CHAPTER

52

Object and Face Recognition

Martha Farah, G.W. Humphreys, and Hillary R. Rodman

As objects and viewers move with respect to one another, the images cast by objects on the viewer's retinas can change radically. Not only will image position and size change, but object surfaces may be foreshortened, occluded, or newly revealed. Similarly, changes in lighting conditions can reveal new surfaces, hide old ones in shadow, and introduce spurious edges where light ends and shadow begins. Yet the visual system is able to extract some reasonably constant representation of the shapes of objects and faces across these transformations. Image processing by the retina, lateral geniculate nucleus of the thalamus, and occipital visual areas makes some headway toward stable object representations by extracting features and representing them over increasingly large receptive fields (see Chapter 28). However, in going from the occipital to the temporal visual cortex, the stimulus representation undergoes its most radical transformation away from image properties and toward intrinsic object shape properties. In this chapter, we examine the role of the temporal cortex in object and face recognition.

EVIDENCE FOR VISUAL AREAS BEYOND THE OCCIPITAL LOBE

The earliest clues about the neural bases of object recognition came from the study of patients with visual agnosia. Visual agnosia is a blanket term for a wide array of visual disorders affecting object recognition, in which elementary visual functions such as acuity and visual fields are grossly intact or at least adequate to allow recognition (see Farah[1] for a taxonomy and detailed review). Lissauer[2] is usually credited with the first detailed discussion of visual agnosia, and early attempts to generalize about the lesions in such cases are found mainly in the German neurology literature of the early decades of this century.[3] Writers as early as Potzl[4] recognized the importance of inferior temporal regions in visual agnosia, including the lingual and fusiform gyri, as well as adjacent occipital regions. Lesions in patients with visual agnosia are typically bilateral or right-sided; left-sided lesions are found primarily in those who retain face recognition ability.[5,6]

Object Recognition in Monkeys Depends on the Inferior Temporal Cortex

One of the earliest experimental studies in animals on the role of the temporal cortex in visual recognition was that of Klüver and Bucy,[7] on what is now known as the *Klüver–Bucy syndrome.* When these researchers removed the temporal lobes of monkeys bilaterally, the monkeys demonstrated complex changes in visual, social, sexual, and eating behavior. These changes appeared to result from a combination of perceptual, mnemonic, and motivational impairments. Later researchers attempted to identify the areas of the temporal lobe responsible for the visual functions disrupted in these monkeys, eventually identifying the inferior temporal gyrus, also known as **inferior temporal** (IT) cortex, as the critical area for producing the visual deficits.[8-10] Monkeys with IT lesions were impaired with respect to learning visual discriminations among different patterns or objects and retaining previously acquired visual discriminations.[11] Figure 52.1 shows the location of this area in the macaque brain.[12]

What type of stimulus representation is impaired after IT lesions? Much research has been aimed at answering this question. The general approach has been to infer which stimulus properties are normally encoded (or not encoded) in IT representations by show-

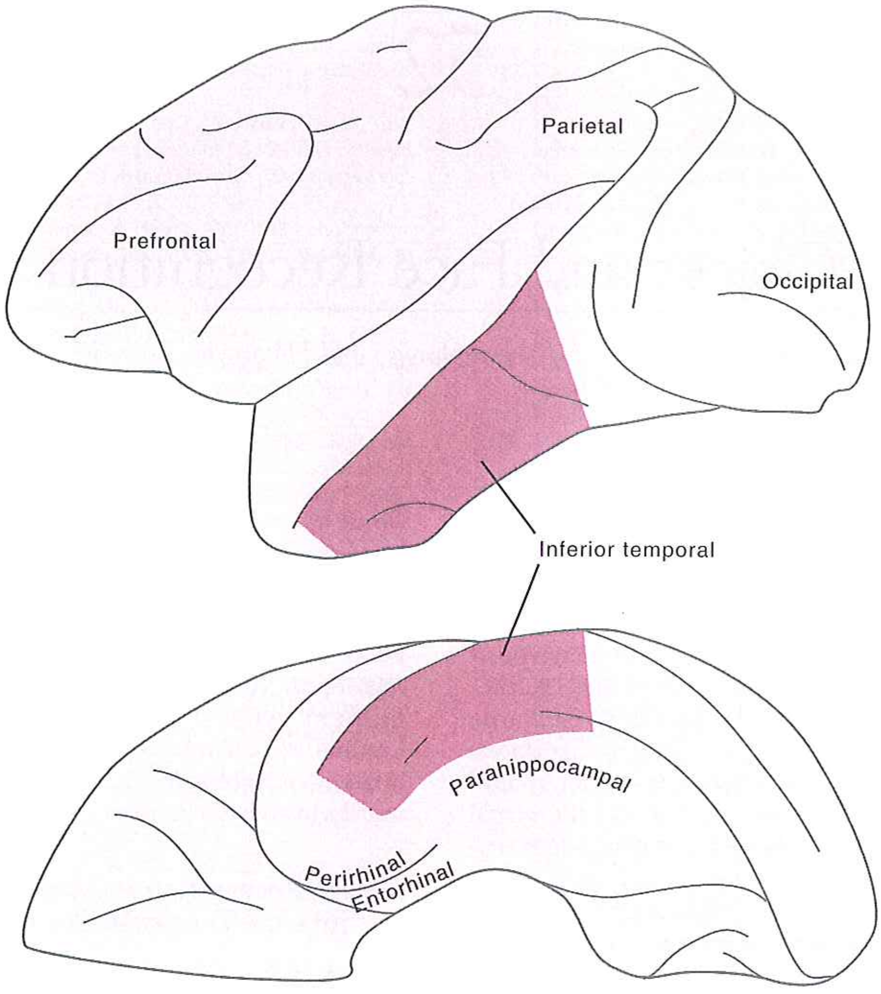


FIGURE 52.1 The inferior temporal region in a rhesus monkey. (Top) Lateral view. (Bottom) Ventral view. Adapted from Dean.[12]

ing which stimulus properties monkeys with IT lesions are impaired at using (or not impaired at using) as a basis for discrimination. On the basis of our current knowledge, a reasonable short answer might be that IT cortex represents aspects of the intrinsic shape of the stimulus that are useful for recognition (e.g., the salient features in a face and their relative positions) and omits most aspects of stimulus appearance that depend on viewing conditions (e.g., absolute size and luminance). We present a few representative studies here.

Position in the visual field is one visual property that is clearly not necessary for objective recognition, and normal monkeys will easily generalize a visual discrimination learned in one hemifield to the other. In other words, if monkeys learn to tell apart stimuli in one half of visual space, they will easily transfer this discrimination when the same stimuli are switched to the other half. The fact that monkeys with bilateral IT lesions are impaired at this generalization implies that they have lost representations that are invariant across visual field locations.[13,14] The retinal image size of the stimulus is another visual property that depends on viewing conditions and not just intrinsic object geometry, and this is another property that monkeys with IT lesions have trouble ignoring. For example, in one experiment, monkeys were trained to discriminate between two disks of different absolute sizes, varying their distance and hence their retinal image size.[15] Monkeys with IT lesions were unable to relearn the discrimination, responding instead on the basis of retinal size or distance. This result implies that representations within the IT cortex normally encode the absolute size of an object, an intrinsic object property useful for recognition, rather than its distance per se or its retinal image size.

Monkeys with IT lesions are also typically impaired at generalizing across views of the same stimulus in a different orientation,[16] implying that orientation is yet another of the incidental image properties that has been discarded from representations within IT cortex. Finally, variations in illumination prevent monkeys with IT lesions from seeing the equivalence of objects,[16] implying that representations in IT cortex are unaffected by patterns of shadow and light falling on object surfaces.

Damage to Object Representations in Humans Causes Associative Visual Agnosia

Agnosias are commonly divided into **apperceptive** and **associative** varieties.[2] Agnosic patients whose perception is obviously impaired, despite intact or at least adequate visual sensory function, are classified as apperceptive agnosics on the basis that their impairment is a perceptual one, that is, in the stage of apperception. Agnosic patients whose perception seems grossly intact are classified as associative agnosics on the basis that their impairment is in the stage of association of percept and memory. The apperceptive and associative distinction is valid in the sense that there are agnosic patients with and without obvious perceptual impairments, and their underlying problems do appear to be different. However, it is probably not true that the underlying problem in associative agnosia is one of association, as opposed to perception. In fact, whereas apperceptive agnosia is thought to result from relatively early impairments in image processing, associative agnosia can be conceptualized as an impairment of perceptual object representation *per se.*

To be considered an associative agnosic, a patient must demonstrate the following features: First, he or she must have difficulty recognizing visually presented objects, as measured by both naming and nonverbal tests of recognition, such as sorting objects by category (e.g., putting kitchen utensils together, separate from sports equipment) or pantomiming the objects' functions. Second, the patient must demonstrate that knowledge of the objects is available through modalities other than vision, for example, by tactile or auditory recognition or by verbal questioning (e.g., "What is an egg beater?"). Third, the patient must be able to see the object clearly enough to describe its appearance, draw it, or answer whether it is the same as or different in appearance from a second stimulus. A good illustrative case of associative visual agnosia has been described by Rubens and Benson[17] and is summarized in Box 52.1. Figure 52.2 shows some copies made by their patient, who did not recognize the pictures he copied.

Summary

How can someone be of sound mind, see pictures clearly enough to produce the copies shown in Fig. 52.2, and yet not recognize the pictures? This constellation of abilities and impairments seems almost paradoxical, and perhaps for this reason the very existence of visual agnosia has been doubted.[18,19] The resolution of this paradox lies in the realization that copying, matching, and so forth can be accomplished using representations based on the specific appearance of the image. In contrast, as discussed at the beginning of this chapter, such representations are too changeable to suffice for object recognition. Indeed, closer examination of the manner in which agnosics copy and match pictures suggests that they are abnormally tied to the local structure of the image rather than guided by the shape of the object itself: their copies are executed in a slow and slavish manner, and they may classify two pictures as different because of a small flaw in the printing of one (see Farah[1] for a review of these and other sources of evidence for perceptual impairment in associative agnosia). Unfortunately, the effects of transformations in size, orientation, lighting, and so forth have not been studied systematically in associative agnosics, as they have been in monkeys with IT lesions and in recordings from single IT neurons. Nevertheless, the evidence available from copying, matching, and other tasks is consistent with an impairment in representing the intrinsic shapes of objects and a consequent overreliance on structural image representations.

VISUAL PATHWAYS FOR OBJECT RECOGNITION

The Visual Cortex of Primates Can Be Divided into Dorsal and Ventral Streams

A major portion of the neocortex of primates is devoted to the processing of visual information. This expanse of visual cortex has been differentiated into a large number of separate regions, which can be thought of as organized in a hierarchical fashion. Visual cortical areas of primates can be divided roughly into a **dorsal stream,** concerned with the analysis of the locations of objects and their movements in space, and a **ventral stream,** responsible for the processing and storage of information about their identity—their shape, color, and other salient physical features. Although the two cortical streams are not completely segregated anatomically or functionally, this formulation, first put forth in 1982 by Ungerleider and Mishkin[20] (Fig. 52.3A)[21]

BOX 52.1

CASE STUDY OF ASSOCIATIVE OBJECT AGNOSIA

The subject was a 47-year-old man who had suffered an acute loss of blood pressure with resulting brain damage. His mental status and language abilities were normal, and his visual acuity was 20/30, with a right homonymous hemianopia (blindness in the right visual hemifield). His one severe impairment was an inability to recognize most visual stimuli.

For the first 3 weeks in the hospital, the patient could not identify common objects presented visually and did not know what was on his plate until he tasted it. He identified objects immediately on touching them.

When shown a stethoscope, he described it as "a long cord with a round thing at the end" and asked if it could be a watch. He identified a can opener as a key. Asked to name a cigarette lighter, he said, "I don't know" but named it after the examiner lit it. He said he was "not sure" when shown a toothbrush.

He was never able to describe or demonstrate the use of an object if he could not name it. If he misnamed an object, his demonstration of use would correspond to the mistaken identification. Identification improved very slightly when given the category of the object (e.g., something to eat) or when asked to point to a named object instead of being required to give the name. When told the correct name of an object, he usually responded with a quick nod and often said, "Yes, I see it now." Then, often he could point out various parts of the previously unrecognized item as readily as a normal subject (e.g., the stem and bowl of a pipe and the laces, sole, and heel of a shoe). However, if asked by the examiner, "Suppose I told you that the last object was not really a pipe, what would you say?" He would reply, "I would take your word for it. Perhaps it's not a pipe." Similar vacillation never occurred with tactilely or aurally identified objects.

After he had spent 3 weeks on the ward, his object-naming ability improved so that he could name many common objects, but this was variable; he might correctly name an object at one time and misname it later. Performance deteriorated severely when any part of the object was covered by the examiner. He could match identical objects but could not group objects by categories (clothing, food). He could draw the outlines of objects (key, spoon, etc.) that he could not identify.

He was unable to recognize members of his family, the hospital staff, or even his own face in the mirror. Sometimes he had difficulty distinguishing a line drawing of an animal face from a man's face but always recognized it as a face.

The ability to recognize pictures of objects was greatly impaired, and after repeated testing he could name only 1 or 2 of 10 line drawings. He was always able to name geometrical forms (circle, square, triangle, cube). Remarkably, he could make excellent copies of line drawings and still fail to name the subject. He easily matched drawings of objects that he could not identify and had no difficulty discriminating between complex nonrepresentational patterns, differing from each other only subtly. He occasionally failed in discriminating because he included imperfections in the paper or in the printer's ink. He could never group drawings of objects by class unless he could first name the subject.

Reading, both aloud and for comprehension, was greatly limited. He could read, hesitantly, most printed letters but often misread "K" as "R," and "L" as "T" and vice versa. He was able to read words slowly by spelling them aloud. (Excerpted from Rubens and Benson.[17])

and subsequently elaborated by these authors and other workers,[22,23] remains one of the most valuable principles in our understanding of the visual cortex. The following section deals with the organization of the ventral stream and how cells in areas within it represent objects in terms of their distinguishing individual features and the configurations of these features, which together make up coherent wholes.

The Ventral Stream of Monkeys Consists of an Interconnected Set of Visual Cortical Areas

Most of our detailed knowledge of the organization of primate visual cortex, including that of pathways specialized for object recognition, comes from studies of Old World monkeys of the genus *Macaca*. Similar areas and connections appear to exist in New World

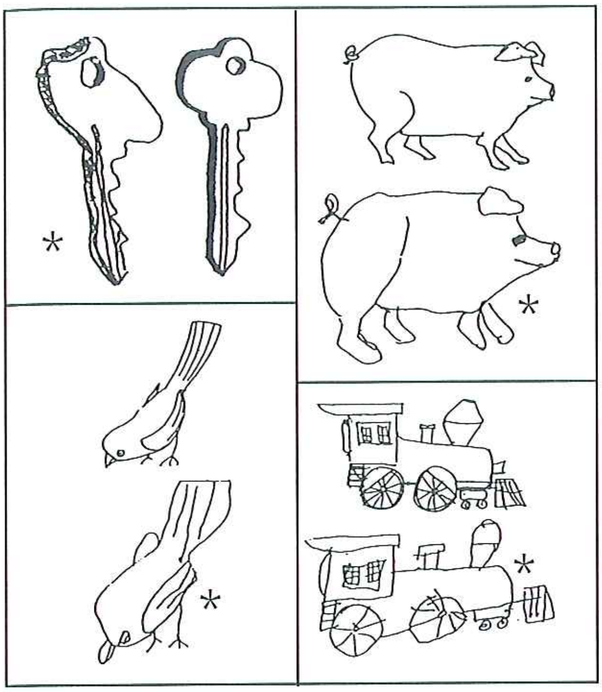


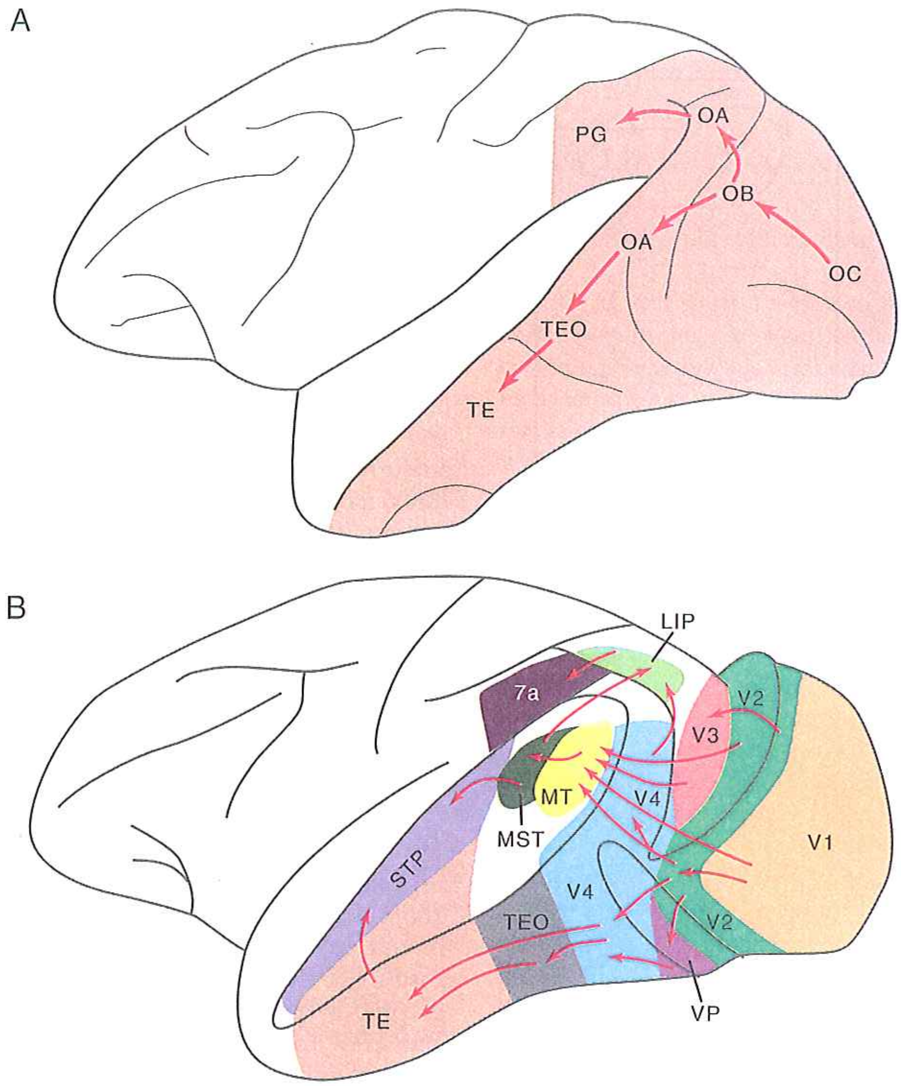
FIGURE 52.2 Pictures that a patient with associative visual agnosia did not recognize but was able to copy. Asterisks indicate the patient's copies. From Rubens and Benson.[17]

monkeys. Figures 52.3B and 52.3C illustrate the locations of areas and overall flow of information within the ventral stream of the macaque monkey. Analysis of objects begins in V1, where information about edges and about the color and brightness of stimuli is extracted from visual inputs from the dorsal lateral geniculate nucleus (Chapter 28). This information is then sent on to certain subdivisions of area V2. From V2, the object recognition pathway takes a largely ventral, occipitotemporal course. Visual information about objects is sent from V2 to area V4 on the lateral and ventromedial surfaces of the hemisphere and to a posterior inferior temporal area, TEO, just anterior to V4. From these regions, information proceeds forward to area TE, located in the anterior inferior temporal cortex. Together, areas TEO and TE constitute IT cortex as described in the previous section. From there, visual information is sent to the most ventral and anterior reaches of temporal lobe neocortex, namely, temporal polar areas TG and perirhinal area 36. These areas and adjacent parahippocampal (areas TF and TH) and entorhinal (area 28) cortices are interconnected with medial temporal lobe structures, most notably, the hippocampus, which are crucial for forming memories of visual objects.

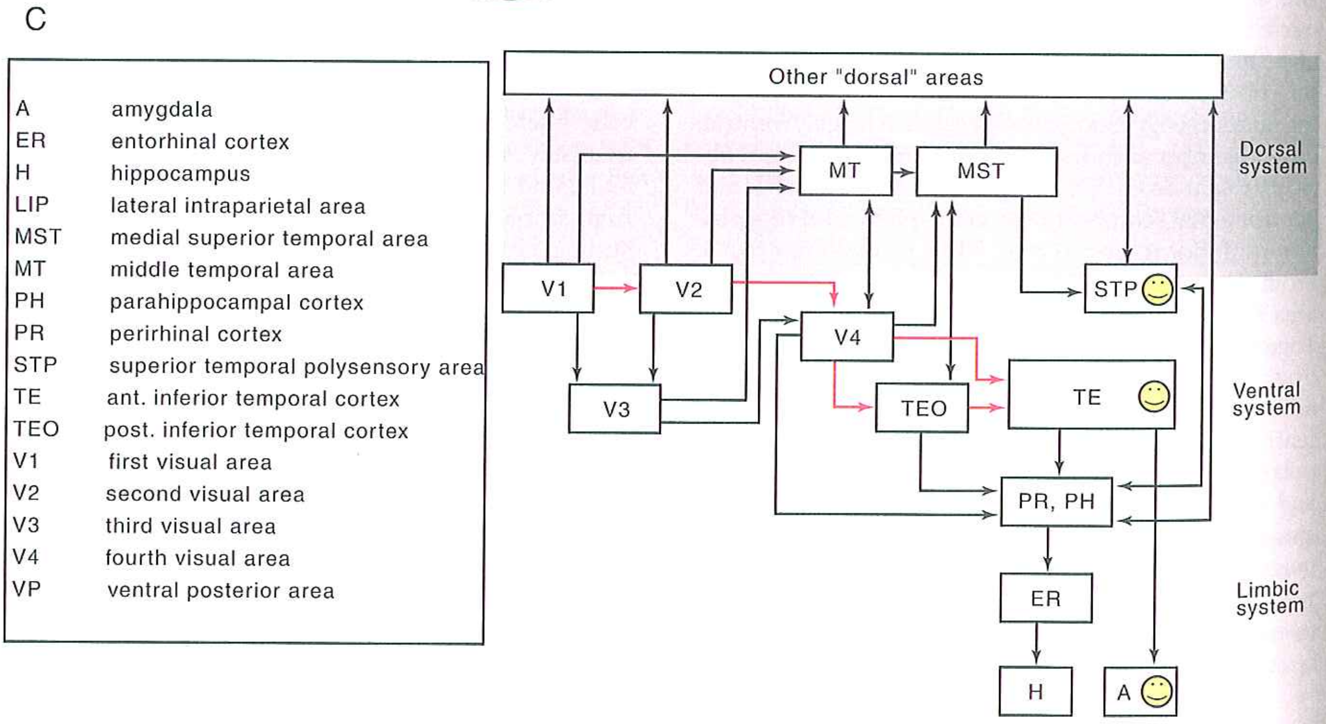
At all stages of the pathway, connections tend to be reciprocal, so that an area receiving feedforward projections from an area earlier in the ventral hierarchy provides feedback projections to that area. Long-range nonreciprocal connections, such as those from area TE back to V1, also exist. In addition, the ventral object recognition pathway involves interconnections of temporal areas, such as TEO and TE, with ventral portions of the frontal lobe neocortex and with portions of the superior temporal polysensory area (STP), which appears to contribute to both object and motion analyses on the basis of converging projections from both dorsal and ventral streams.[24] Further, all areas in the ventral stream are interconnected with subcortical visual structures, most notably with the pulvinar and portions of the amygdala, claustrum, and basal ganglia. Finally, all areas in the ventral stream are interconnected with their counterparts and other ventral stream areas of the opposite hemisphere through the corpus callosum or, in the anterior and ventral regions, through the anterior commissure.

Trends in Anatomical Features Are Evident along the Ventral Stream

Paralleling the flow of information just described, several trends in anatomical features are evident as one progresses forward along the ventral stream of monkeys.[23,25,26] First, at successive stages, patterns of cortical connectivity become successively less topographic or "point-to-point." Inputs to area TE from V4 and TEO, for example, have no obvious retinotopic organization. Similarly, connections with subcortical structures, both visual and nonvisual, become progressively more diffuse at successive levels of this pathway. For example, whereas areas V2 and V4 are topographically interconnected with portions of the pulvinar that themselves show retinotopic organization, areas such as TE and STP have nontopographic connections with nonretinotopic pulvinar subdivisions. Third, commissural pathways interconnecting areas of the ventral stream on the two sides of the brain tend to be restricted to the representation of the midline of visual space early in the pathway, but their representation becomes more widespread at higher levels in the pathway. Other physical gradients are also present. The concentration of opiate receptors increases in a graded fashion along the ventral stream, perhaps reflecting an increasing role for affect-based attentional processes at successive levels of the system.[27] Successive levels of the ventral stream also show increasing levels of activation of several proteins that have been implicated in synaptic plasticity and long-term storage of information in the brain.[28]



A	amygdala
ER	entorhinal cortex
H	hippocampus
LIP	lateral intraparietal area
MST	medial superior temporal area
MT	middle temporal area
PH	parahippocampal cortex
PR	perirhinal cortex
STP	superior temporal polysensory area
TE	ant. inferior temporal cortex
TEO	post. inferior temporal cortex
V1	first visual area
V2	second visual area
V3	third visual area
V4	fourth visual area
VP	ventral posterior area



Trends in Neuronal Response Properties Are Also Evident along the Ventral Stream

In monkeys, the different cortical areas in the ventral stream share a number of physiological characteristics. Consistent with their role in object recognition, all areas in the pathway contain populations of cells sensitive to the shape, color, or texture (regular pattern of detail) of visual stimuli or some combination of these features.[23,29] Four major physiological trends are evident as one moves forward along the pathway. First, the nature of the visual field representation changes from a precise map of the contralateral visual field to a more diffuse mapping of overall visual space. For example, moving from V1 to V2 to V4 and then to areas TEO and TE, neuronal receptive fields (i.e., the portions of visual space from which a stimulus can evoke a neuronal response) become larger (Fig. 52.4)[30–33] and their organization into a map less precise, reflecting changes in anatomical connectivity patterns, so that by area TE no clear retinotopic organization is present. The amount of the ipsilateral visual field represented in the receptive fields (RFs) in a given area also increases at successive levels of the pathway, which is again consistent with anatomical trends; RFs vary from 1 or 2 degrees of the visual field in areas V1 and V2 to the large, bilateral RFs found in areas TE and STP. The second trend is an increase in the degree to which particular stimulus features, such as an appropriately oriented edge, are signaled regardless of their exact retinal location; in other words, there is an increase in the extent of invariance of neuronal responses, progressing from a degree or two of such tolerance in the RFs of complex cells in V1 to the maintained stimulus selectivity across wide expanses of visual space exhibited by cells in TE and STP.[34] The third trend is toward selectivity for increasingly complex stimulus features or combinations of features, progressing from the selectivity for simple dimensions such as the orientation of edges and wavelength in V1 to the preference for complex objects, including biologically meaningful stimuli, such as faces and body movements, shown by many neurons in late stages of the ventral pathway. Finally, progressing along the ventral stream and continuing into the frontal lobe, there is an increase in the extent to which cells contribute information related to the recentness and familiarity of visual objects (visual memory traces).[35]

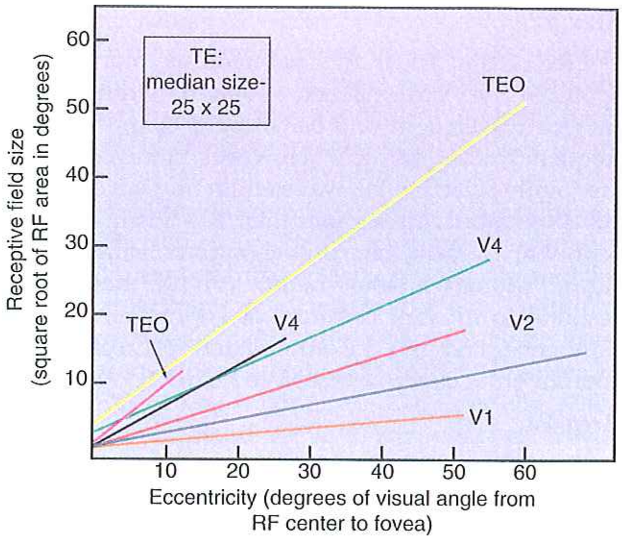


FIGURE 52.4 Regression lines relating multiunit receptive field size (square root of receptive field area) to receptive field eccentricity in various areas along the ventral stream. The two lines for V2 (red and blue) show data separately for two different animals. Adapted from Boussaoud *et al.*,[30] Fenstemaker *et al.*,[31] Gattass *et al.*,[32,33] and Gross *et al.*[23]

FIGURE 52.3 (A) Schematic diagram of the monkey brain illustrating the two cortical visual system model of Ungerleider and Mishkin.[20] According to this model, there are two major processing pathways, or "streams," in the visual cortex, both originating in the primary visual cortex: a ventral stream, directed into the temporal lobe and crucial for the identification of objects, and a dorsal stream, directed into the parietal lobe and crucial for spatial perception and visuomotor performance. The shaded region indicates the extent of visual cortex in the monkey, and the labels on the figure (OC, OB, OA, TEO, TE, PG) indicate cytoarchitectonic areas according to the nomenclature of von Bonin and Bailey.[21] Adapted from Mishkin *et al.*[22] (B) A lateral view of the monkey brain illustrating the multiplicity of functional areas within the visual cortex and the connections between them. For simplicity, only projections from lower-order to higher-order areas are shown, but each of these feedforward projections is reciprocated by a feedback projection. (C) Some of the pertinent connections of inferior temporal cortex with other cortical areas and medial temporal lobe structures. Red lines indicate the main afferent pathway to area TE, which includes areas V1, V2, V4, and TEO. Faces indicate areas in which neurons selectively responsive to faces have been found. Adapted from Gross *et al.*[23]

Novel Neuronal Response Properties Emerge in Successive Areas in the Ventral Stream

There is considerable variation in the degree to which different components of the ventral stream have been explored physiologically in monkeys. Overall, areas V2, V4, and TE have received the most attention. Neuronal activity related to object recognition in areas STP, and in areas 36 and TG (sometimes considered the most anterior-ventral portion of IT cortex), has also become a recent focus of research. Less is currently known about the object-related visual response properties of neurons in area TEO, parahippocampal areas TF and TH, and the ventral frontal cortical regions to which ventral temporal visual areas project.

Area V2

Like cells in V1, individual neurons in area V2 respond in a selective fashion to length, width, orientation, or wavelength of a bar of light or to the spatial frequency of a grating.[36,37] However, some cells in V2 are jointly selective for wavelength and small dimensions of stimuli throughout their RFs; such **spot cells** represent an early stage of integration of stimulus features.[36] Moreover, some neurons in V2 (but not V1) respond to **illusory contours** of appropriate orientation generated by the stimulation of regions surrounding the classical receptive field.[38]

Area V4

A number of fundamental elaborations become evident in area V4. First, many neurons in V4 are conjointly tuned for both the length and the width of a bar stimulus[39] and maintain such selectivity over a considerably wider RF than at earlier stages of the visual pathway. Second, many cells in V4 show a combination of varied shape selectivity with wavelength selectivity. Third, beginning in V4, a small number of cells are found with optimal stimuli clearly more complex than oriented edges or spots, consisting instead of features such as edge conjunctions, irregular borders, and concentrically organized areas of contrast (Fig. 52.5).[40,41] A fourth and critical elaboration in V4 is the appearance of large, **silent surrounds** adjacent to the excitatory or classical RF.[29] These surrounds presumably reflect the increase seen at the level of V4 in commissural inputs and feedback projections from more anterior areas with large RFs. Stimulation of the surrounds typically influences activity of the cell only when carried out in concert with stimulation of the classical RF, and the selectivities of the surround and classical RF for form, wavelength, spatial frequency, and so on are usually antagonistic. Accordingly, such cells tend to respond maximally only when the stimulus stands out from its background on the basis of a difference in such features. These antagonistic surrounds may represent an important step in figure-ground separation—that is, the breakdown of scenes into objects to be analyzed.

A fifth development at the level of V4 has to do with the perception of color. Originally, V4 had been thought of as an area specialized for color processing, but it now appears that the percentage of cells in V4 tuned for wavelength (and the sharpness of their tuning) does not differ significantly from that in other areas in the ventral stream. Moreover, both the physiological studies described previously and the effects of lesions underscore the crucial role of this area in form perception. Nevertheless, many cells in monkey V4 have been shown to exhibit color constancy.[42] In other words, like human observers, they perceive (i.e., respond maximally) to the apparent color of an object defined by the relative reflectances of the object and its surround, independent of the actual wavelengths reflected by the object itself. The complementary wavelength selectivities of the classical RF and of the surround of V4 cells have been suggested to form the basis for the behavior of these cells.

A final noteworthy development at the level of V4 is the sensitivity of neuronal firing patterns to several

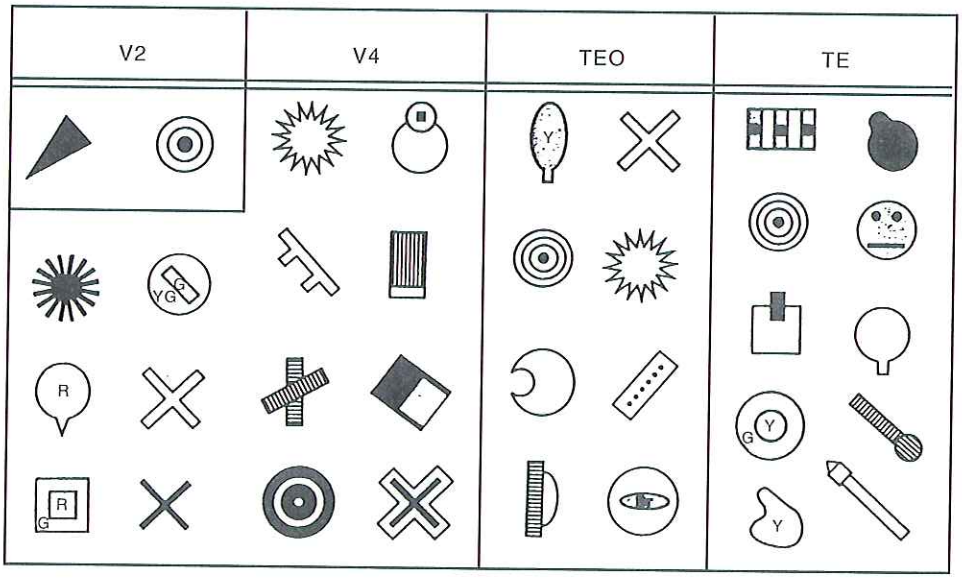


FIGURE 52.5 Examples of the complex object features preferred for four ventral stream areas. Y, yellow; R, red; G, green; YG, yellow green; Br, brown. Adapted from Kobatake and Tanaka.[41]

types of nonvisual signals. For example, whether or not an animal pays attention to a particular visual stimulus located in the RF of a V4 neuron can dramatically affect the strength of a cell's response to the stimulus.[43] There is also some evidence that requiring an animal to generate a visual mental image of a cell's preferred stimulus, based on input in other sensory modalities, can actually activate V4 neurons.[44]

Inferior Temporal Cortex

An additional set of crucial elaborations in the response properties of ventral stream neurons takes place as one moves forward into the IT cortex. First, moving successively into area TEO and then into posterior and anterior portions of area TE, there is a progressive increase in both the complexity of the critical features needed to activate the neurons in each area and the proportion of neurons that are selectively driven by some type of complex pattern or object. Although some proportion of cells selectively responsive to simple stimulus parameters of length, width, orientation, wavelength, and spatial frequency remain in each of these regions, by the level of anterior area TE many neurons respond only to relatively complicated stimuli or classes of stimuli, often three-dimensional stimuli with particular patterns of internal detail.[45] Among cells selective for particular classes of complex patterns or objects, cells that respond selectively to faces are particularly common and have been studied extensively (see following section). Attempts by various workers[45,46] to determine whether simpler critical features within complex objects are crucial for eliciting the response have shown that this is true for only some of those TE neurons that respond maximally to complex stimuli.

Second, while some TE cells respond only to particular complex stimuli or stimulus classes, other TE neurons fire equally well to a great variety of stimuli, albeit with a varying response strength. Still others show what has been called **paradoxical selectivity:** they fire best or only in response to a small number of complex stimuli with few obvious common features.[23] For the first time in the ventral pathway, the critical stimulus dimension for a majority of neurons remains unspecified. All of these observations suggest that TE neurons do not code objects either by serving as complex feature detectors or by responding to any shape primitives thus far tested (an exception to this may be *face-selective* neurons described later). Rather, most workers in the field believe that by the level of area TE, visual stimuli are represented by a pattern of activity across a population or ensemble of cells, each member of which may participate in representing several different objects. Despite this cooperative activity across cell types, cells in area TE with similar types of selectivity tend to cluster into clumps or modules, possibly organized into vertical columns like the ocular dominance and orientation columns demonstrated in V1. (See Chapter 28.)

A third critical development that appears by the level of TE is the considerable response invariance achieved by many neurons. Many neurons throughout the ventral pathway tend to prefer a given stimulus type or parameter independent of the size or exact position of the stimulus within their RFs. However, in area TE, RFs can be very large—20 degrees on a side or more—and extend far into the ipsilateral visual field, such that a proportion of cells in area TE respond to relatively complex objects over a wide expanse of the visual world and do so despite changes in location in depth, which alter retinal image size. A fairly precise characterization of the information represented by individual area TE cells can be obtained from experiments in which responses from a cell are recorded while specific properties of a stimulus are varied. The results of these experiments are generally consonant with the conclusions of the lesion studies reviewed earlier and with the general view that area TE represents intrinsic aspects of object shape and not incidental aspects of appearance that are determined by viewing conditions. For example, the position,[45] retinal image size,[47] and picture plane orientation[45] have relatively small effects on a cell's response to an optimal shape. Moreover, neurons that are selective for a particular shape defined by luminosity differences (e.g., a white star on a black background) are selective for the same shape defined by texture cues and motion cues [e.g., a star-shaped region of speckles with the same average luminosity as its background, defined by larger sparser speckles (texture cues) moving in a different direction from the background (motion cues)[48]]. It is these generalizing properties in particular that are believed to be responsible for the changes in object recognition following inferior temporal damage.

A fourth development—beginning in area TE, continuing into anterior–ventral areas, such as perirhinal area 36, and also into the frontal lobe—is the appearance of response patterns related not to an object per se but rather to its memory—that is, the fact that an object has been seen before, has been seen recently, or has been seen shortly after the appearance of another object.[49,50] For example, for some neurons in area TE, responses to an object that is presented repeatedly, thus losing its novelty, decrease systematically, possibly signaling that the object has been seen in the recent past and/or redirecting attention toward other, novel objects (Fig. 52.6). Some cells in area TE (and also in

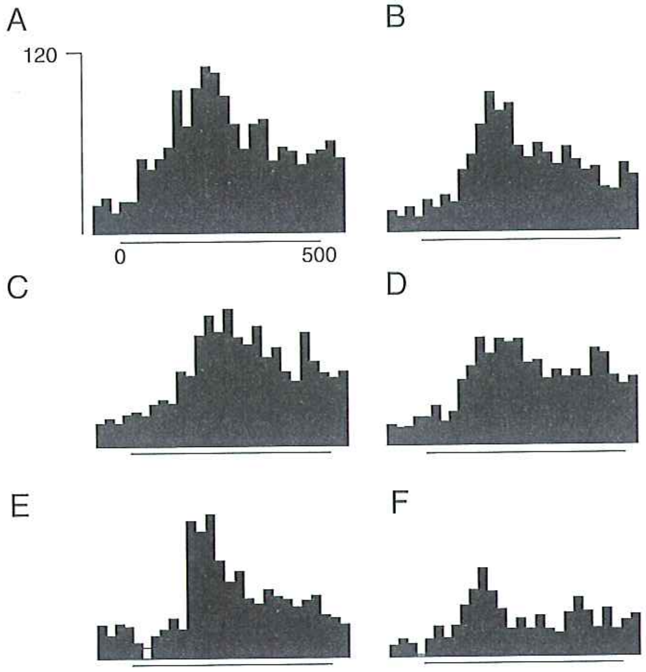


FIGURE 52.6 Examples of responses of an individual IT neuron to a set of 20 initially novel sample visual stimuli. (A, first presentation of 20 novel stimuli; B, second presentation; etc.) Responses to all 20 stimuli have been averaged together into single composite histograms. The six histograms show the average response to the first six presentations of all stimuli. The horizontal line under the histograms indicates the 500-ms time period when the stimuli were on. The vertical scale indicates firing rate in spikes per second, and the bin width is 10 ms. From Li *et al.*[49]

area TE's projection zones in prefrontal cortex) continue to fire during the delay period of a matching task after an object has been removed from view, but still needs to be remembered if the task is to be performed correctly; such activity has been interpreted as a visual memory trace or "holding" signal. Finally, some neurons in area 36, and possibly also area TE, develop conjoint and stable selectivity for unrelated objects that are repeatedly presented together or associated in time.[51]

Functional Imaging and Physiological Studies Demonstrate There Is a Homolog of the Ventral Stream for Object Recognition in Human Cortex

Recent technological advances have made it possible to ask about the neural substrates of perceptual and cognitive abilities in humans by using functional neuroimaging to identify brain regions activated during particular mental and physical states. In particular, **positron emission tomography** (PET) and **functional magnetic resonance imaging** (fMRI) have allowed scientists to start defining areas in the human brain involved in visual object recognition. For example, comparison of patterns of activity during processing of object location on the one hand and physical object features (shape, color) on the other clearly delineates separable dorsal and ventral pathways associated with "what" and "where"[52] in the human brain; as in monkeys, both dorsal and ventral pathways originate in the striate cortex. These studies highlight specific details of object processing in humans. First, much of the tissue devoted to object recognition in humans appears to be considerably more ventral than that in monkeys. Overall, the areas activated during object, color, and form recognition tasks (including identifying faces and naming objects) include the lingual and fusiform gyri (ventral occipitotemporal cortex) and anterior inferior temporal cortex of both hemispheres, the left lateral occipitotemporal cortex, and the right inferior frontal cortex[53,54] (Fig. 52.7). Second, processing of different types of object features activates regions that are only partially coincident, suggesting some localization of function within the human ventral stream at the level of whole areas. For example, foci activated by color processing appear to be located medial to those activated by perception of faces, and foci activated by nonface objects only partially overlap areas that appear to be involved in processing the facial image.[52] Brain imaging work on the neuroanatomy of processing of facial configurations in the human brain is discussed further in the next section.

Researchers have been able to obtain electrophysiological correlates of visual object perception and recognition from human patients with electrodes chronically implanted directly on the occipitotemporal and the inferior temporal cortex for the primary purpose of monitoring epileptic seizures.[56] Evoked potentials recorded by these electrodes when the patients viewed various categories of objects produced a pattern of results highly consistent with the imaging data discussed earlier (Fig. 52.7). Interestingly, potentials evoked specifically by alphanumeric stimuli (letter strings, words, and numbers) were from regions overlapping those associated with potentials related to other types of object stimuli. These findings support the notion that, at least at early levels of processing, linguistic material is treated within the object recognition system.

Summary

In monkeys, the ventral stream consists of a hierarchically organized set of areas with related patterns of anatomical connectivity, physiological response properties, and contributions to object vision. Activity in this ventral stream begins with the coding of simple parameters in early areas (V1 through V4) and progresses to anterior and ventral temporal areas to provide increasingly more abstract representations of visual objects and, ultimately, responses that may code the memory of previously seen objects. Although all areas in the pathway contain populations of cells selective for such simple parameters as edge orientation, color, and texture and for more complex conjunctions of object features, localization of function for components of object processing also exists in the sense that cells selectively responsive to particular stimulus features or to particular types of complex objects are organized into clumps or modules. Studies using functional neuroimaging methods (e.g., PET and fMRI) and direct monitoring of neural activity (event-related potentials) both show that a similar pathway exists within the ventral extrastriate cortex of the human brain.

FACE RECOGNITION: SPECIALIZED CORTICAL CIRCUITS?

Analysis and retention of facial images are crucial cognitive skills for primates. In many primates, the sense of smell is markedly reduced relative to that of

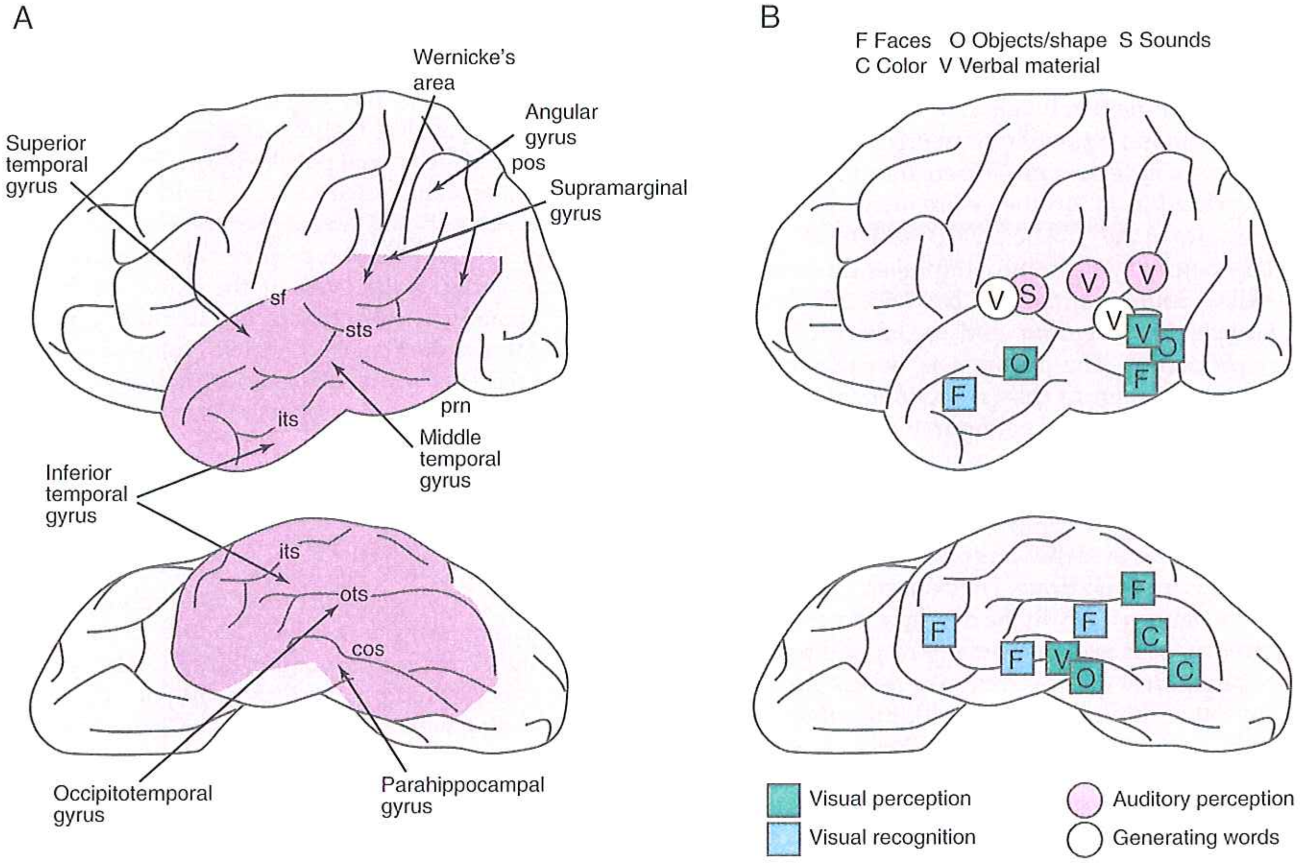


FIGURE 52.7 Schematic lateral (top) and ventral (bottom) views of human cerebral hemispheres. (A) Location of temporal cortex (purple) and major sulci and gyri. The posterior portion of the parahippocampal gyrus is often referred to as the *lingual gyrus,* and the posterior portion of the occipitotemporal gyrus is often referred to as the *fusiform gyrus.* cos, collateral sulcus; its, inferior temporal sulcus; ots, occipitotemporal sulcus; pos, parieto-occipital sulcus; prn, preoccipital notch; sf, sylvian fissure; sts, superior temporal sulcus. (B) Some of the major foci of activation produced in the temporal cortex using stimuli and conditions of the types indicated in neuroimaging and electrical evoked potential studies. Generating words tends to produce a posterior temporal focus of activity only in the left hemisphere, whereas the parahippocampal focus associated with face recognition tends to be found mainly in the right hemisphere. Adapted from Rodman.[55]

other vertebrate groups, and vision is a critical modality for social communication. Visual information inherent in faces thus carries a wealth of critical information about the identity of other individuals, their social positions, emotional states, and intended actions. The survival value of facial-image processing is reflected in the extraordinary memory capacity that humans and other primates show for faces, in the visual preferences for face stimuli shown by infants, and in our remarkable sensitivity to subtle differences among faces. Do face perception and recognition depend on specialized neural circuitry, possibly to some extent inborn, or do they instead reflect general mechanisms of object processing? The major lines of evidence concerning the existence of such dedicated cortical substrates are discussed in this section.

Some Neurons in the Temporal Cortex of Monkeys Respond Selectively to Faces

One of the more striking characteristics of area TE and of the immediately adjacent area STP is the presence of a small population of neurons with responses selective for face stimuli. When these neurons were first reported in 1972,[57] they were interpreted by some as examples of the gnostic units postulated[58] to provide unitary percepts of complex stimuli at the level of single cells.[59] Some years passed between this first report of face-selective neurons and the intensive study of their properties. The question is whether ostensibly face-selective neurons truly respond to visual information unique to the facial configuration or whether their selectivity is more parsimoniously explained as a response to features shared by faces and other types of objects. Although a definitive answer is still elusive, the bulk of the evidence suggests that the former description is more accurate. For example, although face-selective neurons vary in the degree and nature of their preference for face stimuli, many respond to both real faces or pictures of faces, but give nearly no response to any other stimuli tested, including other complex objects, textures, colors, and pictures in which the smaller features making up the face are rearranged or scrambled (Fig. 52.8). Moreover, for many such neurons, the specificity of the response for faces is maintained over a variety of stimulus transformations such as changes in size, position in the RF, angle of lighting of the face, blurring, removal of the natural color of the face, or even replacing the face of a monkey with that of a human. Thus, at least some face-selective neurons appear to be sensitive to very general or global aspects of a face, such as the prototypical arrangement of the eyes and nose into a facelike configuration.

Subsets of Face-Selective Neurons Appear to Participate in Different Aspects of Face Coding

Many face-selective neurons are selective for the orientation of a face relative to the observer; cells tuned to front and profile views are most common. Other face-selective neurons have responses that appear to be specific for particular facial expressions, but which are insensitive to the individual or to other aspects of the face.[60,61] The reverse sort of behavior—face cells that are sensitive to the identity of an individual monkey or human, but which generalize across different facial expressions—has also been documented. However, these cells have been tested with relatively small numbers of different face images, and their preference for a given individual's face within a set of tested stimuli in no way provides evidence for the "grandmother cell" notion of perception. This notion holds that a very specific object, such as one's grandmother, is represented by the activity of a single, very selective neuron. Such neurons may merely be sensitive to one or more distinguishing features within faces. Nevertheless, such cells may well participate in *circuits* responsible for recognizing a particular individual. Some neurons in areas TE and STP do respond to specific face components, such as the presence of eyes per se, the distance between the eyes, or the extent of the forehead.[62] Finally, a subset of face cells that respond to whole faces is particularly sensitive to the direction of gaze of the eyes within a face (looking back or looking to the side), which is an important social signal in both monkeys and humans.[63]

Cells selectively responsive to some aspect of the facial image have a localized distribution within the anterior reaches of the ventral stream in several senses. First, although there are a small number of such neurons (1–5% of all cells recorded) throughout area TE and adjacent areas as a whole, their concentration is much higher in the cortex of the upper and lower banks of the superior temporal sulcus (parts of STP and TE, respectively), where they are found together in clumps and where they may make up as many as 10–20% of all neurons studied. Second, different types of face-selective cells are found in different regions: cells sensitive to facial expression and gaze direction tend to be found within the sulcus, whereas cells more generally selective for faces and for individuals tend to be located in area TE on the inferior temporal gyrus.

Do Face-Selective Cells Indicate a Specialized Neural Machinery?

Do face cells represent a specialized type of neural machinery that has evolved to deal with a specialized

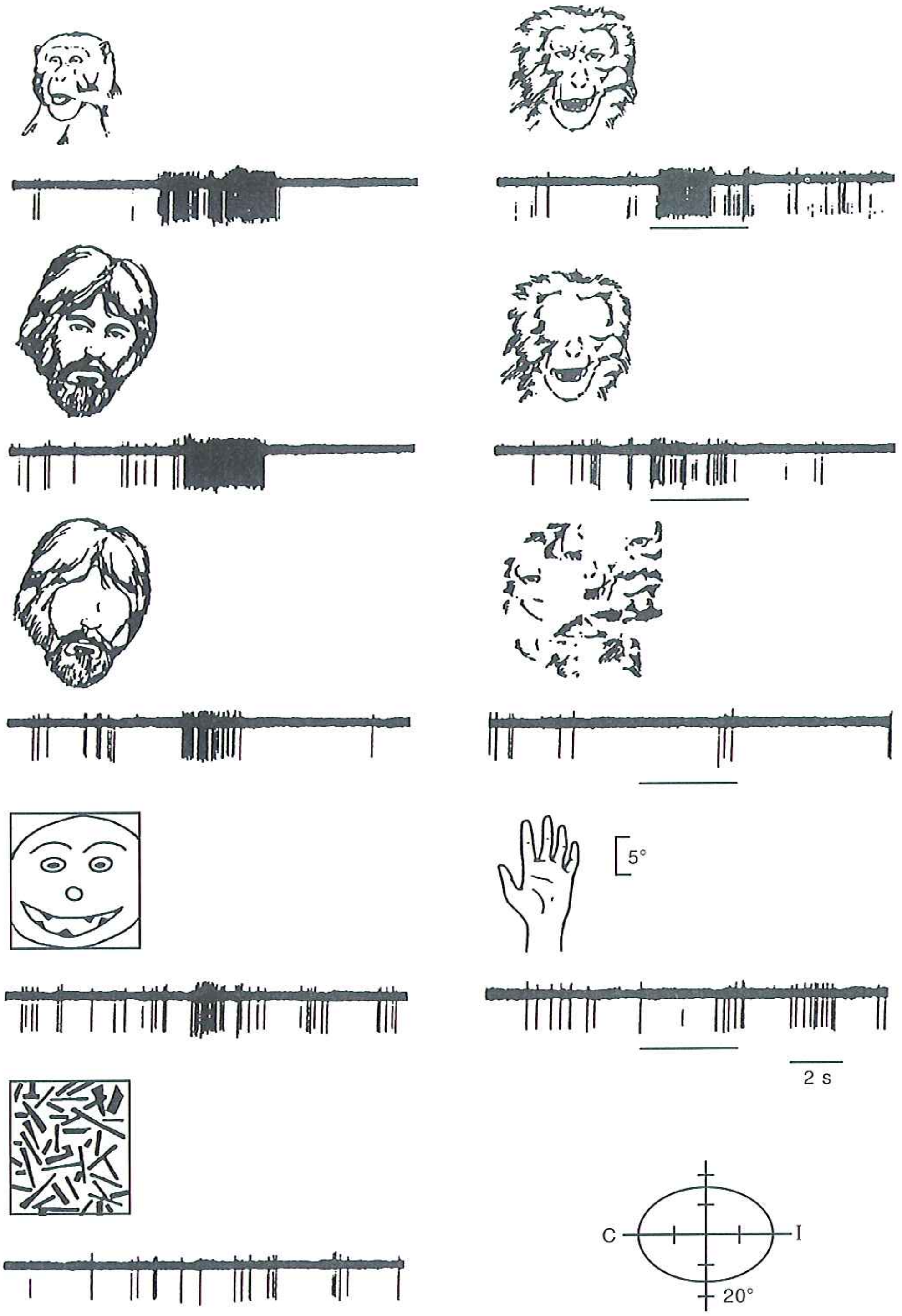


FIGURE 52.8 Activity of an STP neuron that responded better to faces than to all other stimuli tested. Removing eyes on a picture or representing the face as a caricature reduced the response. Cutting the picture into 16 pieces and rearranging pieces eliminated the response. Stimuli represented on the left were traced from a color photograph (monkey face), a black and white photograph (human face), and drawings (caricature and random pattern), which were swept across the fovea at 10 degrees/s. Stimuli represented on the right were traced from color slides, which were projected on the fovea for 3 s (indicated by the horizontal bars). All the unit records are representative ones chosen from a larger number of trials. The receptive field of the neuron is illustrated on the lower right. C, contralateral; I, ipsilateral visual field. From Bruce *et al.*[34]

type of visual stimulus? These neurons can be very selective, and comparably selective cells have yet to be found for other stimulus classes. Moreover, their presence very early in life[64] is consistent with the notion that they represent inborn templates for the detection of faces. However, most cells that respond selectively to faces or face components respond to a variety of individual faces and usually also respond, albeit to a much lesser degree, to other stimuli. In addition, like area TE cells responsive to other given stimulus features (e.g., texture and color), face-selective cells tend to be found together in clumps. Probably, then, the representation of faces, like that of other visual objects, involves a pattern of firing across populations of cells with varying selectivity and some local modular organization. Nevertheless, there may be important quantitative, if not qualitative, differences in the way the ventral stream deals with faces. For example, the large proportion of cells in the anterior portion of the ventral stream that respond to face stimuli, whether in a selective fashion or not, suggests that more cells here may participate in encoding faces than other types of stimuli, reflecting the behavioral significance of this stimulus type. Second, many of the populations or ensembles participating in face recognition may be relatively dedicated to this task, so that they may not participate in coding as many or as wide a variety of nonface patterns as other groups of neurons in the same regions.

Prosopagnosia Is a Specific Deficit of Face Recognition Caused by Brain Damage in Humans

Prosopagnosia, or face agnosia, a selective deficit in recognizing familiar faces, has been documented since at least the turn of the century. Although patients with this syndrome are aware that faces are faces—that is, they know the general type of stimulus category—they fail to reliably identify or even achieve a sense of familiarity from the faces of co-workers, family members, famous persons, and other individuals previously well known to them. Typically, they also have trouble forming memories of new faces, even if other new objects are learned. Accordingly, these persons may resort to remembering individuals by attending to salient details, such as their favorite clothing or their voices. Because the voice of a visually unrecognized person usually enables the patient to identify and feel familiar with that person, prosopagnosia appears to be a specifically visual impairment. A particularly illuminating description of a prosopagnosic patient was given by Pallis.[65] (See Box 52.2.)

A number of explanations for the apparent selectivity of the prosopagnosic deficit have been offered. One explanation is that face perception and recognition are indeed unique behavioral capacities and reflect unique, dedicated neural circuits that can be selectively damaged. A second explanation is that faces are processed and stored in a manner similar to that for other objects, but that the apparent selectivity of prosopagnosia is due to the fact that faces are simply harder to tell apart than other objects; this view is consistent with the observation that prosopagnosia is often accompanied by varying levels of object agnosia of other types. A third and related explanation suggests that face processing reflects subtle discriminations between highly similar exemplars within a category and that it is this general capacity, not the processing of the facial configuration per se, that is disrupted in prosopagnosia. Although there is undoubtedly some truth in these suggestions, one case study[66] has shown that, for at least one patient, processing of faces is still disproportionately impaired when face and nonface stimuli are equated for difficulty or when face discrimination is compared to discrimination of subtly differing exemplars within other classes, such as eyeglass frames or office chairs.

Still another explanation suggests that faces represent a class of stimuli that differ structurally from most other classes of objects, that such structural differences form the basis for different perceptual strategies for faces (such as encoding on the basis of a prototypical configuration), and that these strategies are lost in prosopagnosia. A final and intriguing suggestion, related to the last one, is that face processing represents the acquisition of a type of expertise derived from very protracted experience with a category of complex visual stimuli.[67] Reports of persons with prosopagnosia who have associated deficits in areas of object recognition in which they had previously acquired expertise over long periods of time (e.g., a show dog expert who lost the ability to differentiate breeds) are consistent with this idea.

Different Types of Prosopagnosia Reflect Different Component Subprocesses in Face Recognition

Different patterns of deficits in processing face material occur in prosopagnosia, and these differing patterns have been taken as evidence for a system of dissociable cognitive operations in processing face material by normal subjects. Some patients show considerable sparing of the ability to judge the age, gender, and even the emotional expression of faces whose identity they cannot recognize.[68] In other words, they show intact perception or structural encoding of face infor-

BOX 52.2

CASE STUDY OF PROSOPAGNOSIA

A.H., age 51, a mining engineer and colliery manager, was admitted to the Neurological Department, Cardiff Royal Infirmary, on January 4, 1954. On examination, mitral stenosis and auricular fibrillation were found. The ventricular rate was 70 to 80 per minute. There was a minimal pulse deficit. Blood pressure was 130/90 mm Hg. He was receiving "digoxin." There was no evidence of congestion and no signs of bacterial endocarditis.

He was right-handed. On neurological examination the only abnormalities detected were in the visual fields. There were narrow noncongruous homonymous sector defects in the left and right upper quadrants. (The patient was unaware of these defects.) Confrontation tests revealed no visual inattention. He localized objects accurately in both right and left half-fields. The patient was an excellent witness and gave consistent replies on repeated testing.

He was of above average intelligence and his general level of awareness was extremely keen. His memory was remarkable... His span of digit retention was eight forward and six backward. There was no hesitation in his speech, and he could obey complex orders. He read smoothly, and there was no trouble in understanding and later describing what he had read. He promptly recognized, named, and demonstrated the use of a wide variety of test objects. The significance of line drawings was immediately apparent to him, and he could accurately describe the content of various pictures he was shown.

He mixed readily with the other patients on the ward, but rarely spoke unless spoken to first. He could not identify his medical attendants. "You must be a doctor because of your white coat, but I don't know which one you are. I'll know if you speak." He failed to identify his wife during visiting hours. She was told one day, without his possible knowledge, to walk right past his bed, but he did not show the least sign of recognition. Repeated attempts were made to "catch him out" but none succeeded. If the disability was a feigned one, it was a performance of quite unbelievable virtuosity and consistency. He failed to identify pictures of Mr. Churchill, Mr. Aneurin Bevan, Hitler, Stalin, Miss Marilyn Monroe, or Mr. Groucho Marx. When confronted with such portraits he would proceed deductively, analyzing one feature after another, searching for the "critical detail" that would yield the answer. In human faces, this was rarely forthcoming. He had somewhat less difficulty with animal faces. A goat was eventually recognized by its ears and beard, a giraffe by its neck, a crocodile by its dentition, and a cat by its whiskers.

The patient had analyzed his difficulty in identifying faces with considerable insight. "I can see the eyes, nose, and mouth quite clearly, but they just don't add up. They all seem chalked in, like on a blackboard.... I have to tell by the clothes or by the voice whether it is a man or a woman.... The hair may help a lot, or if there is a mustache.

"At the club I saw someone strange staring at me, and asked the steward who it was. You'll laugh at me. I'd been looking at myself in a mirror." (Excerpted from Pallis.[65])

mation, and their inability to attach these percepts to the appropriate identity of the individual has been interpreted as a deficit in accessing the stored semantic or biographical memories associated with a given facial structure that make possible both identification and the feeling of familiarity. Other patients have severe difficulty with all aspects of face processing; such patients presumably are unable to analyze facial features normally, a necessary precondition to identification. Finally, some patients can perceive structural attributes adequately for judgments about emotion and gender and can even judge whether a face is familiar or not, but have a specific inability to recall the name associated with a familiar individual.[69]

Prosopagnosia Is Caused by Damage to the Ventral Occipitotemporal and Temporal Cortex

Initially, prosopagnosia of various types was associated clinically with right posterior cortical damage, and accordingly face recognition was believed to be dependent exclusively on structures within the posterior right hemisphere, at least in right-handers. In the 1980s, a number of cases came to autopsy. The damage common to these cases lay within the lingual and fusiform gyri, very ventrally and medially within the cortex at the occipitotemporal junction, in roughly the same region of the cortex as that activated in recent neuroimaging studies of object recognition (see Fig. 52.7B). However, in all such cases, this area (or at least the underlying white matter) was damaged bilaterally, and consequently bilateral damage has been thought by many to be a necessary precondition for prosopagnosia.[70] Several cases with prosopagnosia and right cortical damage alone, along with the results of imaging studies, have reopened the debate over the locus of the critical lesion in prosopagnosia. Some of the different patterns of deficits are undoubtedly due to different foci of damage.

A "Monkey Model" of Prosopagnosia Has Not Yet Been Achieved

Prosopagnosia has been conceptualized as the loss of a unique cognitive ability due to localized brain damage. It has often been further interpreted as reflecting the selective loss of ensembles of neurons homologous to face-selective neurons in monkey temporal cortex. Interestingly, the region containing the greatest concentration of face-selective neurons in monkeys—the superior temporal sulcus—is geographically quite distant from the very ventral cortices one would suppose to be homologous to the site of damage revealed in humans with prosopagnosia. Moreover, bilateral ablation of the superior temporal sulcus in monkeys clearly fails to produce more than minor face processing or face recognition impairments on a battery of tasks.[71] Even the minor impairments seen after such lesions—such as a deficit in judging the direction of gaze of a stimulus monkey—appear to be related to the type of behavioral task used and do not appear to be specific to facial material.[72] Face-selective neurons located in more lateral and ventral portions of the temporal lobe (e.g., on the inferior temporal gyrus within area TE) therefore also probably play an important role in face recognition. However, removal of this cortex causes severe impairments not only in face perception but also in perception and recognition of a large number of other classes of objects. One possibility is that the neural substrates for recognizing face and nonface stimuli are not segregated as completely in monkeys as in humans. An alternative possibility is that the still poorly understood ventromedial cortices of monkeys (particularly parahippocampal areas TF and TH) may play a selective role in face recognition that has not yet been appreciated.

Substrates of Dissociable Aspects of Face Recognition in Human Cortex Are Revealed by Functional Imaging and Recording Methods

Neuroimaging studies in normal humans provide strong support both for the involvement of the ventral stream in face recognition and for the existence of dissociable component operations within face processing that are differentially linked to different loci within this pathway and to the two hemispheres. For example, using PET methods, researchers have compared regional activation in normal subjects who were required either to determine the gender of stimulus faces or to judge their identity.[54] In the gender discrimination condition, presumably a perceptual task of facial features, selective activation was found in the right ventral occipitotemporal cortex, to a lesser extent in the same area on the left side, and in a more lateral left focus as well (see foci "F" in Fig. 52.7). These areas overlapped, but were at least partially more anterior to the domains activated by processing of other categories of objects. Judgments of face identity, on the other hand, which presumably require reactivation of specific stored information about individuals, also activated more anterior portions of ventral extrastriate cortex, including the right parahippocampal gyrus, and anterior temporal cortex and the temporal pole on both sides. These studies and those of Haxby and colleagues[53] additionally implicated a portion of the right ventral frontal lobe in perceptual judgments about faces, as well as in working memory for faces.[73] Neuroimaging work also suggests that the frontal cortex is involved in recognizing facial emotion.[74] These functional neuroimaging studies thus show a hierarchical segregation of areas involved with different components of face recognition in humans. The ventral occipitotemporal cortex of the right hemisphere appears to be primarily responsible for the perceptual analysis of faces. Information then appears to be sent forward to more anterior temporal regions for association with stored information about individuals, and frontal cortex appears to contribute both to the analysis of emotional facial content and to holding faces in a short-term memory store. Finally, lesion evidence[69] corroborates the idea that the left lateral focus seen in PET studies of judging face identity does indeed reflect a specialization for attaching stored verbal information (i.e., names) to face percepts.

Electrophysiological correlates of face recognition have also been obtained from patients who have implanted electrodes as part of the preparation for neurosurgery to treat epilepsy (Fig. 52.9).[56,75] A large evoked potential (called the *N200*) is generated by faces (but not by other categories of stimuli) at small sites (i.e., typically at only one electrode within an electrode array) in the ventral occipitotemporal cortex. These small sites, or modules, which vary in their exact location among individuals, are reminiscent of clumps of face neurons found in monkeys. Longer-latency face-specific potentials have also been recorded from the

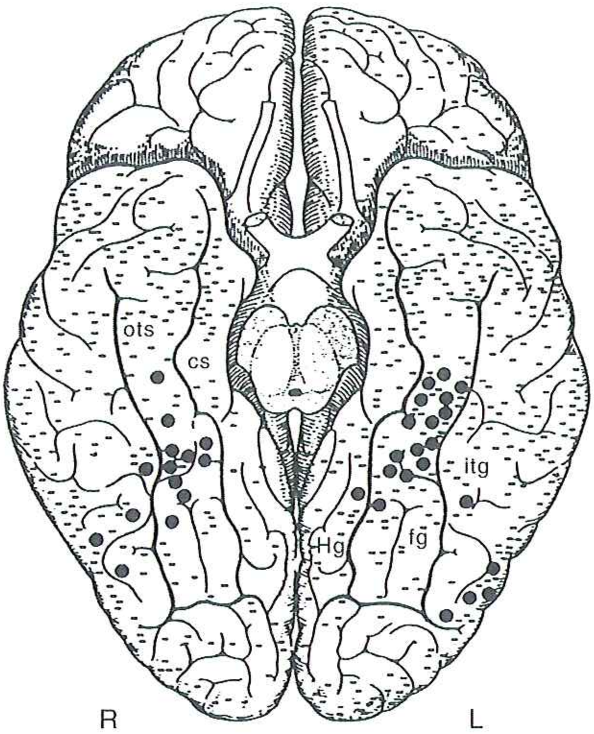


FIGURE 52.9 Summary figure of locations (black dots) on the ventral surface of the human brain from which a surface-negative potential (N200) was recorded only when patients were shown faces and not when other categories of objects were tested. Locations are primarily along the fusiform gyrus. Dashes illustrate other sites tested that were not face-selective. cs, collateral sulcus; fg, fusiform gyrus; itg, inferior temporal gyrus; lg, lingual gyrus; ots, occipitotemporal sulcus. From Allison *et al.*[75]

more anterior portions of inferior temporal cortex that were activated by face recognition in the PET studies. Although sites producing the N200 potential were found bilaterally, there was an interesting hemispheric difference. In normal humans, the right hemisphere tends to process faces as coherent wholes, and accordingly stimuli restricted to the left visual field are processed more slowly when they are presented upside-down, so that the prototypical configuration of features is disrupted. When stimuli are presented to the left hemisphere via the right visual field, no such **inversion effect** is found. Correspondingly, N200 potentials evoked by inverted faces (presented at the center of the visual field) and recorded from the right hemisphere were smaller and longer than those evoked by normally oriented faces; the left hemisphere generated comparable N200s under both conditions. Because of this correspondence between electrophysiology and behavior, the analysis of evoked potentials appears to be a valuable adjunct to other types of functional methods for studying the basis of face and object recognition in humans.

Summary

Striking parallels exist among the results of neurophysiological, neuropsychological, perceptual, and functional neuroanatomical studies of face recognition in primates. These lines of research all indicate that face recognition is a hierarchically organized set of component processes with substrates in the ventral portions of the visual cortex. In both humans and monkeys, faces are processed in subregions of the ventral stream that are at least partially dissociable from those processing other types of visual objects, or the color of objects in isolation. Consequently, the question is now whether the neural machinery dedicated to face recognition represents an adaptation for the processing of faces or for some more general and abstract type of object processing (such as configurational information), of which face recognition is an arbitrary example. In both monkeys and humans, face-processing circuits appear to be localized to relatively small modules, a situation that may reflect a more general tendency of mammalian neocortex to develop category-specific **modules** for behaviorally important classes of stimuli. If so, face recognition could be a valuable model for the understanding of basic mechanisms of object recognition, rather than being a special case.

THEORETICAL ACCOUNTS OF OBJECT RECOGNITION

Biological visual systems are highly successful. Noisy images can be taken as input and, despite the relatively slow firing rate of neurons, objects can be recognized from the images within a few hundred milliseconds.[76] Even the most sophisticated artificial vision systems have still not achieved these capabilities. Yet the goals of biological and artificial visual systems are, in their broadest terms, very similar. For both systems to be successful, processes capable of deriving descriptions of objects from noisy input and of assigning varying input to common categories must be present so that objects are recognized across wide variations in image properties (position, viewing angle, distance, and size). Because of these common goals, theories can address common problems relevant to both biological and artificial visual systems. Full accounts of biological visual systems, however, will also need to explain how the processes of object recognition are implemented in neural terms. Here we discuss several major theories of object recognition, drawing on studies of both artificial and biological visual systems to provide an account of the general processes involved. In addition, we consider proposals about their neural implementation.

Three-Dimensional Representations Arise from "Bottom-Up" Processes (Marr's Approach to Object Recognition)

Perhaps the most influential theoretical account of the processes leading to visual object recognition over the past 15 years is that of David Marr.[77,78] Marr argued that object recognition involves the coding of a series of different representations of objects, with each representation making different kinds of information explicit. Marr's framework starts from the idea that the information used for object recognition is recovered from images on the retina by locating and describing the places where the intensity of the image changes relatively abruptly. These abrupt changes in image intensity signal the presence of edges in the world and can be used to derive useful three-dimensional (3D) attributes such as surface markings on objects, object boundaries, and shadows. Marr proposed that these attributes are specified in something he termed a *primal sketch*. He distinguished two stages of deriving the primal sketch: (1) the raw primal sketch, involving coding and locating individual intensity changes within a map based on the retina; and (2) the full primal sketch, involving a description of the image that is more elaborate, with individual edge fragments being grouped into more meaningful clusters relating to surfaces. Grouping is based on simple physical laws of perceptual organization—for instance, that image elements that are closer or more similar will be grouped more strongly than those that are farther apart and dissimilar. The primal sketch makes explicit information about the presence of two-dimensional (2D) intensity changes in the image. Implicit within this representation is information about the 3D nature of the image features or about the invariant properties of those features across different retinal locations.

At the next stage of coding in Marr's framework, information about the distance and layout of each surface is added, using a variety of depth cues. This representation is termed the *$2\frac{1}{2}$D sketch* because it describes only the visible parts of the scene and thus is not fully 3D. The $2\frac{1}{2}$D sketch may be useful for matching actions to objects, because it specifies information about the relative depths of surfaces with respect to the viewer. However, the representation may be problematic for object recognition. To recognize an object, there needs to be contact between the representation encoded and a stored memory for the object concerned. The $2\frac{1}{2}$D sketch for a given object is specific to the particular viewing conditions; for example, if the observer moves position, or if the object is presented at a different angle, then a different $2\frac{1}{2}$D sketch will be derived. For such a representation to serve for object recognition, stored representations for all views of objects would need to exist. This introduced problems of memory storage and of matching. That is, given the immense number of memory representations needed, how can a given image description be matched to an appropriate memory representation? To lessen these problems, Marr suggested that a further representation—the *3D model description*—be constructed.

In the 3D model representation, the parts of the object are coded relative to some salient part of the object itself (e.g., relative to the main axis of the object). Because the parts of objects generally remain in the same location relative to the main axis of the object, the 3D model description remains the same when objects are presented in different views—provided that the same main axis can be recovered in the different images. Consequently, only a single 3D model representation may need to be stored for each known object. Also, for objects to be identified across different viewpoints, the stored 3D model description needs to specify all the parts of an object, even if the parts are not visible in a given image; hence it is truly 3D (unlike the $2\frac{1}{2}$D sketch). Marr further suggested that the parts coded within the 3D model description take the form of **generalized cones.** Generalized cones are the shapes formed by expanding a constant shape in 3D along a main axis—for instance, a vase can be formed from a circular cross section that expands (e.g., at the base) and contracts (e.g., at the stem) along a main axis. Such generalized cone representations makes explicit information about the 3D structure of objects. Figure 52.10 illustrates Marr's framework.

Marr's approach to object recognition stressed that recognition involves **bottom-up processes,** which extract information from the image and build useful descriptions, without incorporating procedures specific to the individual objects concerned. Object descriptions are derived using rules that make general assumptions

FIGURE 52.10 Marr's framework for visual object recognition.

about the nature of the visual world, but do not call on knowledge about specific objects. Marr's approach contrasts with **top-down approaches,** which suggest that knowledge about specific objects is used to facilitate early stages of object processing. As we shall see, a role for top-down mechanisms is favored in some theories dealing with how object recognition might be realized in networks using neuronlike processing units.

Object Recognition Is Achieved by Encoding Relations between Component Parts (Biederman's Approach)

Biederman and colleagues[79–81] have proposed a somewhat different approach to visual object recognition. This approach differs from the one advocated by Marr in at least two significant ways: One is that, according to Biederman, the component parts of objects can be detected directly from the presence of appropriate 2D edge descriptors in images, without it being necessary to encode properties of the surfaces of objects (as would be the case when a 2½D sketch is derived en route to the construction of a 3D model description). The second difference is that recognition involves coding the spatial relations between the component parts, but without the parts necessarily being encoded with respect to a more global property of the whole object, such as its main axis.

Biederman argues that a finite number of volumetric component parts—he terms these "geons," akin to letters of an alphabet—when combined, form many of the objects that we know (Fig. 52.11).[82] Geons can be encoded directly from the presence of "nonaccidental" relations between 2D edge descriptors. Nonaccidental relations are those that are highly unlikely to occur by chance, given random noise in an image[83]; they include relations such as parallel or co-linear edges and closure. The presence of such relations can be used to infer directly the presence of particular geons. Also, because the same nonaccidental features can be derived across a wide range of image variations, the encoding of geons is relatively robust to changes in viewpoint (viewing angle, position, distance, etc.).

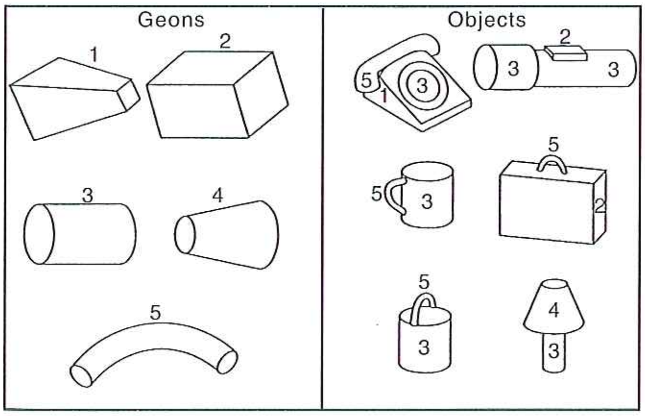


FIGURE 52.11 Illustration of geons and how they are arranged to form objects. (Left) A given view of an object can be represented by an arrangement of simple primitive volumes, or geons, five of which are shown here. (Right) Only two or three geons are required to uniquely specify an object. The relations among the geons matter, as illustrated with the pail and cup. From Biederman.[82]

Experimental Evidence Shows That Both Marr's and Biederman's Approaches Have Validity

Evidence that the spatial relations between component parts are important for object recognition comes from work showing that recognition is impaired when objects are broken into their parts and the spatial relations between the parts are scrambled.[84] Biederman[79] also showed that recognizing line drawings of objects was especially difficult when fragments of the drawings were removed to eliminate regions of discontinuity that might be used to parse the objects into their components (e.g., if the corner regions between the components were removed). This result suggests again that parsing of objects at regions of discontinuity is important for object recognition, although it does not illuminate the nature of the components subsequently encoded.

The approaches of Marr and Biederman differ in the role played by an object's main axis in object recognition. Marr assigns the axis a special role in enabling recognition to become relatively invulnerable to the effects of viewpoint change, whereas Biederman assigns no such special role to the main axis of an object. Biederman and Gerhardstein[81] have reported that large depth rotations produce relatively small effects on object recognition and also that such rotations do not strongly affect the priming effect of one view of an object on the object's subsequent recognition. Strong effects of depth rotation might be expected if the rotations lead to foreshortening of the main axis of the object and if foreshortening is particularly damaging for recognition. On the other hand, researchers[85] have reported that foreshortening disrupts identification even when all the main component parts of objects remain visible. In addition, identification of foreshortened objects is facilitated when the objects are photographed against a textured background giving strong linear perspective cues to the depth of the main axis of the object.[86] These apparently contradictory results may reflect the difficulty of defining the precise component parts used for object recognition, or it may be that recognition is relatively invulnerable to rotations in depth unless the rotations produce severe foreshortening (e.g., when the real main axis of the object is no longer the depicted main axis); evidence[87] indicates such nonlinear effects of foreshortening on people's ability to match different views of objects, thus supporting this view. It may also be that specific component parts and a more global representation of the parts in relation to the main axis of the object can both be used for object recognition[88]; which description is dominant may depend on the task and the context of the discrimination (e.g., whether other similar stimuli may also be presented).

The approaches of Marr and Biederman also differ with respect to the role of surface information in object recognition. Within Marr's framework, surface information can directly affect object recognition, because the $2\frac{1}{2}$D sketch (specifying surface details) is constructed before the 3D model description of an object is generated. In contrast, Biederman's approach emphasizes the role of edge-based information in object recognition. Once again, both approaches appear to be valid to some degree. Biederman and Ju[89] failed to find strong effects of surface information on object recognition and reported that line drawings of objects could be recognized as efficiently as full-color photographs. However, this result depends to some degree on the stimuli used. The effects of surface information on the recognition of man-made objects (*artifacts*) are relatively slight, but are more substantial on the recognition of living things.[90] Living things, as a class, tend to be more visually similar to one another than artifacts and they also tend to have diagnostic colors; surface details appear to be useful when recognition depends on fine-grained discrimination between visually similar items and when surface details are diagnostic for the particular objects involved. The evidence suggests that stored object representations can be accessed directly from edge descriptors, but surface-based representations can contribute when edge-based recognition is relatively inefficient.

Connectionist Approaches Include "Top-Down" Processes for Object Recognition

So far, we have discussed approaches to object recognition that emphasize bottom-up mechanisms, which do not involve knowledge about specific objects in the processes leading up to object recognition. Some attempts to implement object recognition in networks using neuronlike units (so-called connectionist models) move away from this approach, to allow stored knowledge to be used in a top-down manner to guide the recognition process. For instance, Edelman and colleagues[91,92] have trained neural network models (i.e., computer simulations of hypothetical networks constructed of neuronlike units) to recognize wire-framed objects on the basis of a stored set of 2D views. In such models, many different stored representations are formed for a given object, and recognition of the same object from a new view involves interpolation between the stored views. Not only were the models able to recognize the objects they were trained on, but they generated relatively accurate predictions about the success of human recognition of the same objects when rotated. Subjects trained with a set of views in one plane were then presented with the objects either within the same plane of orientation as the old objects (but in new positions) or in a plane orthogonal to that of the old exemplars. There was poor generalization of learning to objects presented in a new, orthogonal plane and much better generalization to objects presented at the new locations in the "old" plane. Bulthoff and Edelman[93] suggest that recognition based on a 3D model of the objects should generalize equally well to objects presented in the old plane and to those presented in the orthogonal plane. In contrast, the finding of a better generalization within the old plane is consistent with the involvement of interpolation based on 2D representations, which is appropriate for stimuli presented in the same plane as the previously presented objects, but not for objects presented in a new plane.

These last results argue against theoretical approaches deriving from Marr's work, which emphasize the utility of having only one stored representation for each known object, along with complex coding procedures to provide accurate access to that object representation. How far Edelman's approach can be taken and whether it will generalize to cover more than the recognition of wire figures are questions awaiting further research.

There Are Open Questions about Objects versus Faces

Most theoretical accounts of object recognition do not make qualitative distinctions between the processes involved in the recognition of different types of objects; nevertheless, neuroscientific evidence suggests that recognition processes can differ for different stimuli.[94] Indeed, surface properties of objects may play a more important role in the recognition of objects belonging to visually homogeneous classes of objects (e.g., living things) than in the recognition of objects from visually more heterogeneous classes (e.g., artifacts). Faces, as a class, are probably more homogeneous than any other general class of visual stimulis, at least with regard to classes of stimuli for which we are able to identify individual exemplars. Does this visual homogeneity mean that face recognition is dependent on the representation of surface features to a greater extent than is the recognition of other types of objects? And, in face recognition, are these surface-based representations differentiated into a parts-based description of the object?[94] These detailed points require further research.

Summary

Lesion studies in monkeys and humans have shown that the ability to identify and recognize objects depends on the inferior temporal cortex. Visual information is relayed to the inferior temporal cortex via a ventrally directed cortical pathway, or "stream," consisting of several functionally distinct visual areas. At successive stations of this ventral stream, the stimulus representation is progressively transformed from one specifying retinal image properties toward one emphasizing intrinsic object shape properties. Functional brain imaging and electrical recording studies in humans suggest that face recognition may involve a specialized neural machinery separate from that responsible for the recognition of other objects. However, it is still unclear whether such a specialized machinery represents an adaptation for the processing of faces per se (e.g., through very protracted experience with faces leading to expert knowledge) or an adaptation for some more abstract type of object processing (e.g., a surface-based representation), of which faces are an especially good example. Theoretical accounts of object recognition have included those of Marr and of Biederman, both of which stress bottom-up mechanisms, as well as that of Edelman and colleagues, in which object recognition is implemented in a network of neuronlike units. One advantage of this latter, connectionist approach is that it allows stored knowledge to be used in a top-down manner to guide the bottom-up recognition process.

References

1. Farah, M. J. (1990). *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision.* MIT Press, Cambridge, MA.
2. Lissauer, H. (1890). Ein fall von seelenblindheit nebst einem beitrage zur theorie derselben. *Arch. Psychiatr. Nervenkr.* **21:** 222–270.
3. Levine, D. N. (1982). Visual agnosia in monkey and man. In *Analysis of Visual Behavior* (D. I. Ingle, M. A. Goodale, and R. J. W. Mansfield, eds.), pp. 629–670. MIT Press, Cambridge, MA.
4. Potzl, O. (1928). *Die Aphasielehre vom Standpunkte der kliniscen Psychiatrie.* Franz Deudicte, Lepzig.
5. Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. *Cognit. Neuropsychol.* **8:** 1–19.
6. Feinberg, T. E., Schindler, R. J., Ochoa, E., Kwan, P. C., and Farah, M. J. (1994). Associative visual agnosia and alexia without prosopagnosia. *Cortex* **30:** 395–411.
7. Klüver, H., and Bucy, P. C. (1937). "Psychic blindness" and other symptoms following bilateral temporal lobectomy in rhesus monkeys. *Am. Physiol. Soc.*, pp. 352–353.
8. Mishkin, M. (1954). Visual discrimination performance following partial ablations of the temporal lobe. II. Ventral surface vs. hippocampus. *J. Comp. Physiol. Psychol.* **47:** 187–193.
9. Mishkin, M. (1966). Visual mechanisms beyond the striate cortex. In *Frontiers in Physiological Psychology* (R. W. Russell, ed.), pp. 93–119. Academic Press, New York.
10. Mishkin, M., and Pribram, K. H. (1954). Visual discrimination performance following partial ablations of the temporal lobe. I. Ventral vs. lateral. *J. Comp. Physiol. Psychol.* **47:** 14–20.
11. Pribram, K. H. (1955). Toward a science of neuropsychology: Method and data. In *Current Trends in Psychology and the Behavioral Sciences* (J. T. Wilson, C. S. Ford, B. F. Skinner, G. Bergmann, F. A. Beach, and K. Pribram, eds.), pp. 115–142. University of Pittsburgh, Pittsburgh, PA.
12. Dean, P. (1982). Visual behavior in monkeys with inferotemporal lesions. In *Analysis of Visual Behavior* (D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield, eds.), pp. 587–628. MIT Press, Cambridge, MA.
13. Gross, C. G., and Mishkin, M. (1977). The neural basis of stimulus equivalence across retinal translation. In *Lateralization in the Nervous System* (S. R. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, and G. Krauthamer, eds.), pp. 109–122. Academic Press, New York.
14. Seacord, L., Gross, C. G., and Mishkin, M. (1979). Role of inferior temporal cortex in interhemispheric transfer. *Brain Res.* **167:** 259–272.
15. Humphrey, N. K., and Weiskrantz, L. (1969). Size constancy in monkeys with inferotemporal lesions. *Q. J. Exp. Psychol.* **21:** 225–238.
16. Weiskrantz, L., and Saunders, R. C. (1984). Impairments of visual object transforms in monkeys. *Brain* **107:** 1033–1072.
17. Rubens, A. B., and Benson, D. F. (1971). Associative visual agnosia. *Arch. Neurol. (Chicago)* **24:** 305–316.
18. Bay, E. (1953). Disturbances of visual perception and their examination. *Brain* **76:** 515–550.
19. Bender, M. B., and Feldman, M. (1972). The so-called "visual agnosias." *Brain* **95:** 173–186.
20. Ungerleider, L. G., and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior* (D. Ingle, M. A. Goodale, and R. J. W. Mansfield, eds.), pp. 549–586. MIT Press, Cambridge, MA.
21. von Bonin, G., and Bailey, P. (1947). *The Neocortex of Macaca mulatia.* University of Illinois Press, Urbana, IL.
22. Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends Neurosci.* **6:** 415–417.
23. Gross, C. G., Rodman, H. R., Gochin, P. M., and Colombo, M. W. (1993). Inferior temporal cortex as a pattern recognition device. In *Computational Learning and Cognition* (E. Baum, ed.), pp. 44–73. SIAM Press, Philadelphia.
24. Baizer, J. S., Ungerleider, L. G., and Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J. Neurosci.* **11:** 168–190.
25. Van Essen, D. C. (1985). Functional organization of primate visual cortex. In *Cerebral Cortex* (A. Peters and E. G. Jones, eds.), pp. 259–329. Plenum, New York.

26. Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1:** 1–47.
27. Lewis, M. E., Mishkin, M., Bragin, E., Brown, R. M., Pert, C. B., and Pert, A. (1981). Opiate receptor gradients in monkey cerebral cortex: Correspondence with sensory processing hierarchies. *Science* **211:** 1166–1169.
28. Nelson, R. B., Friedman, D. P., O'Neill, J. B., Mishkin, M., and Routtenberg, A. (1987). Gradients of protein kinase C substrate phosphorylation in primate visual system peak in visual memory storage areas. *Brain Res.* **28:** 387–392.
29. Desimone, R., and Ungerleider, L. G. (1989). Neural mechanisms of visual processing monkeys. In *Handbook of Neuropsychology* (F. Boller and J. Grafman, eds.), pp. 267–299. Elsevier, Amsterdam.
30. Boussaoud, D., Desimone, R., and Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *J. Comp. Neurol.* **306:** 554–575.
31. Fenstemaker, S. B., Albright, T. D., and Gross, C. G. (1985). Organization and neuronal properties of visual area TEO. *Soc. Neurosci. Abstr.* **11:** 1012.
32. Gattass, R., Gross, C. G., and Sandell, J. H. (1981). Visual topography of V2 in the macaque. *J. Comp. Neurol.* **201:** 519–539.
33. Gattass, R., Sousa, A. P. B., and Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* **8:** 1831–1856.
34. Bruce, C., Desimone, R., and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46:** 369–384.
35. Mishkin, M., and Murray, E. A. (1994). Stimulus recognition. *Curr. Opin. Neurobiol.* **4:** 200–206.
36. Baizer, J. S., Robinson, D. L., and Dow, B. M. (1977). Visual responses of area 18 neurons in awake, behaving monkey. *J. Neurophysiol.* **40:** 1024–1037.
37. Burkhalter, A., and Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J. Neurosci.* **6:** 2327–2351.
38. von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science* **224:** 1260–1262.
39. Desimone, R., and Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *J. Neurophysiol.* **57:** 835–868.
40. Gallant, J. L., Braun, J., and Van Essen, D. C. (1993). Selectivity for polar, hyperbolic and Cartesian gratings in macaque visual cortex. *Science* **259:** 100–103.
41. Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* **71:** 856–867.
42. Zeki, S. (1983). Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience* **9:** 767–781.
43. Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science* **229:** 782–784.
44. Haenny, P. E., Maunsell, J. H., and Schiller, P. H. (1988). State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Exp. Brain Res.* **69:** 245–259.
45. Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **4:** 2051–2062.
46. Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J. Neurophysiol.* **66:** 170–189.
47. Sato, T., Kawamura, T., and Iwai, E. (1980). Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination. *Exp. Brain Res.* **38:** 313–319.
48. Sary, G., Vogels, S. R., and Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* **260:** 995–997.
49. Li, L., Miller, E. K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* **69:** 1918–1929.
50. Miller, E. K., Li, L., and Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J. Neurosci.* **13:** 1460–1478.
51. Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature (London)* **335:** 817–820.
52. Ungerleider, L. G., and Haxby, J. V. (1994). "What" and "where" in the human brain. *Curr. Opin. Neurobiol.* **4:** 157–165.
53. Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., and Grady, C. L. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *J. Neurosci.* **14:** 6336–6353.
54. Sergent, J., Ohta, S., and MacDonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain* **115:** 15–36.
55. Rodman, H. (1998). Temporal cortex. *Encycl. Neurosci.* (in press).
56. Allison, T., McCarthy, C., Nobre, A., Puce, A., and Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb. Cortex* **4:** 544–554.
57. Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* **35:** 96–111.
58. Konorski, J. (1967). *Integrative Activity of the Brain.* University of Chicago, Chicago.
59. Gross, C. G., and Sergent, J. (1992). Face recognition. *Curr. Opin. Neurobiol.* **2:** 156–161.
60. Baylis, G. C., Rolls, E. T., and Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res.* **342:** 91–102.
61. Hasselmo, M. E., Rolls, E. T., and Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* **32:** 203–218.
62. Yamane, S., Kaji, S., and Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp. Brain Res.* **73:** 209–214.
63. Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., and Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. London B* **223:** 293–317.
64. Rodman, H. R., Scalaidhe, S. P., and Gross, C. G. (1993). Response properties of neurons in temporal cortical visual areas of infant monkeys. *J. Neurophysiol.* **70:** 1115–1136.
65. Pallis, C. A. (1955). Impaired identification of faces and places with agnosia for colors. *J. Neurol. Neurosurg. Psychiatry* **18:** 218–224.
66. Farah, M. J., Levinson, K. L., and Klein, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia* **33:** 661–674.
67. Diamond, R., and Carey, S. (1986). Why faces are and are not special: An effect of expertise. *J. Exp. Psychol. Gen.* **115:** 107–117.

68. Tranel, D., Damasio, A. R., and Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology* **38:** 690–696.

69. Flude, B. M., Ellis, A. W., and Kay, J. (1989). Face processing and name retrieval in an anemic aphasic: Names are stored separately from semantic information about familiar people. *Brain Cognit.* **11:** 60–72.

70. Damasio, A. R., Tranel, D., and Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annu. Rev. Neurosci.* **13:** 89–109.

71. Heywood, C. A., and Cowey, A. (1992). The role of the "face cell" area in the discrimination and recognition of faces by monkeys. *Philos. Trans. R. Soc. London B* **335:** 31–37.

72. Eacott, M. J., Heywood, C. A., Gross, C. G., and Cowey, A. (1993). Visual discrimination impairments following lesions of the superior temporal sulcus are not specific for facial stimuli. *Neuropsychologia* **31:** 609–619.

73. Courtney, S. M., Underleider, L. G., Keil, K., and Haxby, J. V. (1997). Transient and sustained activity in a neural system for human working memory. *Nature* **386:** 608–611.

74. George, M. S., Ketter, T. A., Gill, D. S., Haxby, J. V., Ungerleider, L. G., Herscovitch, P., and Post, R. M. (1993). Brain regions involved in recognizing facial emotion or identity: An oxygen-15 PET study. *J. Neuropsychiatry Clin. Neurosci.* **5:** 384–394.

75. Allison, T., Ginter, H., McCarthy, G., Nobre, A., Puce, A., Luby, M., and Spencer, D. D. (1994). Face recognition in human extrastriate cortex. *J. Neurophysiol.* **71:** 821–825.

76. Thorpe, S. J., and Imbert, M. (1989). Biological constraints on connectionist modelling. In *Connectionism in Perspective* (R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, and L. Steels, eds.), pp. 63–92. Elsevier/North-Holland, Amsterdam.

77. Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman, San Francisco.

78. Marr, D., and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. London B* **200:** 269–294.

79. Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychol. Rev.* **94:** 115–147.

80. Biederman, I., and Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognit. Psychol.* **23:** 393–419.

81. Biederman, I., and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *J. Exp. Psychol. Hum. Percept. Perform.* **19:** 1162–1182.

82. Biederman, I. (1990). Higher-level vision. In *Visual Cognition and Action* (D. H. Osherson, S. M. Kosslyn, and J. M. Hollerbach, eds.), Vol. 2, pp. 41–72. MIT Press, Cambridge, MA.

83. Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artif. Intell.* **31:** 355–395.

84. Cave, C. B., and Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception* **22:** 229–248.

85. Humphreys, G. K., and Jolicoeur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object identification. *Q. J. Exp. Psychol.* **A 46:** 137–159.

86. Humphreys, G. W., and Riddoch, M. J. (1984). Routes to object constancy: Implications from neurological impairments of object constancy. *Q. J. Exp. Psychol.* **A 36:** 385–415.

87. Lawson, R., and Humphreys, G. W. (1998). View-specificity in object processing: Evidence from picture matching. *J. Exp. Psychol. Hum. Percept. Perform.* (in press).

88. Jolicoeur, P. (1992). Identification of disoriented objects: A dual-systems theory. In *Understanding Vision: An Interdisciplinary Perspective* (G. W. Humphreys, ed.), pp. 180–198. Blackwell, Oxford.

89. Biederman, I., and Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognit. Psychol.* **20:** 38–64.

90. Price, C. J., and Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Q. J. Exp. Psychol.* **A 41:** 797–827.

91. Poggio, T., and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature* (*London*) **343:** 263–266.

92. Edelman, S., and Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biol. Cybernet.* **64:** 209–219.

93. Bulthoff, H. H., and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. U.S.A.* **89:** 60–64.

94. Bruce, V., and Humphreys, G. W. (1994). Recognizing objects and faces. *Visual Cognit.* **2/3:** 141–180.