

High-Level Visual Processing: From Vision to Cognition

High-Level Visual Processing Is Concerned With Object Recognition

The Inferior Temporal Cortex Is the Primary Center for Object Recognition

Clinical Evidence Identifies the Inferior Temporal Cortex as Essential for Object Recognition

Neurons in the Inferior Temporal Cortex Encode Complex Visual Stimuli and Are Organized in Functionally Specialized Columns

The Primate Brain Contains Dedicated Systems for Face Processing

The Inferior Temporal Cortex Is Part of a Network of Cortical Areas Involved in Object Recognition

Object Recognition Relies on Perceptual Constancy

Categorical Perception of Objects Simplifies Behavior

Visual Memory Is a Component of High-Level Visual Processing

Implicit Visual Learning Leads to Changes in the Selectivity of Neuronal Responses

The Visual System Interacts With Working Memory and Long-Term Memory Systems

Associative Recall of Visual Memories Depends on Top-Down Activation of the Cortical Neurons That Process Visual Stimuli

Highlights

AS WE HAVE SEEN, LOW-LEVEL visual processing is responsible for detecting various types of contrasts in the patterns of light projected onto the retina. Intermediate-level processing is concerned with

the identification of so-called visual primitives, such as contours and fields of motion, and the segregation of surfaces. High-level visual processing integrates information from a variety of sources and is the final stage in the visual pathway leading to visual perception.

High-level visual processing is concerned with identifying behaviorally meaningful features of the environment and thus depends on descending signals that convey information from short-term working memory, long-term memory, and executive areas of cerebral cortex.

High-Level Visual Processing Is Concerned With Object Recognition

Our visual experience of the world is fundamentally object-centered. We can recognize the same object even when the patterns of light it casts onto the retina vary greatly with viewing conditions, such as lighting, angle, position, and distance. And this is the case even for visually complex objects, those that include a large number of conjoined visual features.

Moreover, objects are not mere visual entities, but are commonly associated with specific experiences, other remembered objects, and sensations—such as the hum of the coffee grinder or the aroma of a lover’s perfume—and a variety of emotions. It is the behavioral significance of objects that guides our action based on visual information. In short, object recognition establishes a nexus between vision and cognition (Figure 24–1).

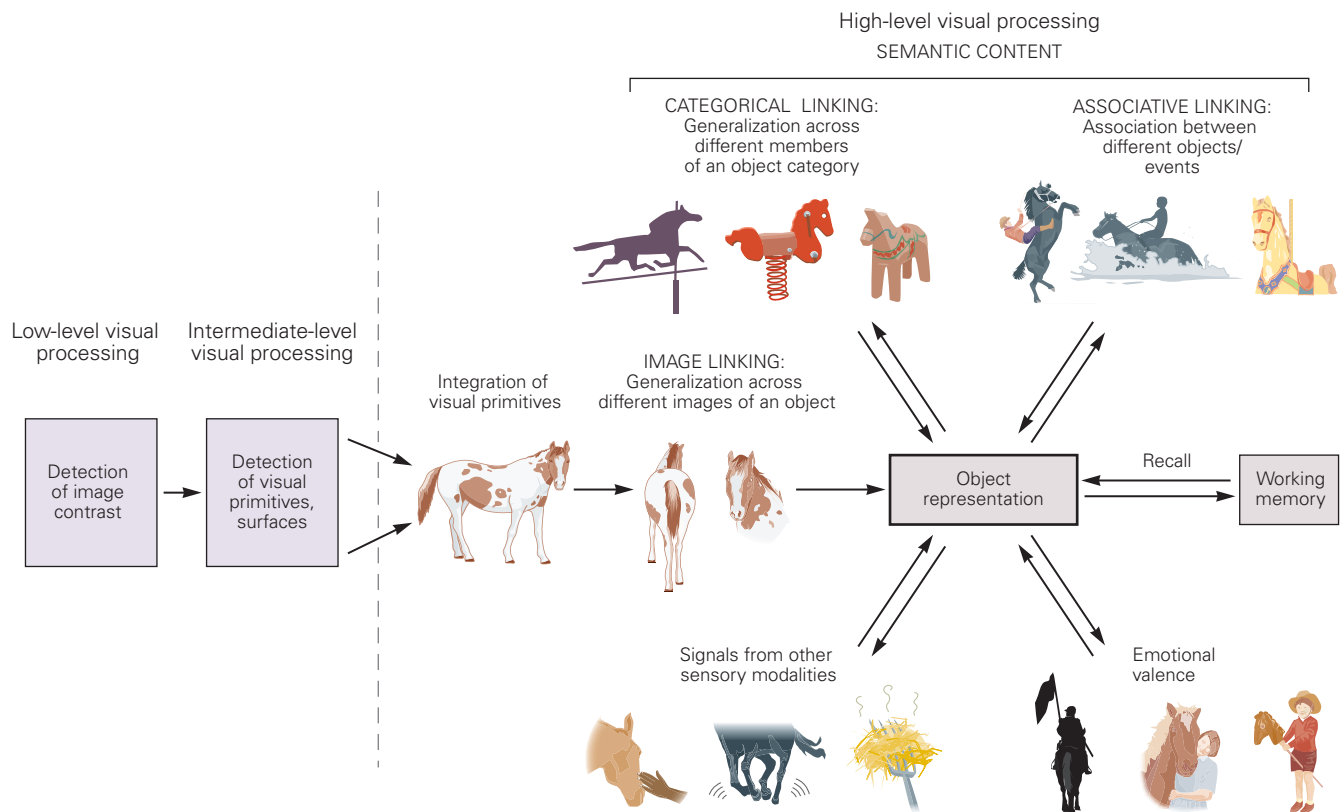


Figure 24–1 Representation of entire objects is central to high-level visual processing. Representation of entire objects involves integration of visual features extracted at earlier stages in the visual pathways. This integration is a generalization of the numerous retinal images generated by the same object and of

different members of an object category. The representation also incorporates information from other sensory modalities, attaches emotional valence, and associates the object with the memory of other objects or events. Object representations can be stored in working memory and recalled in association with other memories.

The Inferior Temporal Cortex Is the Primary Center for Object Recognition

Primate studies implicate neocortical regions of the temporal lobe, principally the inferior temporal cortex, in object perception. Because the hierarchy of synaptic relays in the cortical visual system extends from the primary visual cortex to the temporal lobe, the temporal lobe is a site of convergence of many types of visual information.

Neuropsychological studies have found that damage to the inferior temporal cortex can produce specific failures of object recognition. Neurophysiological and functional imaging studies have, in turn, yielded remarkable insights into the ways in which the activity of inferior temporal neurons represents objects, how these representations relate to perceptual and cognitive events, and how they are modified by experience.

Visual signals originating in the retina are processed in the lateral geniculate nucleus of the thalamus

before reaching the primary visual cortex (V1). Ascending visual pathways from V1 follow two main parallel and hierarchically organized streams: the ventral and dorsal streams (Chapter 21). The ventral stream extends ventrally and anteriorly from V1 through V2, via V4, into inferior temporal cortex, which, in macaque monkeys, comprises the lower bank of the superior temporal sulcus and the ventrolateral convexity of the temporal lobe (Figure 24–2). Neurons at each synaptic relay in this ventral stream receive convergent input from the preceding stage. At the top of the hierarchy, inferior temporal neurons are in a position to integrate a large and diverse quantity of visual information over a vast region of visual space.

The inferior temporal cortex is a large brain region. The patterns of anatomical connections to and from this region indicate that it comprises at least two main functional subdivisions—the posterior area temporo-occipital cortex and the anterior area temporal cortex—and functional evidence suggests further subdivisions

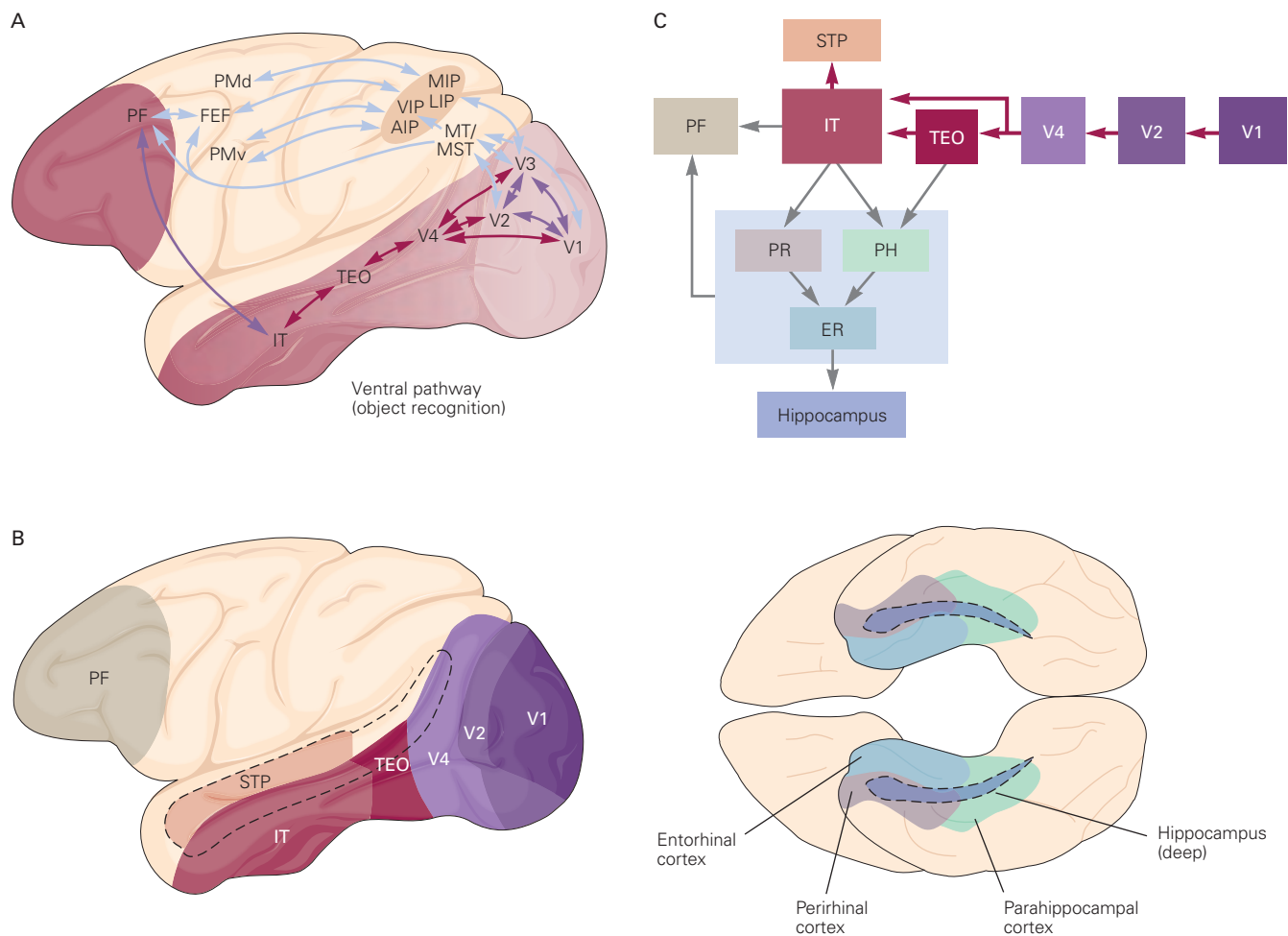


Figure 24-2 Cortical pathway for object recognition.

A. A lateral view of the macaque brain shows the major pathways involved in visual processing, including the pathway for object recognition (red). (Abbreviations: AIP, anterior intraparietal cortex; FEF, frontal eye fields; IT, inferior temporal cortex; LIP, lateral intraparietal cortex; MIP, medial intraparietal cortex; MST, medial superior temporal cortex; MT, middle temporal cortex; PF, prefrontal cortex; PMd, dorsal premotor cortex; PMv, ventral premotor cortex; TEO, temporo-occipital cortex; VIP, ventral intraparietal cortex.)

B. Lateral and ventral views of the macaque monkey brain show the cortical areas involved in object recognition.

(Abbreviations: IT, inferior temporal cortex; PF, prefrontal cortex; STP, superior temporal polysensory area; TEO, temporo-occipital cortex.)

C. The inferior temporal cortex (IT) is the end stage of the ventral stream (red arrows) and is reciprocally connected with neighboring areas of the medial temporal lobe and prefrontal cortex (gray arrows). This chart illustrates the main connections and predominant direction of information flow. (Abbreviations: ER, entorhinal cortex; PF, prefrontal cortex; PH, parahippocampal cortex; PR, perirhinal cortex; STP, superior temporal polysensory area; TEO, temporo-occipital cortex.)

into multiple functionally specialized areas. As we shall see, the distinction between anterior and posterior parts of the inferior temporal cortex is supported by both neuropsychological and neurophysiological evidence.

Clinical Evidence Identifies the Inferior Temporal Cortex as Essential for Object Recognition

The first clear insight into the neural pathways mediating object recognition was obtained in the late 19th century when the American neurologist Sanger Brown

and the British physiologist Edward Albert Schäfer found that experimental lesions of the temporal lobe in primates abolished the ability to recognize objects. Unlike the deficits that accompany lesions of occipital cortical areas, temporal lobe lesions do not impair sensitivity to basic visual attributes, such as color, motion, and distance. Because of the unusual type of visual loss, the impairment was originally called *psychic blindness*, but this term was later replaced by *visual agnosia* (“without visual knowledge”), a term coined by Sigmund Freud.

In humans, there are two basic categories of visual agnosia, apperceptive and associative, the description of which led to a two-stage model of object recognition in the visual system. With apperceptive agnosia, the ability to match or copy complex visual shapes or objects is impaired (Figure 24–3). This impairment results from disruption of the first stage of object recognition: integration of visual features into sensory representations of entire objects. With associative agnosia, the ability to match or copy complex objects remains intact, but the ability to identify objects is impaired. This impairment results from disruption of the second stage of object recognition: association of the sensory

representation of an object with knowledge of the object’s meaning or function.

Consistent with this functional hierarchy, apperceptive agnosia is most common following damage to the posterior inferior temporal cortex, whereas associative agnosia, a higher-order perceptual deficit, is more common following damage to the anterior inferior temporal cortex. Neurons in the anterior subdivision exhibit a variety of memory-related properties not seen in the posterior area.

More focal lesions within temporal cortex can lead to specific deficits. Damage to a small region of the human temporal lobe results in an inability to

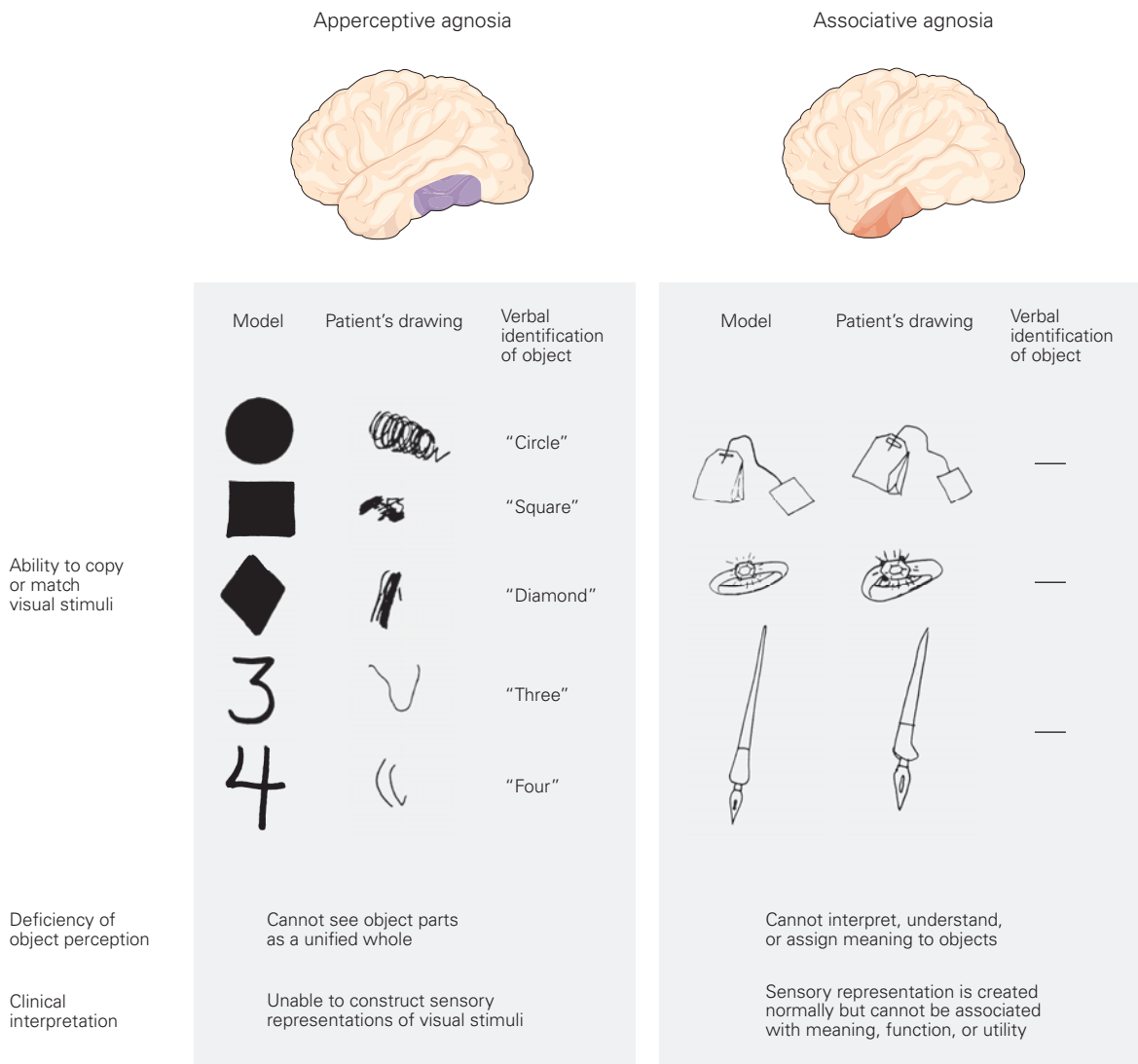


Figure 24–3 Neurons in the temporal lobe of humans are involved in object recognition. Damage to the inferior temporal cortex impairs the ability to recognize visual objects, a condition known as visual agnosia. There are two major categories

of visual agnosia: Apperceptive agnosia results from damage to the posterior region, and associative agnosia results from damage of the anterior region. (Reproduced, with permission, from Farah 1990. © 1990 Massachusetts Institute of Technology.)

recognize faces, a form of associative agnosia known as *prosopagnosia*. Patients with prosopagnosia can identify a face as a face, recognize its parts, and even detect specific emotions expressed by the face, but they are unable to identify a particular face as belonging to a specific person.

Prosopagnosia is an example of a *category-specific agnosia*, in which patients with temporal lobe damage fail to recognize particular items belonging to a specific semantic category. Category-specific agnosias for living things, fruits, vegetables, tools, or animals have also been reported. Owing to the pronounced behavioral significance of faces and the normal ability of people to recognize an extraordinarily large number of faces, prosopagnosia may simply be the most commonly diagnosed variety of category-specific agnosia.

Neurons in the Inferior Temporal Cortex Encode Complex Visual Stimuli and Are Organized in Functionally Specialized Columns

The coding of visual information in the temporal lobe has been studied extensively using electrophysiological

techniques, beginning with the work of Charles Gross and colleagues in the 1970s. Neurons in this region have distinctive response properties. They are relatively insensitive to simple stimulus features such as orientation and color. Instead, the vast majority possess large, centrally located receptive fields and encode complex stimulus features. These selectivities often appear somewhat arbitrary. An individual neuron might, for example, respond strongly to a crescent-shaped pattern of a particular color and texture. Cells with such unique selectivities likely provide inputs to higher-order neurons that respond to specific meaningful objects.

In fact, within the inferior temporal cortex, several small subpopulations of neurons are activated by highly meaningful objects, such as faces and hands (Figure 24–4), as Charles Gross discovered. For cells that respond to the sight of a hand, individual fingers are particularly critical. Among cells that respond to faces, the most effective stimulus for some cells is a frontal view of the face, whereas for others it is a side view. Although some neurons respond preferentially to faces in general, others respond only to specific

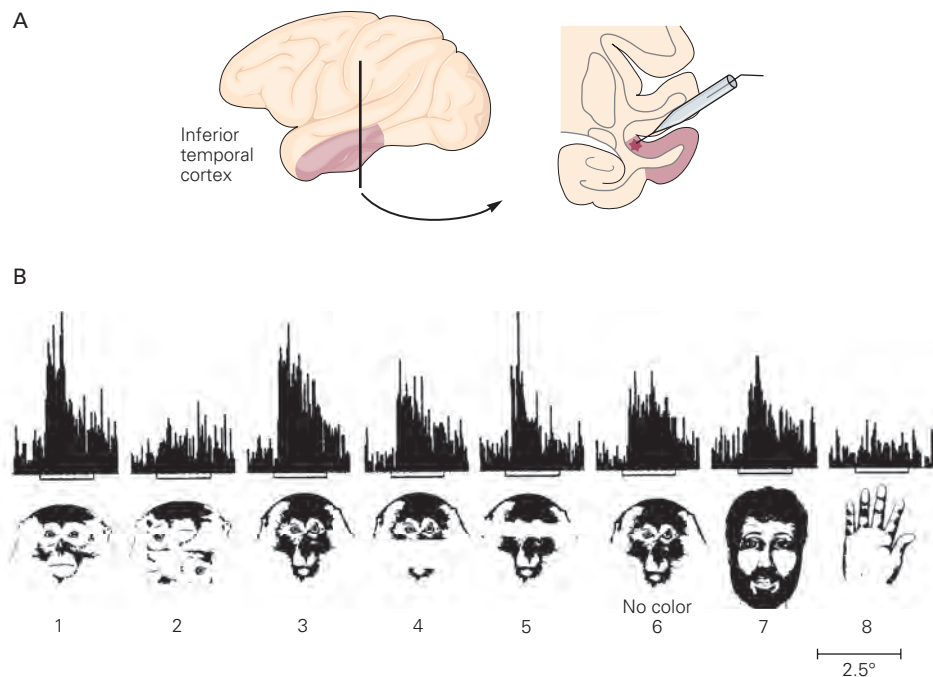


Figure 24–4 Neurons in the inferior temporal cortex of the monkey are involved in face recognition. (Reproduced, with permission, from Desimone et al. 1984. Copyright © 1984 Society for Neuroscience.)

A. The location of the inferior temporal cortex of the monkey is shown in a lateral view and coronal section. The colored area is the location of the recorded neurons.

B. Peristimulus time histograms illustrate the frequency of action potentials in a single neuron in response to different images (shown below the histograms). This neuron responded selectively to faces. Masking of critical features, such as the mouth or eyes (4, 5), led to a substantial but not complete reduction in response. Scrambling the parts of the face (2) nearly eliminated the response.

facial expressions. It seems likely that such cells contribute directly to face recognition.

In initial relays in the cortical visual system, neurons that respond to the same stimulus features, such as orientation or direction of motion, but from different parts of the visual field are organized in columns. Cells within the inferior temporal cortex are similarly organized. Columns of neurons representing the same or similar stimulus properties commonly extend throughout the cortical thickness and over a range of approximately 400 μm . The columns are arranged such that different stimuli that possess some similar features are represented in partially overlapping columns (Figure 24–5). Thus, one stimulus can activate multiple columns. Horizontal connections can span many millimeters and may facilitate the formation of distributed networks for encoding objects.

The Primate Brain Contains Dedicated Systems for Face Processing

Prosopagnosia often occurs in the absence of any other form of agnosia. Such a highly specific perceptual deficit could be explained by focal lesions of face-selective neurons located in exclusive clusters. This idea was strengthened by the discovery of face-selective regions in the human brain by Nancy Kanwisher and colleagues using functional magnetic resonance imaging (fMRI) and by Gregory McCarthy and colleagues using direct electrophysiological recordings from the surface of the human brain. Kanwisher and colleagues found that during the presentation of pictures of faces and other objects one area in the human temporal lobe, the fusiform face area, responded significantly more during the presentation of faces compared to other objects.

Subsequently, several more face-selective areas were found, primarily in temporal but also in prefrontal cortex. Early studies of these areas provided circumstantial evidence for clustering of face-selective neurons. In later studies, Doris Tsao, Winrich Freiwald, and colleagues directly demonstrated such clustering and showed that face processing might be performed by a dedicated face-processing network spanning from the posterior part of inferior temporal cortex to prefrontal cortex. Using fMRI, they found six areas in temporal cortex and three in prefrontal cortex of the macaque monkey that responded more selectively to faces than to other objects. These areas, called face patches, are found at highly consistent locations across individuals and thus are named based on their location. Each face patch is a few millimeters in diameter and thus differs organizationally from the inferior temporal columns. Intracellular recordings from the face patches

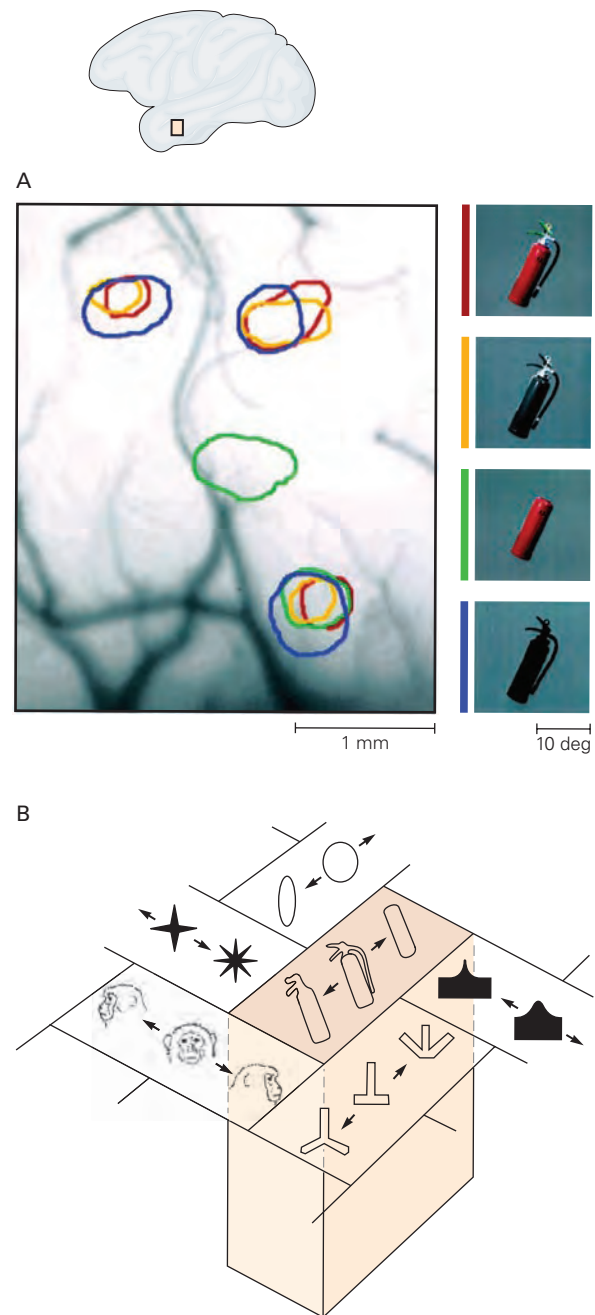


Figure 24–5 Neurons in the anterior portion of the inferior temporal cortex that respond to complex visual stimuli are organized into columns. (Reproduced, with permission, from Tanaka 2003. Copyright © 2003 Oxford University Press.)

A. Optical images of the surface of the anterior inferior temporal cortex illustrate regions selectively activated by the objects shown at the right.

B. Neurons of the inferior temporal cortex are organized in functionally specialized columns that extend from the surface of the cortex. According to this model, each column includes neurons that respond to a specific visually complex object. Columns of neurons that represent variations of an object, such as different faces or different fire extinguishers, constitute a hypercolumn.

revealed that the vast majority of cells respond selectively more to faces than to other objects. Thus, millions of face cells are clustered into a fixed number of small areas. These areas are directly connected to each other, thus forming a face-processing network. Within this network, each node appears to be functionally specialized. From posterior to anterior locations within the temporal lobe, the initial face patches respond to particular views of the face, and then face patches become gradually more selective to identity and less selective for angle of view. Furthermore, dorsal face areas within the temporal lobe exhibit a selectivity for natural facial motion, which ventral areas lack. Thus, a highly specialized network, located primarily in temporal cortex, processes the multiple dimensions of information conveyed by a face (Figure 24–6).

The Inferior Temporal Cortex Is Part of a Network of Cortical Areas Involved in Object Recognition

Object recognition is intimately intertwined with visual categorization, visual memory, and emotion, and

the outputs of the inferior temporal cortex contribute to these functions (see Figure 24–2). Among the principal projections are those to the perirhinal and parahippocampal cortices, which lie medially adjacent to the ventral surface of the inferior temporal cortex (Figure 24–2C). These regions project, in turn, to the entorhinal cortex and the hippocampal formation, both of which are involved in long-term memory storage and retrieval. A second major projection from the inferior temporal cortex is to the prefrontal cortex, an important site for high-level visual processing. As we shall see, prefrontal neurons play important roles in object categorization, visual working memory, and memory recall.

The inferior temporal cortex also provides input—directly and indirectly via the perirhinal cortex—to the amygdala, which is believed to apply emotional valence to sensory stimuli and to engage the cognitive and visceral components of emotion (Chapter 42). Finally, the inferior temporal cortex is a major source of input to multimodal sensory areas of cortex such as the superior temporal polysensory area (Figure 24–2B), which lies dorsally adjacent to the inferior temporal cortex.

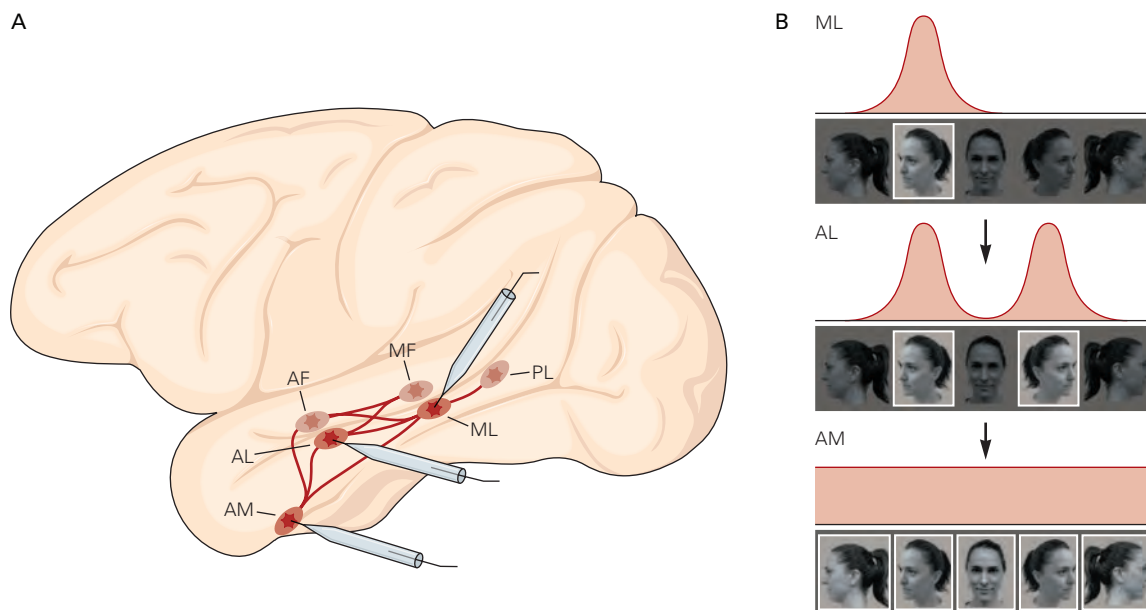


Figure 24–6 The temporal lobe contains a network of face-selective areas.

A. Functional magnetic resonance imaging of macaque monkeys watching pictures of faces and other objects identified six face-selective areas in the temporal lobe, inside and around the superior temporal sulcus. These areas occur at the same locations across subjects and have been given names based on their anatomical location (PL, posterior lateral; ML, medial lateral; MF, medial fundus of the superior temporal sulcus;

AL, anterior lateral; AF, anterior fundus; AM, anterior medial). These areas are interconnected to form a face-processing network.

B. Single-neuron recordings from areas ML, AL, and AM show tuning to head orientation. ML cells are tuned to specific head orientations, many AL cells are tuned to multiple orientations that are mirror-symmetric versions of each other, and AM cells are broadly and more weakly tuned to head orientation. These three representations in interconnected areas can be thought of as transformations of each other (arrows).

Object Recognition Relies on Perceptual Constancy

The ability to recognize objects as the same under different viewing conditions, despite the sometimes markedly different retinal images, is one of the most functionally important requirements of visual experience. The invariant attributes of an object—for example, spatial and chromatic relationships between image features or characteristic features such as the stripes of a zebra—are cues to the identity and meaning of the objects.

For object recognition to take place, these invariant attributes must be represented independently of other image properties. The visual system does this with proficiency, and its behavioral manifestation is termed *perceptual constancy*. Perceptual constancy has many forms ranging from invariance across simple transformations of an object, such as changes of size or position, to more difficult ones, such as rotation in depth or changes in lighting, and even to the sameness of objects within a category: All zebras look alike.

One of the best examples is *size constancy*. An object placed at different distances from an observer is perceived as having the same size, even though the object produces images of different absolute size on the retina. Size constancy has been recognized for centuries, but only in the past several decades has it been possible to identify the neural mechanisms responsible. An early study found that lesions of the inferior temporal cortex lead to failures of size constancy in monkeys, suggesting that neurons in this area play a critical role in size constancy. Indeed, one of the most striking properties of individual inferior temporal neurons is the invariance of their shape selectivity even to very big changes in stimulus size (Figure 24-7A).

Another type of perceptual constancy is *position constancy*, in which objects are recognized as the same regardless of their location in the visual field. The pattern of selective response of many inferior temporal neurons does not vary when an object changes position within their large receptive fields (Figure 24-7B). *Form-cue invariance* refers to the constancy of a form when the cues that define the form change. The silhouette of Abraham Lincoln's head, for example, is readily recognizable whether it is black on white, white on black, or red on green. The responses of many inferior temporal neurons do not change with changes in contrast polarity (Figure 24-7C), color, or texture.

Viewpoint invariance refers to the perceptual constancy of three-dimensional objects observed from different angles. Because most objects we see are

three-dimensional and opaque, when looked at from different viewpoints, some parts become invisible, while others are revealed, and all others change in appearance. Yet despite the limitless range of retinal images that might be cast by a familiar object, an observer can readily recognize an object independently of the angle at which it is viewed. There are notable exceptions to this rule, which generally occur when an object is viewed from an angle that yields an uncharacteristic retinal image, such as a bucket viewed from directly above.

Thus, object recognition mechanisms must infer the identity of objects from apparent complex shapes. Many neurons in inferior temporal cortex do not exhibit viewpoint invariance. In fact, many are systematically tuned to viewing angle. Yet at more anterior locations, neurons are not only more size and position invariant, but they also exhibit greater invariance to viewpoint. The face-processing system is a case in point. Neurons in posterior face patches are tuned to viewing angle, while neurons in anterior face patches exhibit great robustness to changes in viewpoint. Thus, population responses in posterior face areas contain more information about head orientation than those in anterior areas, while the anterior face patches provide more information about face identity across head orientations compared to posterior face areas. The degree of viewpoint invariance achieved in anterior inferior temporal cortex, by individual neurons and populations of neurons, might be sufficient to account for perceptual viewpoint invariance. But this has not been directly shown yet. Alternatively, viewpoint invariance may be achieved at a higher stage of cortical processing, such as the prefrontal cortex.

Studies of the conditions under which viewpoint invariance fails may lead to insights into the neural mechanisms of the behavior. One such condition is presentation of mirror images. Although mirror images are not identical, they are frequently perceived as such, a confusion reflecting a false-positive identification by the system for viewpoint invariance. Carl Olson and colleagues examined the responses of neurons in a particular region of the inferior temporal cortex to stimuli that were mirror images. Consistent with the perceptual confusion, many inferior temporal neurons responded similarly to both images. Similarly, in one face area between the posterior and anterior ones described earlier, profile-selective cells respond similarly to the left and right profile of a face. These results reinforce the conclusion that activity in the inferior temporal cortex reflects perceptual invariance, albeit incorrectly in this case, rather than the actual features of a stimulus.

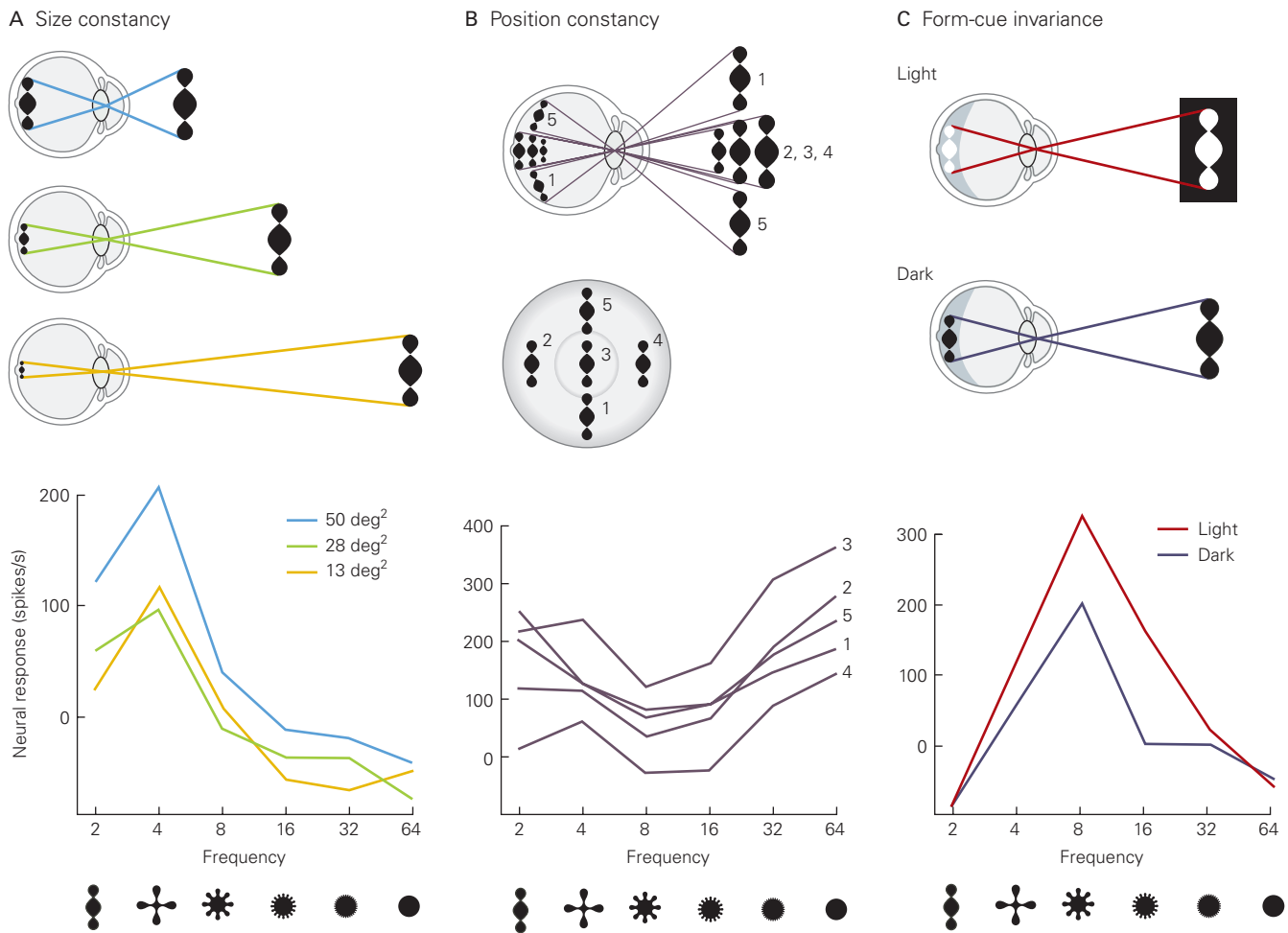


Figure 24-7 Perceptual constancy is reflected in the behavior of neurons in the inferior temporal cortex. The responses of many inferior temporal neurons are selective for stimuli with a particular frequency (number) of lobes but invariant to object size, position, and reflectance. (Reproduced, with permission, from Schwartz et al. 1983.)

A. Size constancy. An object is perceived to be the same even when the retinal image size decreases with the distance of the object in the visual field. The response of the vast majority of inferior temporal neurons to substantial changes in retinal image size is invariant, as illustrated here by the record of a single cell.

B. Position constancy. An object is perceived to be the same despite changes in position in the retinal image. Almost all inferior temporal neurons respond similarly to the same stimulus in different positions in the visual field, as illustrated here by the record of a single neuron.

C. Form-cue invariance. An object is perceived to be the same despite changes in reflectance. Most inferior temporal neurons respond similarly to the two images illustrated, as shown in the record of an individual neuron.

Categorical Perception of Objects Simplifies Behavior

All forms of perceptual constancy are the product of the visual system's attempts to generalize across different retinal images generated by a single object. A still more general type of constancy is the perception of individual objects as belonging to the same semantic category. The apples in a basket or the many appearances of the letter *A* in different fonts, for example, are

physically distinct but are effortlessly perceived as *categorically identical*.

Categorical perception is classically defined as the ability to distinguish objects of different categories better than objects of the same category. For example, it is more difficult to discriminate between two red lights that differ in wavelength by 10 nm than to discriminate between red and orange lights with the same wavelength difference.

Categorical perception simplifies behavior. For example, it usually does not matter whether an apple