

Visual Perception: How Your Brain Interprets Reality

Authored by
mohammad looti

June 16, 2026

RECOMMENDED CITATION

mohammad looti (2026). *Visual Perception: How Your Brain Interprets Reality*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=38113>

Visual perception is the ability to interpret the surrounding environment using light in the visible spectrum reflected by the objects in the environment.

The resulting perception is also known as visual perception, eyesight, sight, or vision (adjectival form: visual, optical, or ocular). The various physiological components involved in vision are referred to collectively as the visual system, and are the focus of much research in linguistics, psychology, cognitive science, neuroscience, and molecular biology, collectively referred to as vision science.

Visual System

The visual system in animals allows individuals to assimilate information from their surroundings. The act of seeing starts when the cornea and then the lens of the eye focuses light from its surroundings onto a light-sensitive membrane in the back of the eye, called the retina. The retina is actually part of the brain that is isolated to serve as a transducer for the conversion of light into neuronal signals. Based on feedback from the visual system, the lens of the eye adjusts its thickness to focus light on the photoreceptive cells of the retina, also known as the rods and cones, which detect the photons of light and respond by producing neural impulses. These signals are processed via complex feedforward and feedback processes by different parts of the brain, from the retina upstream to central ganglia in the brain.

Note that up until now much of the above paragraph could apply to octopuses, mollusks, worms, insects and things more primitive; anything with a more concentrated nervous system and better eyes than say a jellyfish. However, the following applies to mammals generally and birds (in modified form): The retina in these more complex animals sends fibers (the optic nerve) to the lateral geniculate nucleus, to the primary and secondary visual cortex of the brain. Signals from the retina can also travel directly from the retina to the superior colliculus.

The perception of objects and the totality of the visual scene is accomplished by the visual association cortex. The visual association cortex combines all sensory information perceived by the striate cortex which contains thousands of modules that are part of modular neural networks. The neurons in the striate cortex send axons to the extrastriate cortex, a region in the visual association cortex that surrounds the striate cortex.

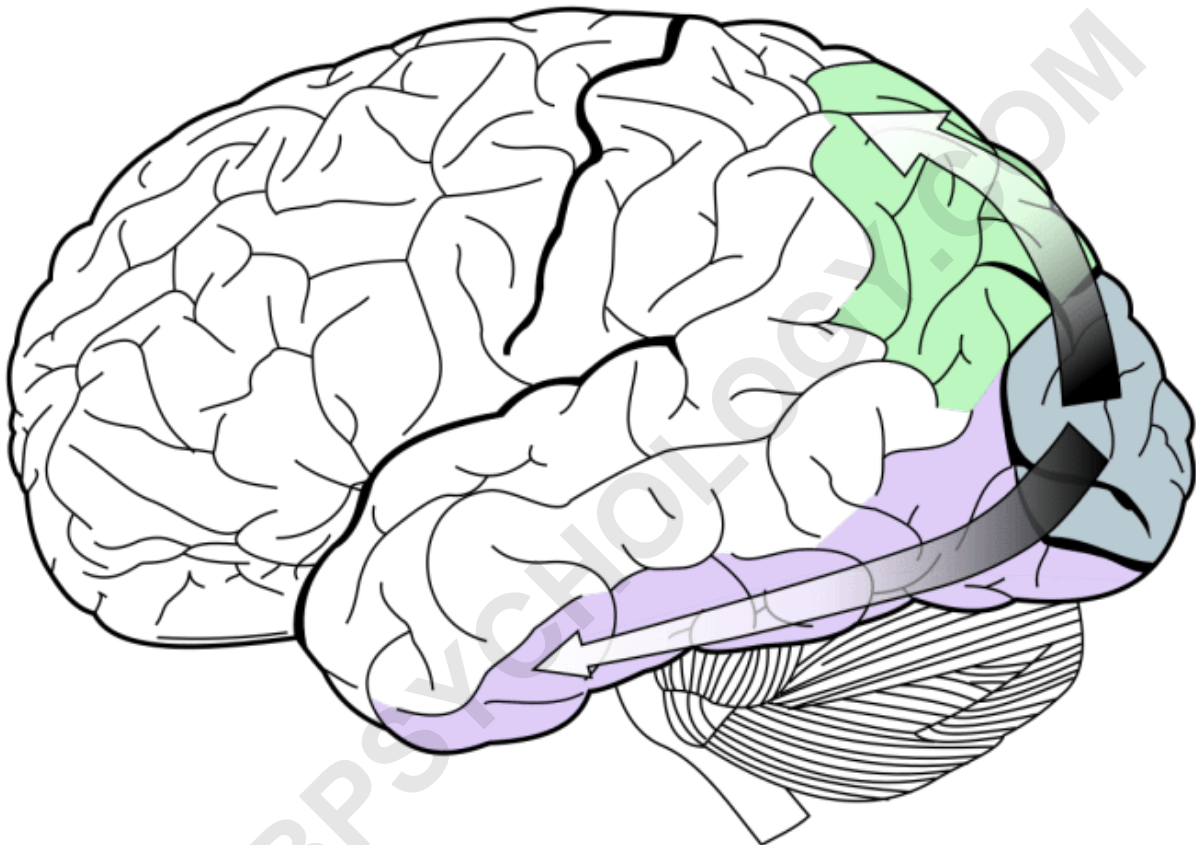
The human visual system perceives visible light in the range of wavelengths between 370 and 730 nanometers (0.00000037 to 0.00000073 meters) of the electromagnetic spectrum.

Study

The major problem in visual perception is that what people see is not simply a translation of retinal

stimuli (i.e., the image on the retina). Thus people interested in perception have long struggled to explain what visual processing does to create what is actually seen.

Early Studies



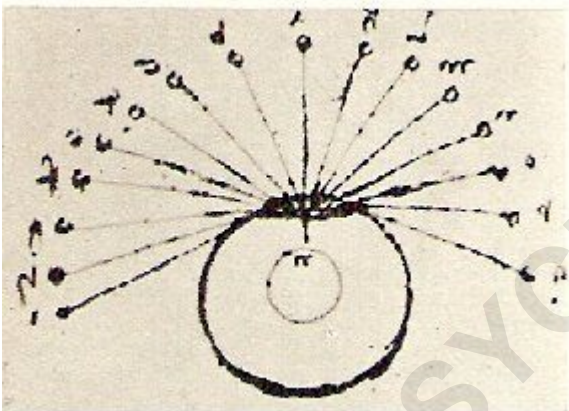
The visual dorsal stream (green) and ventral stream (purple) are shown. Much of the human cerebral cortex is involved in vision.

There were two major ancient Greek schools, providing a primitive explanation of how vision is carried out in the body.

The first was the "emission theory" which maintained that vision occurs when rays emanate from the eyes and are intercepted by visual objects. If an object was seen directly it was by 'means of rays' coming out of the eyes and again falling on the object. A refracted image was, however, seen by 'means of rays' as well, which came out of the eyes, traversed through the air, and after refraction, fell on the visible object which was sighted as the result of the movement of the rays from the eye. This theory was championed by scholars like Euclid and Ptolemy and their followers.

The second school advocated the so-called 'intro-mission' approach which sees vision as coming from something entering the eyes representative of the object. With its main propagators Aristotle, Galen and their followers, this theory seems to have some contact with modern theories of what vision really is, but it remained only a speculation lacking any experimental foundation. (In eighteenth-century England, Isaac Newton, John Locke, and others, carried the intromission/intromittist theory forward by insisting that vision involved a process in which rays--composed of actual corporeal matter--emanated from seen objects and entered the seer's mind/sensorium through the eye's aperture.)

Both schools of thought relied upon the principle that "like is only known by like", and thus upon the notion that the eye was composed of some "internal fire" which interacted with the "external fire" of visible light and made vision possible. Plato makes this assertion in his dialogue *Timaeus*, as does Aristotle, in his *De Sensu*.



Leonardo da Vinci: The eye has a central line and everything that reaches the eye through this central line can be seen distinctly.

Alhazen (965 - c. 1040) carried out many investigations and experiments on visual perception, extended the work of Ptolemy on binocular vision, and commented on the anatomical works of Galen. He was the first person to explain that vision occurs when light bounces on an object and then is directed to one's eyes.

Leonardo da Vinci (1452-1519) is believed to be the first to recognize the special optical qualities of the eye. He wrote "The function of the human eye ... was described by a large number of authors in a certain way. But I found it to be completely different." His main experimental finding was that there is only a distinct and clear vision at the line of sight--the optical line that ends at the fovea. Although he did not use these words literally he actually is the father of the modern distinction between foveal and peripheral vision.

Issac Newton (1642-1726/27) was the first to discover through experimentation, by isolating individual colors of the spectrum of light passing through a prism, that the visually perceived color of objects appeared due to the character of light the objects reflected, and that these divided colors could not be changed into any other color, which was contrary to scientific expectation of the day.

Unconscious Inference

Hermann von Helmholtz is often credited with the first study of visual perception in modern times. Helmholtz examined the human eye and concluded that it was, optically, rather poor. The poor-quality information gathered via the eye seemed to him to make vision impossible. He therefore concluded that vision could only be the result of some form of unconscious inferences: a matter of making assumptions and conclusions from incomplete data, based on previous experiences.

Inference requires prior experience of the world.

Examples of well-known assumptions, based on visual experience, are:

light comes from above

objects are normally not viewed from below

faces are seen (and recognized) upright.

closer objects can block the view of more distant objects, but not vice versa

figures (i.e., foreground objects) tend to have convex borders

The study of visual illusions (cases when the inference process goes wrong) has yielded much insight into what sort of assumptions the visual system makes.

Another type of the unconscious inference hypothesis (based on probabilities) has recently been revived in so-called Bayesian studies of visual perception. Proponents of this approach consider that the visual system performs some form of Bayesian inference to derive a perception from sensory data. However, it is not clear how proponents of this view derive, in principle, the relevant probabilities required by the Bayesian equation. Models based on this idea have been used to describe various visual perceptual functions, such as the perception of motion, the perception of depth, and figure-ground perception. The "wholly empirical theory of perception" is a related and newer approach that rationalizes visual perception without explicitly invoking Bayesian formalisms.

Gestalt Theory

Gestalt psychologists working primarily in the 1930s and 1940s raised many of the research questions that are studied by vision scientists today.

The Gestalt Laws of Organization have guided the study of how people perceive visual

components as organized patterns or wholes, instead of many different parts. "Gestalt" is a German word that partially translates to "configuration or pattern" along with "whole or emergent structure". According to this theory, there are eight main factors that determine how the visual system automatically groups elements into patterns: Proximity, Similarity, Closure, Symmetry, Common Fate (i.e. common motion), Continuity as well as Good Gestalt (pattern that is regular, simple, and orderly) and Past Experience.

Analysis of Eye Movement

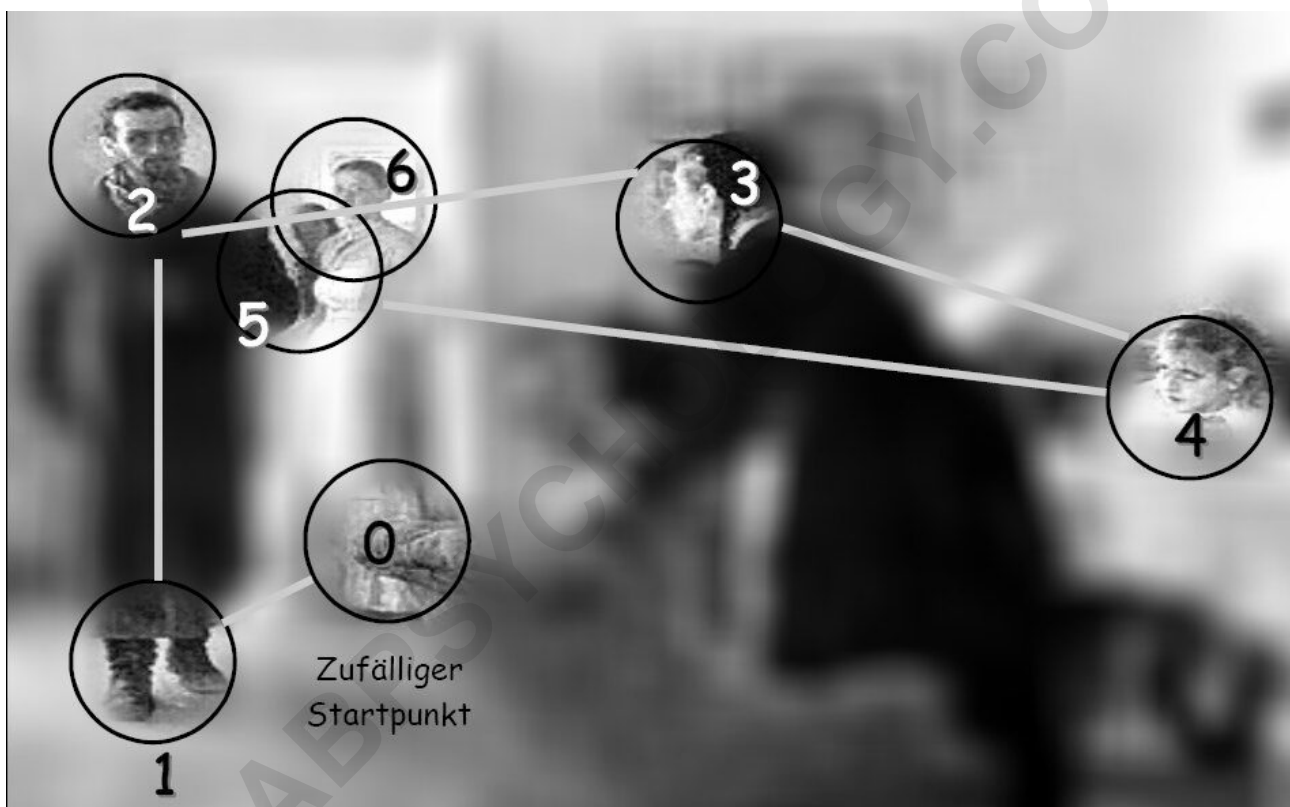


Bild 11: Foveale Erganzung durch die ersten 6 Fixationen (nach Daten von *Yarbus, 1967*)

Eye movement first 2 seconds (Yarbus, 1967)

During the 1960s, technical development permitted the continuous registration of eye movement during reading in picture viewing and later in visual problem solving and when headset-cameras became available, also during driving.

The picture to the right shows what may happen during the first two seconds of visual inspection. While the background is out of focus, representing the peripheral vision, the first eye movement goes to the boots of the man (just because they are very near the starting fixation and have a

reasonable contrast).

The following fixations jump from face to face. They might even permit comparisons between faces.

It may be concluded that the icon face is a very attractive search icon within the peripheral field of vision. The foveal vision adds detailed information to the peripheral first impression.

It can also be noted that there are three different types of eye movements: vergence movements, saccadic movements and pursuit movements. Vergence movements involve the cooperation of both eyes to allow for an image to fall on the same area of both retinas. This results in a single focused image. Saccadic movements is the type of eye movement that makes jumps from one position to another position and is used to rapidly scan a particular scene/image. Lastly, pursuit movement is smooth eye movement and is used to follow objects in motion.

Face and Object Recognition

There is considerable evidence that face and object recognition are accomplished by distinct systems. For example, prosopagnosic patients show deficits in face, but not object processing, while object agnostic patients (most notably, patient C.K.) show deficits in object processing with spared face processing. Behaviorally, it has been shown that faces, but not objects, are subject to inversion effects, leading to the claim that faces are "special". Further, face and object processing recruit distinct neural systems. Notably, some have argued that the apparent specialization of the human brain for face processing does not reflect true domain specificity, but rather a more general process of expert-level discrimination within a given class of stimulus, though this latter claim is the subject of substantial debate. Using fMRI and electrophysiology Doris Tsao and colleagues described brain regions and a mechanism for face recognition in macaque monkeys.

The Cognitive and Computational Approaches

In the 1970s, David Marr developed a multi-level theory of vision, which analyzed the process of vision at different levels of abstraction. In order to focus on the understanding of specific problems in vision, he identified three levels of analysis: the computational, algorithmic and implementational levels. Many vision scientists, including Tomaso Poggio, have embraced these levels of analysis and employed them to further characterize vision from a computational perspective.

The computational level addresses, at a high level of abstraction, the problems that the visual system must overcome. The algorithmic level attempts to identify the strategy that may be used to solve these problems. Finally, the implementational level attempts to explain how solutions to these problems are realized in neural circuitry.

Marr suggested that it is possible to investigate vision at any of these levels independently. Marr described vision as proceeding from a two-dimensional visual array (on the retina) to a three-dimensional description of the world as output. His stages of vision include:

A 2D or primal sketch of the scene, based on feature extraction of fundamental components of the scene, including edges, regions, etc. Note the similarity in concept to a pencil sketch drawn quickly by an artist as an impression.

A 2½ D sketch of the scene, where textures are acknowledged, etc. Note the similarity in concept to the stage in drawing where an artist highlights or shades areas of a scene, to provide depth.

A 3 D model, where the scene is visualized in a continuous, 3-dimensional map.

Marr's 2.5D sketch assumes that a depth map is constructed, and that this map is the basis of 3D shape perception. However, both stereoscopic and pictorial perception, as well as monocular viewing, make clear that the perception of 3D shape precedes, and does not rely on, the perception of the depth of points. It is not clear how a preliminary depth map could, in principle, be constructed, nor how this would address the question of figure-ground organization, or grouping. The role of perceptual organizing constraints, overlooked by Marr, in the production of 3D shape percepts from binocularly-viewed 3D objects has been demonstrated empirically for the case of 3D wire objects, e.g. For a more detailed discussion, see Pizlo (2008).

Transduction

Transduction is the process through which energy from environmental stimuli is converted to neural activity for the brain to understand and process. The back of the eye contains three different cell layers: photoreceptor layer, bipolar cell layer and ganglion cell layer. The photoreceptor layer is at the very back and contains rod photoreceptors and cone photoreceptors. Cones are responsible for color perception. There are three different cones: red, green and blue. Rods, are responsible for the perception of objects in low light. Photoreceptors contain within them a special chemical called a photopigment, which are embedded in the membrane of the lamellae; a single human rod contains approximately 10 million of them. The photopigment molecules consist of two parts: an opsin (a protein) and retinal (a lipid). There are 3 specific photopigments (each with their own color) that respond to specific wavelengths of light. When the appropriate wavelength of light hits the photoreceptor, its photopigment splits into two, which sends a message to the bipolar cell layer, which in turn sends a message to the ganglion cells, which then send the information through the optic nerve to the brain. If the appropriate photopigment is not in the proper photoreceptor (for example, a green photopigment inside a red cone), a condition called color vision deficiency will occur.

Opponent Process

Transduction involves chemical messages sent from the photoreceptors to the bipolar cells to the ganglion cells. Several photoreceptors may send their information to one ganglion cell. There are two types of ganglion cells: red/green and yellow/blue. These neuron cells constantly fire--even when not stimulated. The brain interprets different colors (and with a lot of information, an image) when the rate of firing of these neurons alters. Red light stimulates the red cone, which in turn stimulates the red/green ganglion cell. Likewise, green light stimulates the green cone, which stimulates the red/green ganglion cell and blue light stimulates the blue cone which stimulates the yellow/blue ganglion cell. The rate of firing of the ganglion cells is increased when it is signaled by one cone and decreased (inhibited) when it is signaled by the other cone. The first color in the name of the ganglion cell is the color that excites it and the second is the color that inhibits it. i.e.: A red cone would excite the red/green ganglion cell and the green cone would inhibit the red/green ganglion cell. This is an opponent process. If the rate of firing of a red/green ganglion cell is increased, the brain would know that the light was red, if the rate was decreased, the brain would know that the color of the light was green.

Artificial Visual Perception

Theories and observations of visual perception have been the main source of inspiration for computer vision (also called machine vision, or computational vision). Special hardware structures and software algorithms provide machines with the capability to interpret the images coming from a camera or a sensor. Artificial Visual Perception has long been used in the industry and is now entering the domains of automotive and robotics.