide chasm in extrastriate cortex at the horizontal meridian; therefore, it is likely that visual signals from upper and lower visual fields are processed separately in extrastriate cortex. This does not apply to V1, in which neurons are mapped smoothly across the horizontal meridian. This can be
Figure 3  Diagram of the stimuli used in the phase mapping technique for fMRI. Radial checkerboards are the basic stimuli. In A, used for eccentricity mapping, rings of the underlying checkerboards are propagating inwards to a focus of contraction. In B, a rotating wedge of checkerboard is shown, and the wedge’s azimuth varies with time [reproduced with permission from Wandell (1999). Computational neuroimaging of human visual cortex. *Annu. Rev. Neurosci.* 22, 145–173.]

Figure 4  Retinotopic mapping in the human cortex measured with fMRI. Phase mapping as in Fig. 3 was used. In this example, only azimuth is mapped, as indicated in the inset, which gives the key to angle. The map is onto a mathematically flattened cortex and shows the transitions between retinotopic areas along the meridians [reproduced with permission from Wandell (1999). Computational neuroimaging of human visual cortex. *Annu. Rev. Neurosci.* 22, 145–173.]
seen best in Fig. 1. The continuous mapping in V1 and the discontinuous mapping in V2–V4 are likely to have functional consequences for vision.

In the human visual cortex, mapping experiments such as those in Figs. 3 and 4 indicate that the right and left hemifields are kept distinct through several extrastriate regions, including V1–V4. Signals from the right and left visual fields are not combined until one reaches the lateral occipital cortex (LOC), an area discussed later. Therefore, there are two “seams” in the mapping of the visual world onto early visual cortical processing areas from V1–V4 and possibly beyond. One seam is the vertical meridian that separates the right and left visual fields. The second seam is the horizontal meridian that separates the upper from the lower visual fields. Our visual perception appears to be seamless, subjectively. The unified worldview we have must be the result of a combination of visual signals across these seams in the mapping in a manner that remains to be discovered. The interhemispheric mapping of the visual world in LOC and anterior visual areas is likely to be significant in producing the seamless appearance of the visual world.

III. SEGMENTATION

There is a great transformation that takes place in visual perception between the analog representation of the visual image in V1 and the symbolic representation of surfaces and objects as they appear to us. In V1, as in the retina, there seems to be an analog map of the brightness and colors in the image. However, when we perceive a meaningful natural image, we see a countable number of surfaces and discrete objects that are segregated from the background and from other objects. There are probably many stages of this transformation, but one stage of this process is known to be of great importance: visual segmentation. Segmentation is a process that involves the grouping of object parts together in a separated figure that is distinct from what is around or near it. Often, this process must group together parts of an object in an image that are separated from each other by an occluding object in front of the object to be segmented. Segmentation is necessary for correct organization of a scene because it allows the organizing process to know which are the surfaces that must be ordered in depth. Also, the action of segmentation could contribute to the scene organization computation because it needs to resolve occlusions that are a strong clue to depth order and therefore, scene organization.

As with many brain computations, we can understand segmentation better by observing its action when it deals with an exceptionally difficult task. Usually, segmentation is so smooth and efficient that we are unaware it is happening. However, for certain special visual images, the segmentation process becomes evident. This is the reason for the fascination with these special images, the so-called illusory contours (ICs). An example of such a visual image is shown in Fig. 5, which presents an image that is referred to as a Kanizsa triangle, named after the Italian Gestalt psychologist Gaetano Kanizsa. In Fig. 5, the perception of a bright white triangle is very strong, but if one scrutinizes the boundaries of the triangle it becomes evident that there is no difference in the amount of light coming to the eye from the regions inside and outside the perceived triangle. However, we see the inside as a bright surface segmented from its background by sharp contours along the boundary of the triangle. In this sense, the boundary between inside and outside the triangle is an illusory contour. This image is a classic example in favor of the basic concept of the Gestalt psychologists—that the brain is searching for meaningful patterns. In this case, the brain manufactures a perceptual triangle from fragmentary information because a meaningful pattern, an occluding triangle, is consistent with the available image information even though other perceptions are possible. It is reasonable...
to believe that the segmentation computations that the visual system performs on these exceptional Kanizsa images are the same as for more typical images.

From electrophysiological single-cell recordings in awake monkeys, von der Heydt, Peterhans, and colleagues found that Kanizsa-type images and other illusory contour images could excite spike activity in neurons in early visual cortex. An example of one of their experiments is shown in Fig. 6. They recorded from a single neuron in area V2 of the macaque visual cortex. The neuron’s activity is depicted as a raster plot of spikes vs time during repeated presentations of the stimuli. The neuron responds with excitation to a luminance contour that crosses its receptive field. When an illusory contour (as perceived by us) crosses its receptive field, the cell produces a slightly delayed excitatory response resembling the response to a real contour. As a control that the response is not merely a weak response to the remote features of the IC stimulus, the investigators made a small image manipulation (closing the inducing boundary) and this eliminated the neuron’s response. Von der Heydt and Peterhans also performed several quantitative studies on these IC-responsive V2 neurons, particularly the measurement of the orientation tuning for illusory and real contours on the same population of V2 neurons; they found that real and illusory contours produced similar orientation tuning in IC-responsive neurons in V2. Thus, these neurons seem to be a candidate neural substrate for illusory contour perception. There have also been reports of IC responses in neurons in V1. This is a controversial issue since von der Heydt, Peterhans, and colleagues maintained that they observed very few V1 neurons that produced IC responses. The discrepancy may occur in part perhaps because of the use of different stimuli and in part because of different views of what constitutes an illusory contour. For present purposes, it is enough to conclude that IC responses can be observed in retinotopic areas in the monkey’s brain—areas that are traditionally thought of as stimulus driven.

The connection of the monkey V2 neurons with IC perception in humans needs to be established more firmly. We do not know the nature or quality of IC

Figure 6  Responses of three macaque V2 neurons to ICs. Stimuli are shown in the insets next to raster plots of neuronal responses, in which each white dot is a nerve impulse. The raster shows multiple repeats of the same stimulus swept over the receptive field repeatedly. The crosses indicate the fixation point, and ellipses mark the classical receptive field of each neuron. In each case, row 1 is a conventional luminance stimulus, row 2 is the IC stimulus, and row 3 is a control stimulus, where a small image manipulation destroys the IC percept. Row 4 is a blank control to show the spontaneous firing rate [reproduced with permission from Peterhans and von der Heydt (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. J. Neurosci. 9, 1749–1763.]
perception in these animals because there are insufficient animal data from rigorous experiments to test for IC perception. Once we know whether monkeys can exhibit behavior that proves they see illusory contours in Kanizsa-type images, further experiments would be necessary to determine whether or not the V2 neurons have the same sort of parameter dependence on size, contrast, and retinal location as the behavior. How do the V2 neurons respond to an illusory contour that crosses the horizontal meridian? Humans respond as well to such illusory contours as they do to ICs that do not cross the horizontal or meridian. The reader can observe this for himself or herself by fixating the middle of the Kanizsa triangle in Fig. 5 and observing the robust lateral ICs that traverse the horizontal meridian in his or her visual field. However, in V2, there is a marked separation between neurons that represent the visual field just above and just below the meridian. Therefore, one might expect some deleterious effect on IC responses for meridian-crossing ICs in V2 neurons. If that decrease in IC sensitivity in V2 neurons were observed, it might cast doubt on the role of V2 alone in IC perception. Moreover, as discussed later, the human data point to other brain areas as the major processing sites for ICs. The monkey results could be interpreted to indicate that similar IC-related activity is occurring in human V2, but fMRI and other techniques used on humans are too insensitive to measure it. A second possibility is that the V2 activity seen in monkeys is related to IC perception but is not the central mechanism involved in the percept. Another possibility is that human and monkey perception and neural mechanisms are fundamentally different at this midlevel stage of visual processing.

Human responses to ICs have been measured with fMRI techniques. Most fMRI studies have involved the measurement of the activation of Kanizsa squares or diamonds compared with the same pacman-shaped inducers rotated outwards or all in the same direction. Pacman is a term that refers to the shape of an agent in a video game from the early 1980s. The shape of “Pacman” was exactly the same as the cut-off circles used by Kanizsa in the IC figures he originated. Early studies by Hirsch and colleagues and by Ffytche and Zeki found that there was activation of the occipital cortex lateral to V1 by Kanizsa-type figures, but they could not pinpoint the cortical location because the IC experiments were not combined with retinotopic mapping. Therefore, these studies established that signals related to segmentation were present in occipital cortex, but further work was needed to determine more precise localization.

The extensive research of Janine Mendola and colleagues established that IC-related signals were observed in retinotopic area V3 and also in LOC, the lateral occipital area previously discovered by Malach and colleagues. Figure 7 is an fMRI image indicating the large region of cortical activation evoked by the Kanizsa diamonds used as stimuli in Mendola’s study. The early retinotopic areas V1 and V2 did not produce statistically significant activation. Mendola and colleagues also used different inducers for ICs, such as aligned line endings, and found a similar pattern of brain activation in V3 and LOC. These results are important in implicating extrastriate cortex in the process of visual segmentation in humans.

Although ICs are often chosen for studying visual segmentation, other visual phenomena can lead us to an understanding of this very important process. The assignment of an image region as figure or ground, is one such phenomenon. As Edgar Rubin, the famous perceptual psychologist, pointed out in 1921, such assignment is automatic and inescapable. However, ambiguous figures exist in which figure and ground assignments flip back and forth, and perception changes when that happens. The familiar face/vase figure from E. Rubin is the most reproduced example, but there are other examples from Rubin that illustrate the consequences of figure/ground assignments even better. One of these is the Maltese cross figure shown in Fig. 8. This example is described in Koffka’s 1935 book but not depicted in it. The diamond-shaped arms of the cross can appear grouped in fours, with a vertical and horizontal pair grouped together as figures in front (resembling a propeller in shape) and then two diagonal pairs grouped together as figures in front (and the vertical–horizontal pairs are then in back). The contrasts in the figure are arranged such that the vertical–horizontal propeller shape looks darker in front than it does when it is perceived “in back,” appearing as a light gray diamond behind the white tilted propeller. This is because of the enhanced effect of brightness contrast across borders that define a figure and on the regions to which such borders are attached, as Rubin noted. Similar effects can be seen in color. This is only one of many illustrations of the deep consequences of figure/ground assignment. For instance, another consequence of the importance of figure/ground is that people remember the shapes of figures, not grounds. Therefore, understanding the neural basis for this phenomenology is an important clue to the function of the visual system.

Figure/ground assignment is a special case of a more general problem in vision—the assignment of border
assignment of a region as figure or ground is all one has to do if there is only one figure surrounded by the background. However, if there are many figures, and if one is in front of another so that it partly occludes the shape of the second figure in the visual image, then the visual system must decide on the basis of image information which surface is in front along the boundary between the two figures in the image. Briefly, the brain has to decide which figural region owns the border between them (i.e., the front surface). Assignment of border ownership is a problem that must be solved for almost every visual image.

There have been only a few investigations of neural mechanisms for border ownership and figure/ground assignments. Zhou and colleagues studied single cells in V1 and V2 cortex of macaque monkeys. By keeping local edge contrast the same but varying the global stimulus so that different regions own the boundary between them perceptually, Zhou et al. tested sensitivity to border ownership in single cortical neurons. The experimental design and results in an archetypal border ownership cell are shown in Fig. 9. A substantial number of cells such as this are encountered in monkey V2 cortex. Baylis and Driver reported that many neurons in monkey inferotemporal (IT) cortex respond differentially to figure or ground; thus, these
also must reflect signals about border ownership. Since the IT cortex is supposed to be involved in object recognition, it is reasonable that neurons in this area should be affected by border ownership that is necessary for accurate object recognition in the real world.

Studies by Kleinschmidt and colleagues of figure/ground reversals in human cortex using fMRI revealed activation over a number of areas in occipital, temporal, parietal, and even frontal cortex. The involvement of temporal, parietal, and frontal cortex seems to imply that activation of top-down influences from “high-level” cortical areas could be necessary for figure/ground reversal of border ownership. However, as in the case of ICs, it is also possible that there may be signals associated with figure/ground assignment in “early” retinotopic areas, such as V1 or V2, that are undetectable with fMRI.

An important part of segmentation in human visual perception is the phenomenon of amodal completion—that is, completion and grouping together of the parts of a partially occluded object that are visible. This completion process is crucial for normal object perception in the real world. Evidence that amodal completion affects the firing rates of V1 neurons in macaque V1 was obtained by Y. Sugita by manipulating apparent occlusion using stereopsis (Fig. 10). Only a small fraction of V1 neurons were affected by amodal completion, but it is still a significant result.

IV. MT AND MOTION

Two of the strongest stimulus cues for segmentation are motion and color. This makes sense ecologically since it is unlikely that separate things will move together for any length of time, and similarity in color is associated with similarity in surface properties. The visual cortex appears to agree with this reasoning because it devotes a large amount of specialized cortical processing for motion and for color. We first consider motion processing in the middle temporal area MT (or V5). This brain area was initially defined by Zeki in his early work in the macaque extrastriate visual cortex. Zeki noted the large number of directionally selective cells in what he called V5, the middle temporal area later called MT by others. In human cortex, a motion-sensitive homolog to macaque MT was suspected from neuropsychological work on brain-damaged patients. Zihl and colleagues described a patient who had a lesion in dorsolateral occipital cortex and who had lost the ability to see motion. The motion-blind patient reinforced the concept of Zeki and others that there was a discrete brain region, assumed to include MT, that was obligatory for the perception of motion.

Functional imaging in normal human subjects, with visual stimulation by moving dots compared to stationary flashing dots, indicated that there was a discrete region in dorsolateral occipital cortex (lateral, dorsal, and anterior to LOC) that had most differential activity. Figure 11 shows the location bilaterally of what is now assigned to be human MT (V5). This is shown in a drawing of a slice of brain, not a flat map.

Figure 9 This is a representative neuron from V2 cortex in a macaque monkey that reveals sensitivity to border ownership. The same polarity edge border is in the pairs A and B and C and D, but the cell responds most strongly when the border “belongs” to a figural region down and to the left of the cell’s receptive field (black line). Each vertical stroke in the raster plots stands for a nerve impulse [reproduced with permission from Zhou et al. (2000). Coding of border ownership in monkey visual cortex. J. Neurosci. 20, 6594–6611.]
Many other groups have confirmed this localization of MT in human cortex. The location of MT in the flat map representation of fMRI images of occipital cortex is given in Fig. 7 (top), in which MT’s location is mapped and the locations of the retinotopically mapped areas are also indicated.

Classic results on MT in monkey cortex tended to confirm the notion that MT neurons were devoted to extracting velocity (direction and speed) of contours or patterns over the large receptive fields of the MT neurons. A significant number of MT neurons are directionally selective when stimulated with drifting bars or grating patterns. Using plaid patterns similar to those that Hans Wallach introduced to study motion coherency and transparency in human perception, Tony Movshon and colleagues found that some MT neurons were tuned to the coherent motion direction of a plaid pattern, whereas other MT cells were only selective for the motion directions of the component gratings that summed to make the plaid. These two kinds of neurons were given the labels “pattern” and “component” neurons, respectively. The assignment of two separate names might be supposed to carry the implication that there are distinct classes of neurons in MT. However, this was not claimed. In fact, existing data tend to suggest that there are not two distinct types of neurons in MT but rather a continuum of pattern vs component selectivity. There is definitely a large group of “mixed” cells that are partially component and partially pattern selective.

Although many macaque MT cells respond to grating patterns and contours, most MT cells respond most vigorously to small dots or populations of dots moved through their receptive fields. Also, it was shown that MT neurons would respond to coherent motion of a proportion of dots embedded within random dot kinematograms, and that the coherency

Figure 10  Macaque V1 neurons that respond to amodal completion. The neuron responds to a long contour in its receptive field as in a. It responds to both eyes (b) and to the right eye alone (c) but not to the left eye alone (d). The interesting manipulation is in f–h. In f, the neuron does not respond to two unconnected segments. In g, it does not respond to the same two segments when they are perceived as being in front of a gray region. However, there is a response when the retinal disparity is such that the gray region is in front and occluding the two line segments (h) [reproduced with permission from Sugita (1999). Grouping of image fragments in primary visual cortex. Nature 401, 269–272.]
thresholds for individual MT neurons were similar to behavioral thresholds for the monkey. This work in Bill Newsome’s laboratory suggested that MT was involved in the perception of coherent motion in a definite direction. The work on cortical microstimulation in macaque MT cortex by Salzman and colleagues in Newsome’s laboratory showed that differential activation of subpopulations of macaque MT neurons could bias perceptual judgments of coherency. Indeed, when coherency was low, MT microstimulation could cause the monkey to behave as if it perceived the random dot kinematogram with no coherent dot motions to have significantly suprathreshold coherency. These results were consistent with the interpretation that MT activity determined the perceptual judgment of coherency and direction.

Braddick and colleagues studied human fMRI response to coherency of motion in random dot kinematograms, in analogy to the experiments done by Newsome’s group on macaque MT. Their experimental design was broader in that they were comparing the pattern of activation in different cortical areas in response to different kinds of stimuli that were designed to elicit responses to motion coherency and to form coherency. However, for the current discussion, we focus on the motion coherency results. Their stimuli for motion coherency were in essence the same as those used by Salzman et al. The comparison stimulus in their fMRI experiments was totally incoherent random motion of a field of dots. They found a widespread pattern of activation caused by motion coherency including MT (V5), the retinotopic area V3, as well as isolated regions of temporal and parietal cortex.

The role of human MT in motion perception has been examined in another kind of experiment the correlation of brain activation with the motion after-effect by Roger Tootell and colleagues. Prolonged viewing of motion in one direction leads to a familiar perceptual aftereffect when the stimulus motion ceases. Stationary objects appear to move in the opposite direction. Tootell et al.’s fMRI study of the motion aftereffect used expanding concentric rings as a motion stimulus and stationary concentric rings as a control. They found that at the termination of the motion stimulus, MT activation continued for another 20–30 sec. This prolonged afteractivation paralleled almost exactly the subjective perception of the motion aftereffect (apparent contraction of the concentric rings). Because there was no physical motion during
the aftereffect period, but there was subjective experience of motion counter to that seen during the stimulation period, this is a very strong result indicating that MT activation is correlated with the subjective experience of perceived motion. Other studies have confirmed this result, finding motion aftereffect activation in MT.

A number of recent studies indicate that MT cortex is not simply responding to motion direction or speed but, rather, may be playing a role in visual segmentation. For example, at the Salk Institute, Stoner and Albright, working with Ramachandran, initiated a series of studies combining experiments on human perception and on single cells in macaque MT that indicate that MT neurons are affected by form cues as well as by motion. They used plaid patterns and studied human perception of coherency as they changed the brightness at the intersections between the lines of the plaid. When the intersection brightness was not compatible with optical transparency of the perceived overlaying line, the probability of perceiving coherency was high and the probability of perceiving transparency was low. When the intersection brightness was compatible with transparency, the probability of perceiving motion transparency became greater.

This perceptual effect was also confirmed in macaque MT neurons, as shown in Fig. 12. The direction selectivity of neurons in MT was changed by the same manipulation of intersection brightness and hence of consistency with physical transparency. Thus, in macaque monkey MT there is evidence that individual neuron activity is determined by segmentation (or combination) of motion signals contingent on cues for transparency.

Another line of evidence that macaque MT neurons are responding to moving surfaces and not simply a velocity vector derives from experiments on the phenomenon of the kinetic depth effect (KDE). This is a classical perceptual phenomenon introduced and explored perceptually by Hans Wallach and one of his students in 1953. A visual image formed by the projection of a moving 3D object on a flat screen elicits a strong perception of 3D structure. The perceptual ability to see KDE is the most important means for judging 3D shape for objects that are too distant for a subject to use binocular stereopsis for depth and shape judgments (i.e., for distance > 5 m). Such depth from motion is combined with stereo for depth judgments at any distance. David Bradley and colleagues found that many macaque MT neurons

![Figure 12](image_url)

**Figure 12** The effect of static form cues on motion coherency in MT neurons in macaque visual cortex. Plaid patterns were moved through MT receptive fields. Each point in each graph represents a single neuron and the scatter plots represent the entire population of MT neurons studied. The graphs plot the calculated component index (CI), which estimates the amount of a neuron’s directional tuning curve that is explained by transparent (or “sliding”) motion as opposed to coherent plaid motion. The horizontal axis in each graph is the CI measured and calculated for stimuli that should be perceived as coherent because the static brightness of the intersection is not consistent physically with transparency. The vertical axis is for a case in which the intersection brightness is consistent with transparency. The tendency of the points to lie above the line indicates that when physical transparency is possible, motion transparency is perceived more often [reproduced with permission from Stoner and Albright (1992). Neural correlates of perceptual motion coherence. *Nature* 358, 412–414.]
were tuned for depth. This finding was confirmed and amplified by the later work of DeAngelis and coworkers, who showed that there was an orderly mapping of depth in MT cortex. For example, some MT neurons would be responsive only when the moving target was in front of the fixation plane. Bradley then showed that such depth-sensitive MT neurons were not only responsive to depth based on stereopsis but also signaled depth from KDE (Fig. 13). The KDE stimuli used by Bradley et al. were random dot kinematograms like those used in the 1970s for studying KDE perception by Shimon Ullman and others. The moving dot flow fields are ambiguous stimuli that could be perceived in a number of ways. Bradley et al. demonstrated that the neurons responded to KDE when the monkey perceived the KDE as creating a surface at the appropriate depth for the neuron’s depth tuning. This is an important result that indicates the perceptual sophistication of MT neurons. However, in this case the homology with human MT may be less close than in other interspecies comparisons. In an experiment on KDE in humans, Paradis and colleagues used fMRI. It revealed that many areas of the brain responded differentially to KDE, but MT was not among those KDE regions. Rather, V3 and regions in parietal and temporal cortex were activated more by KDE than by random dot motion. These authors speculated that in humans, the higher form-related functions of MT were taken over by parietal cortex.

Another important result more directly indicates the linkage between form and motion in human MT. Rainer Goebel and colleagues, studying fMRI activation in humans by different apparent motion stimuli, found that apparent motion of organized Kanizsa squares elicited much more fMRI activation than did apparent motion of shuffled pacmen. This result also corresponds with perception: There is a much stronger percept of apparent motion with moving Kanizsa squares than with the same pacmen pointing outwards. These results suggest that the nature of human MT’s response to moving visual surfaces needs to be further investigated. For instance, experiments such as those by Bradley combining motion flow fields with variations in depth organization should be attempted on human MT and compared with Bradley’s results in macaques.

V. COLOR IN VENTRAL OCCIPITAL BRAIN AREAS

The idea that there is a separate localized area in human extrastriate cortex that is specialized for mediating color perception is derived from the phenomenon of cerebral achromatopsia (cortical color blindness). Achromatopsia is usually caused by stroke lesions in ventral occipital–temporal cortex. It is a variable clinical syndrome, with the common feature that patients cannot pass tests indicating they can perceive colors normally. The critical test is the
Farnsworth–Munsell 100 hue test, which involves color arrangement. Normal humans perform this test very accurately and achromatopsics can be at chance in this test of arranging hues in an orderly sequence. Failure on the 100 hue test correlates very well with subjective reports that the patients cannot tell what color they are looking at. Because this neurological syndrome is usually accompanied by lesions in a particular region of cortex, there is strong suspicion that this extrastriate cortex subregion is the color-specialist area, a view stated forcefully by Semir Zeki. However, we review and modify this concept of a color module in the following discussion.

Human imaging studies comparing activation by color vs achromatic patterns are quite consistent in identifying to a ventral occipital–temporal region as the color area. Originally, this region was dubbed V4 by Lueck and colleagues, who used positron emission tomography (PET) to find the regions of the brain that were preferentially activated by color. This was later confirmed with fMRI by Zeki and colleagues. The naming of the color region as V4 cortex was done in analogy with Zeki’s previous studies on macaque cortex, in which he proposed that macaque V4 was specialized for color processing. This proposal of color specialization in macaque V4 has been challenged many times and is still controversial, especially due to the research of Heywood and Cowey, who found that lesions in V4 did not have a major effect on color discriminations in monkeys. Hadjikhani, Tootell, and colleagues mapped the color area in the same manner as Zeki and colleagues, but they also measured retinotopic mapping in the same subjects to find out if the color area Zeki and colleagues discovered was indeed V4 in the retinotopic framework (Fig. 14). They found that it was quite anterior to V4 as measured with retinotopic mapping. There is no disagreement on the location of the color-prefering area, just on its proper name. Hadjikhani and colleagues added an elegance to their study by showing that the color region they found, which they named V8, responded persistently after a color stimulus ceased. This poststimulus activity correlated with the perception of color afterimages. In this way, it resembled the motion aftereffect activity earlier observed in MT by Tootell and colleagues. Figure 15 illustrates the kind of stimulus they used to elicit color afterimages. The strong persistent color afterimage from Fig. 15 evokes the persistent afterstimulus activity drawn in Fig. 16.

There are other methods for testing for color activation besides the differencing method used by Zeki and by Hadjikhani. The differencing method assumes that color perception will be associated with activation that is linked to color but that disappears when achromatic stimulation is used. Implicitly, this method includes the assumption that the neural mechanism for color perception is composed of neurons that only respond to color and do not respond...
VI. OBJECT RECOGNITION IN LOC AND VENTRAL OCCIPITAL–TEMPORAL CORTEX

One of the chief functions of visual perception is object recognition. Although it is still too soon to tell the whole story of how the visual system recognizes and categorizes objects, recent research indicates remarkable specialization of specific regions in visual cortex for recognizing objects and even classes of objects. This topic can be subdivided into a discussion of LOC and ventral occipital–temporal cortex.

Lateral and slightly anterior to retinotopic areas V3A and V4V, there are regions of human occipital cortex that are activated preferentially by pictures of objects compared with pictures of random textures. This has been shown repeatedly in fMRI experiments on human subjects, first by Rafi Malach working with Tootell and then by Malach and colleagues. The ventral region most strongly activated is called LOC. LOC activation has many distinguishing characteristics. In experiments with cut and scrambled pictures of objects, into LOC activation declined only slightly when the pictures were cut into four pieces and scrambled, but it declined precipitously when the pictures were cut and scrambled into smaller pieces (Fig. 17). LOC activation was relatively invariant to change in location and to changes in the size of pictures of the same object. Therefore, activation of LOC seems to resemble neuronal activation in IT cortex of monkeys. LOC seems to have a bias toward visual information from the center of the visual field. From fMRI studies on retinotopic mapping, Tootell and colleagues concluded that the region in which LOC is located receives interhemispheric inputs, meaning that it represents visual signals coming from both right and left visual fields. LOC is also an area activated by the IC stimuli used by Mendola (Fig. 7). These various experimental results suggest that LOC could be an important area for intermediate-level vision, where segmentation and grouping are made explicit in the activity of neurons and neuronal populations.

Further specialization for the recognition of classes of objects has been found in occipital–temporal areas of human cerebral cortex. These areas are located in the ventral cortex and more anteriorly. A much studied region is the anterior fusiform gyrus, where pictures of faces activate the region preferentially, as first shown by Nancy Kanwisher and colleagues. It was shown that pictures of faces compared to pictures of houses activate this fusiform face area (FFA) differentially. Also, pictures of faces compared to inverted face pictures also produce differential activation in a
Nearby but distinct from the FFA is the parahippocampal place area (PPA), a region that prefers pictures of houses to those of faces. This more posterior region is located in the collateral sulcus, also in the ventral occipitotemporal cortex. Malach and colleagues showed that the FFA was biased more toward central vision, whereas the PPA in the collateral sulcus was biased more toward the periphery of the visual field. Also, these occipitotemporal regions overlap with the interhemispheric region mapped by Tootell and thus represent both right and left visual hemifields.

The specialization of the face and house regions of the occipitotemporal cortex was shown in an elegant experiment by Frank Tong and colleagues. They used the phenomenon of binocular rivalry combined with fMRI measurements to explore the modulation of activity with change of perceptual state. Subjects viewed face and house pictures that were present monoptically—for example, a picture of a house to the right eye and a picture of a face to the left eye. In these circumstances, perception is bistable: First one sees a face, and then after some time the picture of the face fades and one sees a house, and the two percepts repeatedly alternate in perception. Tong and colleagues asked subjects to press different keys when they saw a face or house, and they averaged the fMRI signal with respect to the key presses. They observed the results in Fig. 18. FFA activation increased when subjects perceived faces in the rivalrous situation, and PPA activation increased when they saw houses. The respective activations were comparable to what was measured when only faces or only houses were shown. This experiment reveals in a very direct manner that activation of these specific cortical areas is highly correlated with the perception of specific categories of objects.

**VII. CONCLUSION**

Perception requires location and mapping of space, segmentation, sensitivity to motion and color, and recognition of familiar or unfamiliar shapes. We discussed all these aspects of visual perception and the areas of the cerebral cortex that are activated when these are seen. It can be concluded from the experiments we reviewed that the neural basis of visual perception is based on specialized modules in the visual cortex. However, in perception there can be an overriding influence of part of a visual scene on the perception of other parts of the scene, as in the effects of visual cues for segmentation. Therefore, cannot be definitively concluded that specialized modules are all
there is to neural mechanisms of visual perception. A major role may be played by the interaction between different visual areas and between the visual networks in occipital cortex and memory- or decision-related cortical networks elsewhere in the brain. As the Gestalt psychologists used to say about visual images, so also we can say about the visual cortex: The whole is greater than the sum of its parts.
Figure 18  Binocular rivalry between pictures of faces and of houses. The fusiform face area (FFA) and parahippocampal place area (PPA) were monitored with fMRI in human subjects during binocular rivalry between pictures of faces and of houses. There was a clear modulation of the amount of activation during and after the switch from house to face and from face to house, as seen in a. (b) A control by stimulating both eyes with pictures of faces and of houses and measuring the magnitude of fMRI amplitude modulation caused by switching the stimuli and percepts [reproduced with permission from Tong et al. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 21, 753–759.]

**See Also the Following Articles**

**Suggested Reading**


## Encyclopedia of Human Brain

### Article 0361/Vision: Brain Mechanisms

**Queries:** While preparing this paper/manuscript for typesetting, the following queries have arisen.

<table>
<thead>
<tr>
<th>Query No.</th>
<th>Page</th>
<th>Column</th>
<th>Line No.*</th>
<th>Description</th>
<th>Author’s Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>3</td>
<td>2</td>
<td>42 to 45</td>
<td><strong>Author:</strong> Please check the meaning of this sentence.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>8</td>
<td>1</td>
<td>3 to 8</td>
<td><strong>Author:</strong> Please check the meaning of this sentence.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>8</td>
<td>2</td>
<td>39 &amp; 40</td>
<td><strong>Author:</strong> “looking like” has been replaced with “appearing as”. Please check.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>14</td>
<td>2</td>
<td>24</td>
<td><strong>Author:</strong> “It is from the work of” has been deleted before “Rainer Goebel”. Please check.</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>15</td>
<td>2</td>
<td>18</td>
<td><strong>Author:</strong> One sentence after “color afterimages.” has been deleted, as this information is about the color figure which is being produced in black &amp; white.</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>17</td>
<td>2</td>
<td>25</td>
<td><strong>Author:</strong> “repeatedly” has been inserted after “percepts”. Is it OK?</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>17</td>
<td>2</td>
<td>23</td>
<td><strong>Author:</strong> Please check the meaning of this sentence.</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>2, 4, 5, 7, 10 to 15, 17 to 19</td>
<td>Figures 1 to 4, 6, 7, 9 to 14, 16 to 18</td>
<td><strong>Author:</strong> Please provide publisher permissions for Figures 1 to 4, 6, 7, 9 to 14, 16 to 18. Author permissions are on file.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>9</td>
<td>Line 3 of Figure 7 caption</td>
<td><strong>Author:</strong> “from” has been replaced with “as obtained by”. Is it OK?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>17</td>
<td>Figure 16</td>
<td><strong>Author:</strong> “longer than” has been replaced with “compared to”. Is it OK?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Number of lines down from the top of the column, including headings.