

Neural mechanisms of vision

Giuseppe di Pellegrino
Department of Psychology, University of Bologna

g.dipellegrino@unibo.it

Cognition and Neuroscience
Second cycle Degree in Artificial Intelligence – 2021/22

Outline

The anatomical and functional organization of vision

Low-level visual processing: the retina

What is vision?

What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the process of discovering what is present in the world, and where it is. (D. Marr, Vision, 1982)

Vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information. (Marr and Nishihara, 1978).

Vision dominates our perceptions and memories of the world and appears even to frame the way we think.

Vision is used not only for object recognition but also for guiding our movements.

These separate functions are mediated by at least two parallel and interacting pathways.

Vision, and more generally the brain, is a system that analyzes information (information processing device): receives inputs and transforms them into outputs.

Marr's "tri-level hypothesis"

information-processing systems can be analysed in terms of the problems that they solve, the representations and processes by which they solve them, and their physical instantiation.

1.2 Understanding Complex Information-Processing Systems

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1–4. The three levels at which any machine carrying out an information-processing task must be understood.

Marr, Vision (1982)

A central element in Marr's view was that a higher - level was largely independent of the levels below it, and hence computational problems of the highest level could be analyzed independently of understanding the algorithm that executes the computation.

For the same reason the algorithmic problem of the second level was thought to be solvable independently of an understanding of the physical implementation.

In contrast to the doctrine of independence of computation from implementation, current research suggests that considerations of implementation play a vital role in the kinds of algorithms that are devised and the kind of computational insights available to the scientist.

It would be convenient if we could understand the nature of cognition without understanding the nature of the brain itself.

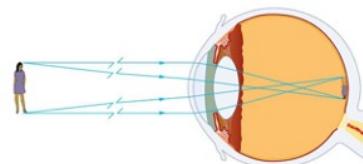
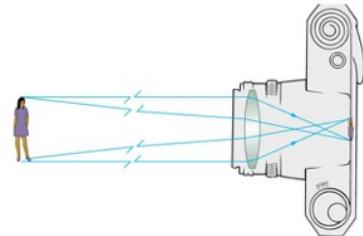
Unfortunately, it is difficult if not impossible to theorize effectively on these matters in the absence of neurobiological constraints.

The primary reason is that computational space is consummately vast, and there are many conceivable solutions to the problem of how a cognitive operation could be accomplished. Neurobiological data provide essential constraints on computational theories, and they consequently provide an efficient means for narrowing the search space (Churchland e Sejnowski, Science, 1988).

Vision is often incorrectly compared to the operation of a camera.

A camera simply reproduces point-by-point the light intensities in one plane of the visual field.

The visual system, in contrast, does something fundamentally different. It interprets the scene and parses it into distinct components, separating foreground from background.



5

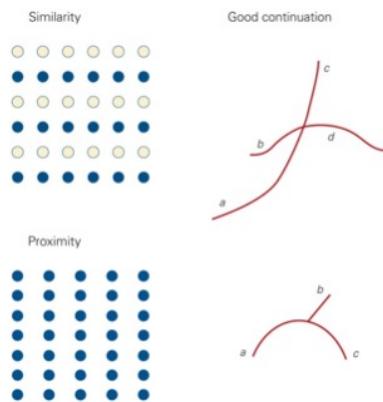
The visual system is less accurate than a camera at certain tasks, such as quantifying the absolute level of brightness or identifying spectral colors.

However, it excels at tasks such as recognizing objects (a charging animal or a speeding car) whether in bright sunlight or at dusk, in an open field or partly occluded by trees (or other cars).

And it does so rapidly (<200ms) to let the viewer respond and, if necessary, escape.

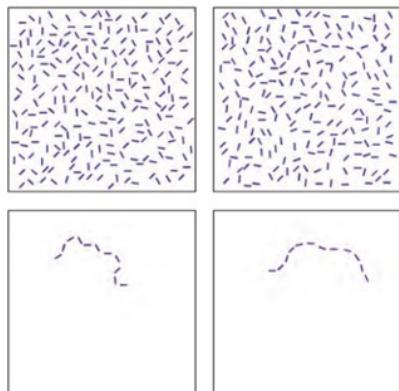
Vision is an active and bidirectional process

Vision is a generative process that involves more than just the information provided to the retina. The brain has a way of looking at the world, a set of expectations about the structure of the world that derives in part from experience and in part from built-in neuronal wiring neural wiring.



To link the elements of a visual scene into unified percepts, the visual system relies on organizational rules such as similarity, proximity, and good continuation.

Contour saliency



The principle of good continuation is also seen in contour saliency. On the right, a smooth contour of line elements pops out from the background, whereas the jagged contour on the left is lost in the background.



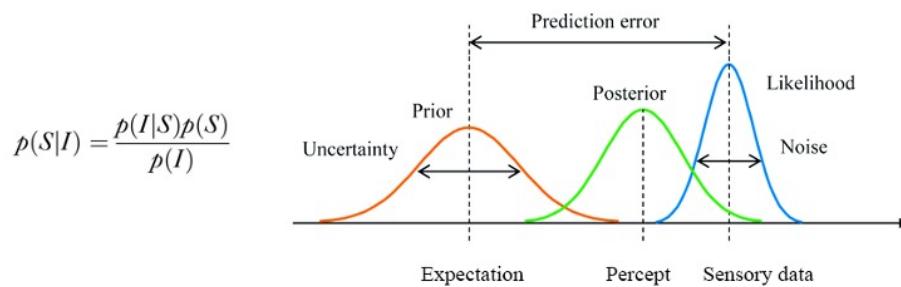
Visual priming

Higher-order representations of shape (in memory) guide lower-order processes of surface segmentation.

Bayesian theories treat the visual system as an ideal observer that uses prior knowledge about visual scenes and information in the image to infer the most probable interpretation of the image.

The posterior probability of a possible real-world stimulus S (i.e., percept) is proportional to the product of the prior probability of S (that is, the probability of S before receiving the stimulus I , e.g., expectation) and the likelihood (the probability of I given S , i.e., sensory data).

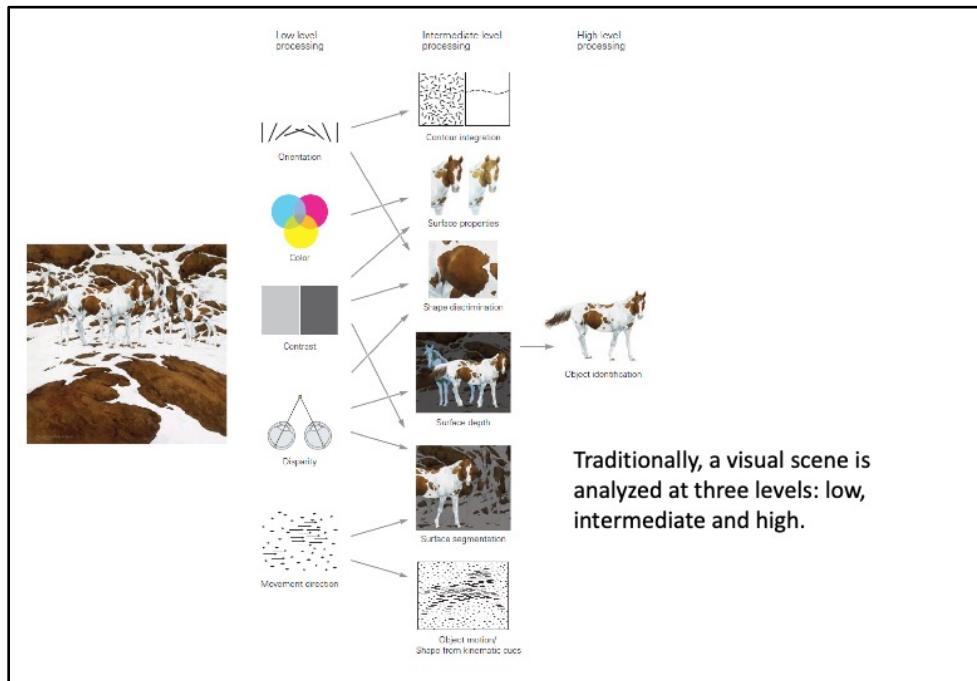
Is it reasonable to assume that the visual system knows the probability calculus and operates according to it?



Prior probability distributions in typical applications of the Bayesian strategy represent knowledge of the regularities governing object shapes, constituent materials, and illumination, and likelihood distributions represent knowledge of how images are formed through projection on the retina. Some examples of prior knowledge are that solids are more likely to be convex than concave and that the light source is above the viewer.

The more ambiguous the image – the greater the influence of prior knowledge in yielding a nonambiguous percept.

Some perceptions may be more data-driven, others more prior knowledge driven.

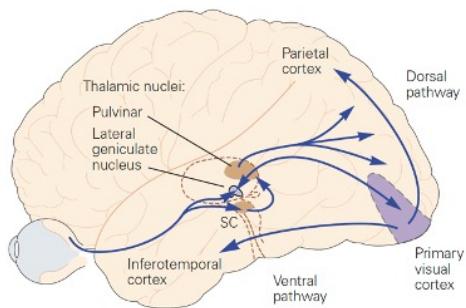


Low-level processing: simple attributes (contrast, orientation, motion, depth, and color) of the visual environment are first analyzed;

Intermediate-level processing: low-level features are used to parse the visual scene. Local orientation is integrated into global contours (contour integration); local visual features are assembled into surfaces, objects are segregated from background (surface segmentation), surface shape is identified from depth, shading and kinematic cues.

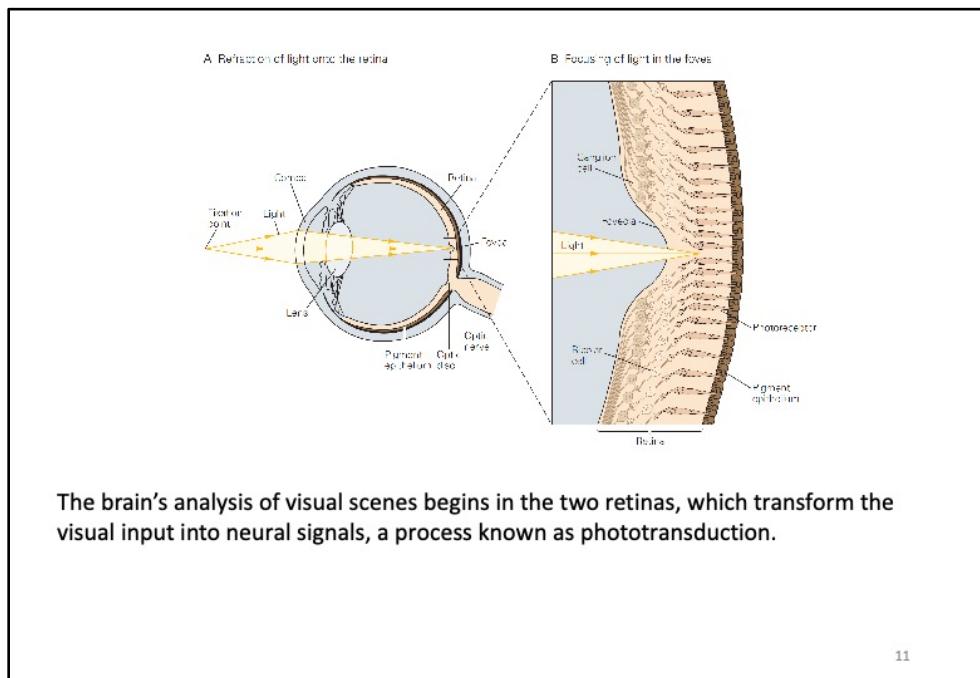
High-level processing: surfaces and contours are used to identify the object

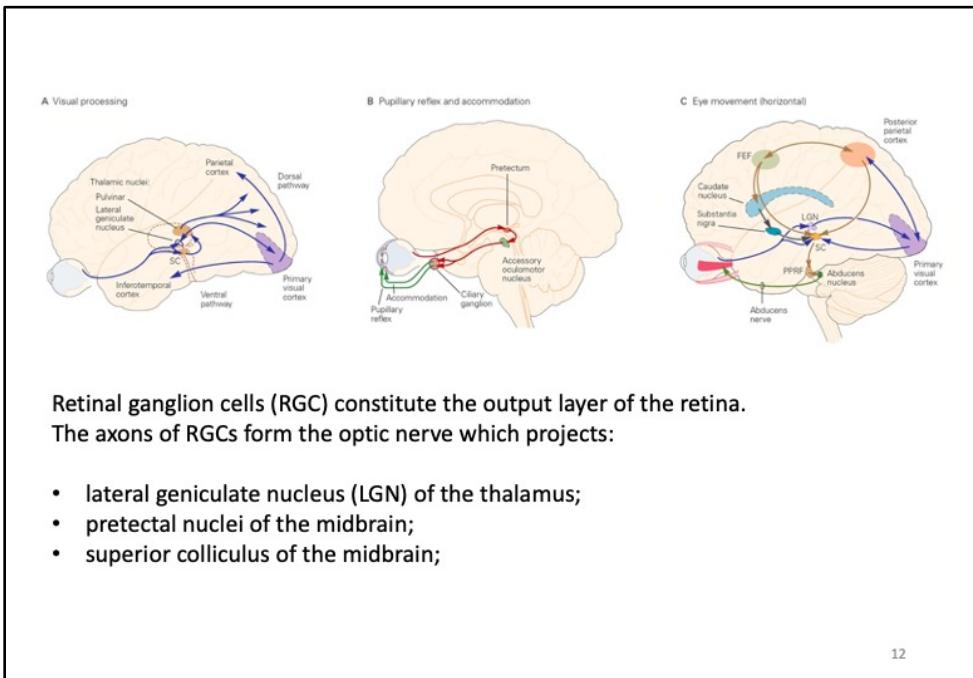
Visual processing is mediated by the retino-geniculo-striate pathway



This pathway includes:

- the retina;
- The lateral geniculate nucleus (LGN) of the thalamus;
- the primary visual cortex (V1) or striate cortex;





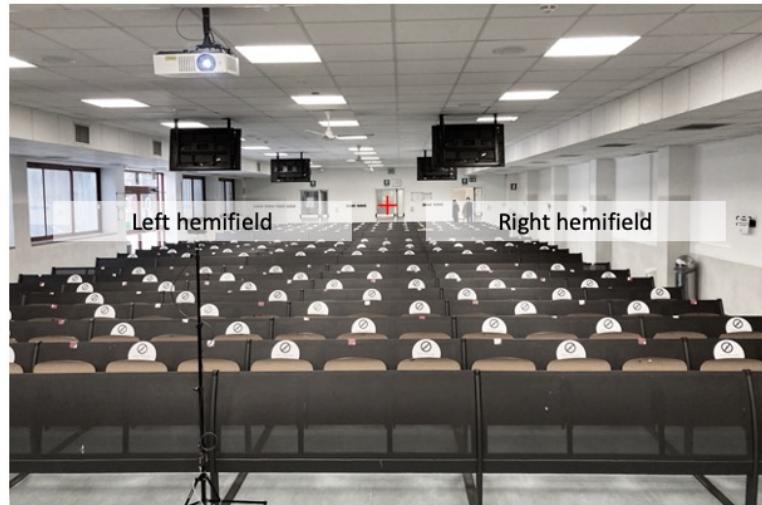
Retinal ganglion cells (RGC) constitute the output layer of the retina.

The axons of RGCs form the optic nerve which projects:

- lateral geniculate nucleus (LGN) of the thalamus;
- pretectal nuclei of the midbrain;
- superior colliculus of the midbrain;

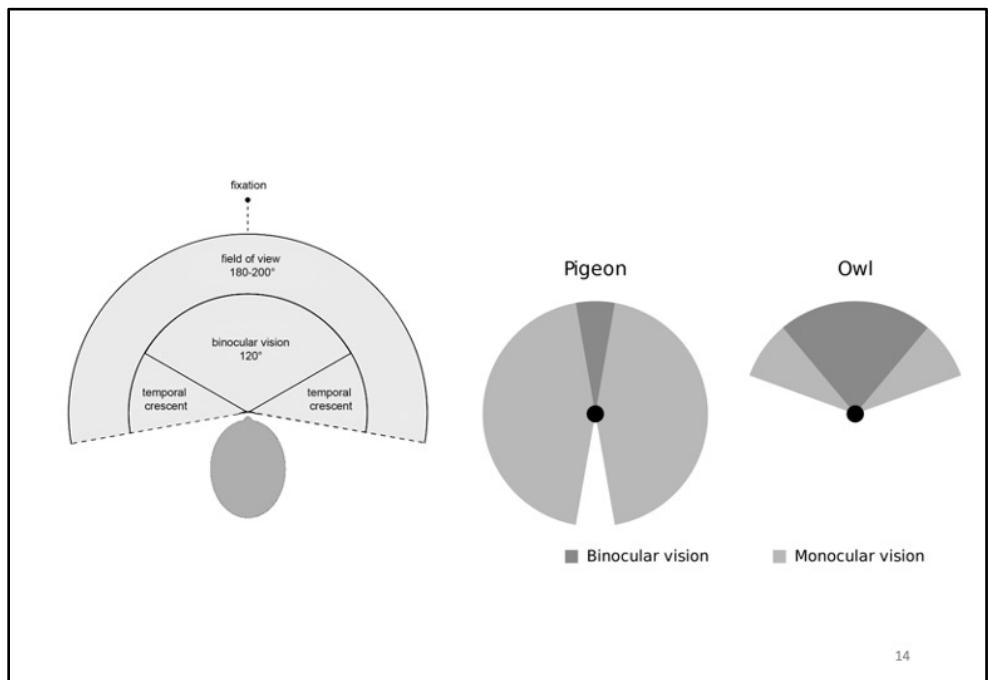
12

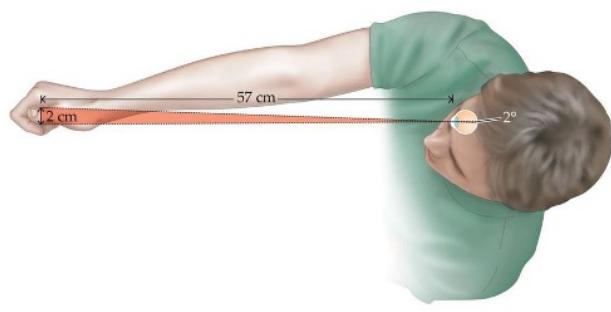
The primary visual pathway is also called the geniculostriate pathway because it passes through the LGN on its way to the primary visual cortex (V1), also known as the striate cortex because of the myelin-rich stripe that runs through its middle layers; A second pathway extends from the retina to the pretectal area of the midbrain, where neurons mediate the pupillary reflexes that control the amount of light entering the eyes; A third pathway from the retina runs to the superior colliculus and is important in controlling eye movements;



Visual field refers to the area that can be seen as you fixate your eyes on a central point (cross).

13



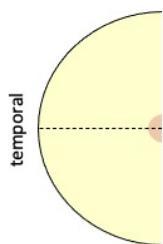


Visual angle of the width of the thumb held at arm's length is about 2 deg.

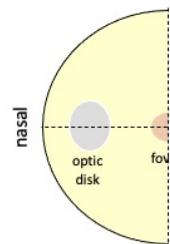
The retina of each eye can be divided into

Nasal hemiretina, medial to the fovea

Temporal hemiretina, lateral to the fovea



Left eye



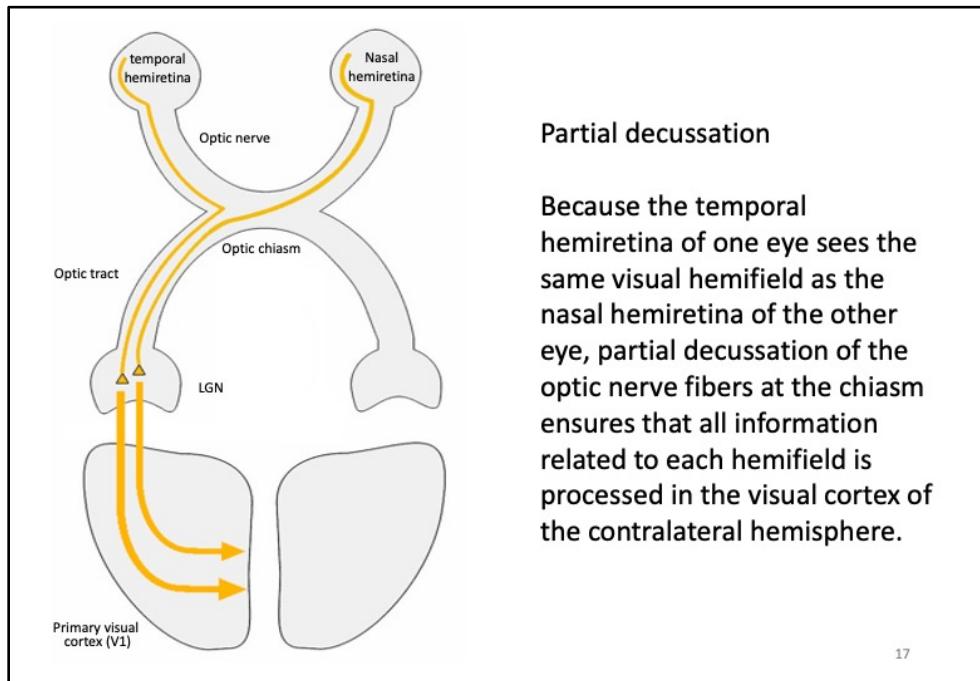
Right eye

The left hemiretinas (temporal of the left eye, and nasal of the right eye) see the right (opposite) visual hemifield

The right hemiretinas (nasal of the left eye, and temporal of the right eye) see the left (opposite) visual hemifield

16

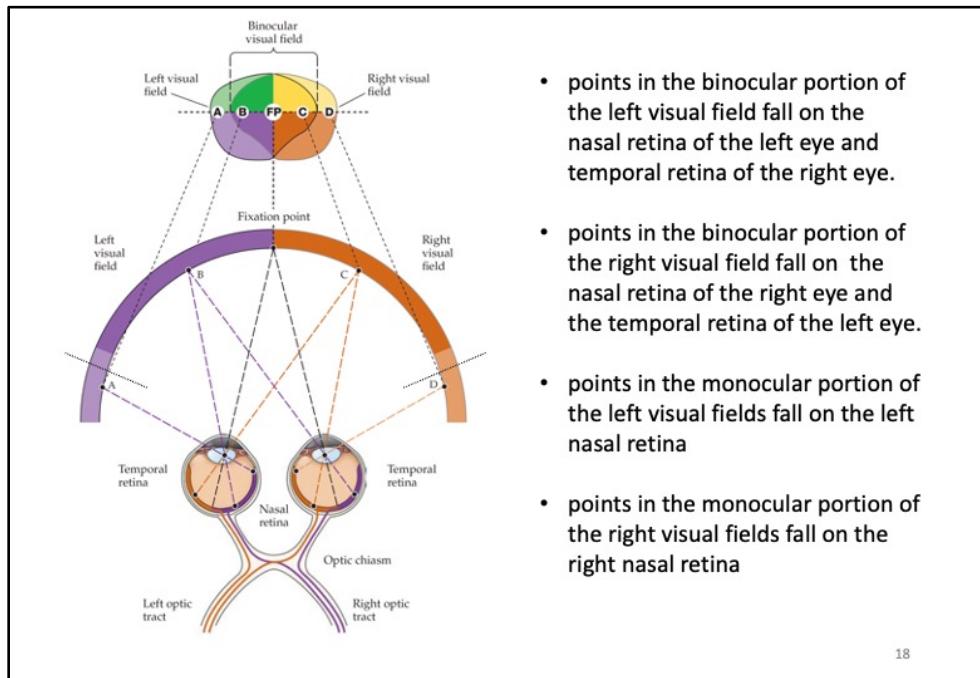
The optic disc, or optic nerve head, is the site where ganglion cell axons converge and exit from the eye forming the optic nerve.



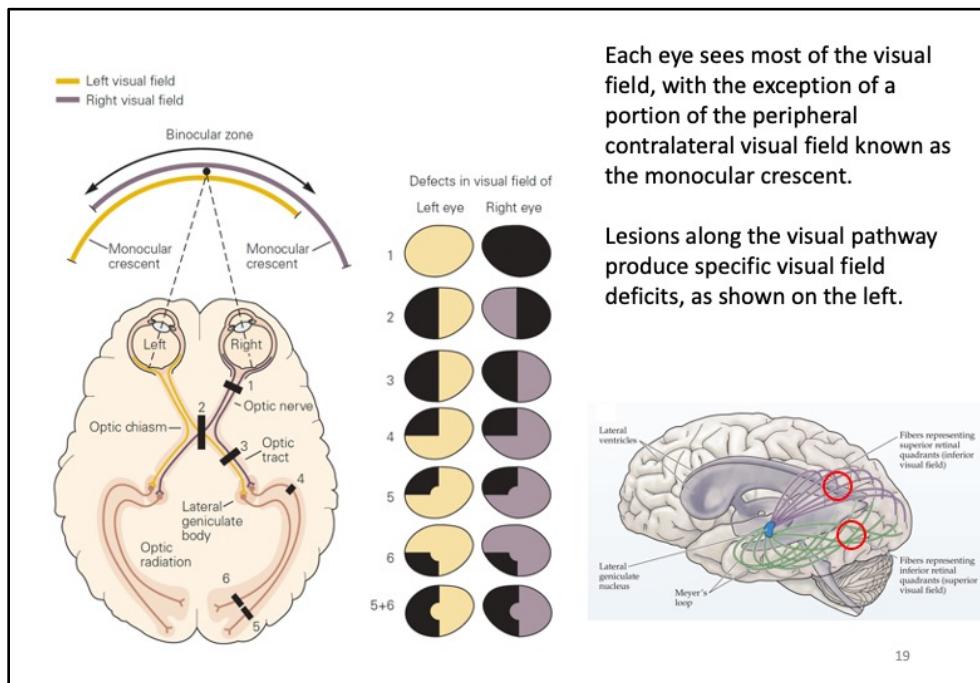
From the eye, the optic nerve extends to a midline crossing point, the optic chiasm.

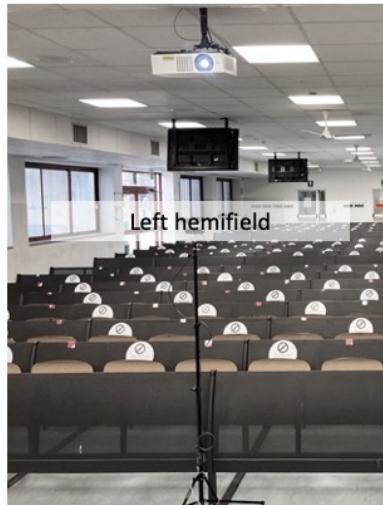
Beyond the chiasm, the fibers from each temporal hemiretina proceed to the ipsilateral hemisphere along the ipsilateral optic tract;

Fibers from the nasal hemiretinas cross to the contralateral hemisphere along the contralateral optic tract.



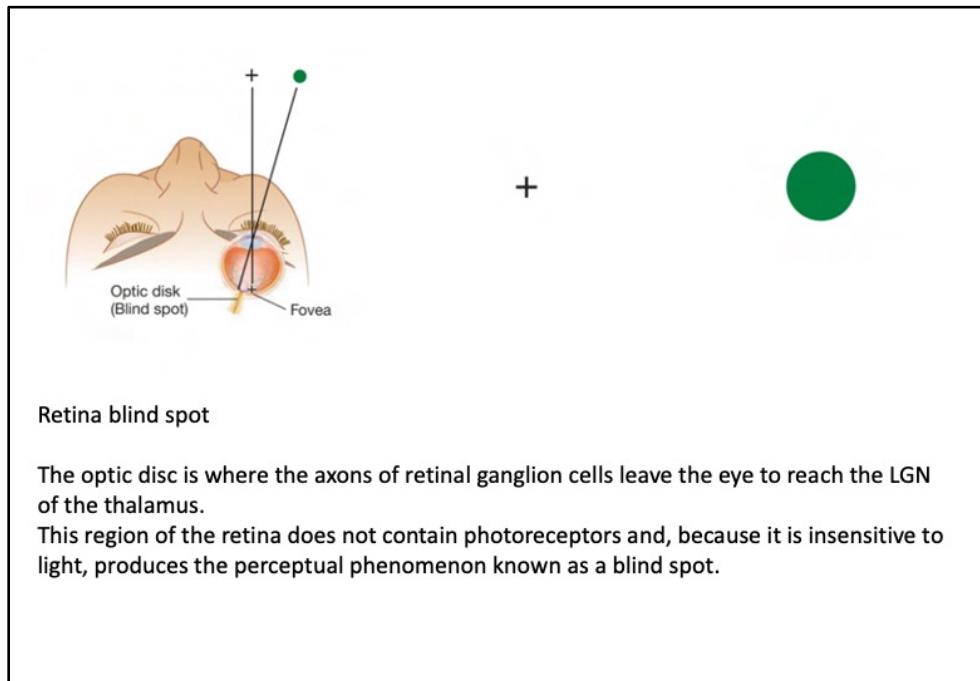
18





Right hemianopia

20



Given the lack of photoreceptors, a defect or scotoma (blind area) of the visual field should be noticed by the observer.

However, the blind spot goes unnoticed for two reasons:

1) Since the optic disc (blind spot) is nasal to the fovea of each eye, light from a single point in the visual field never falls on both blind spots at the same time.
Thus, In binocular vision to compensate for the missing information in the blind spot of one eye, the brain uses information from the homologous region (temporal hemiretinas) of the other eye.

2) In monocular vision, the visual system simply "fills-in" the missing part of the scene (the filling-in phenomenon) using information from the surrounding areas to the blind area.

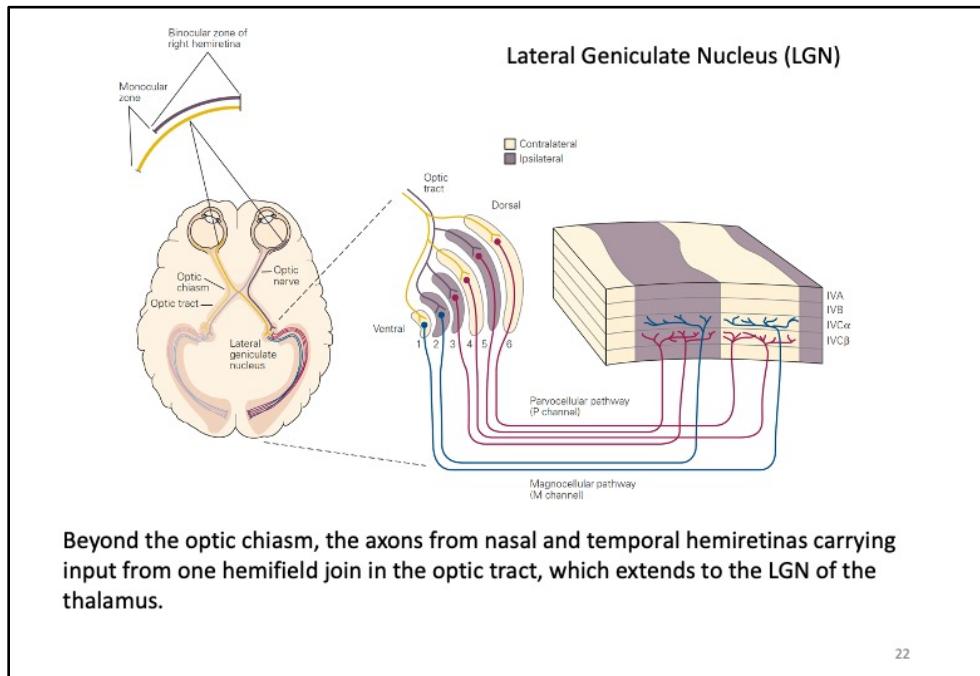
To find your right eye's blind spot:

Close your left eye.

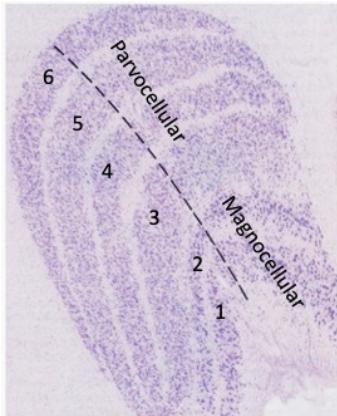
Stare at the cross.

Move closer to the screen (2530 cm), then farther away. Keep doing this until the green circle disappears.

When it disappears, you found your right eye's blind spot.



Lateral Geniculate Nucleus (LGN)



In primates, LGN is a layered structure consisting of six layers, of which two Magnocellular layers (layers 1 and 2), and four Parvocellular layers (layers 3 to 6). Each layer receives input from either the ipsilateral eye (temporal hemiretina, layers 2, 3, 5) or the contralateral eye (nasal hemiretina, layers 1, 4, 6). Since each layer contains a map of the contralateral hemifield, the six maps are stacked on top of each other and in spatial register. The magnocellular layers project to the IVC α layer while the Parvocellular layers project to the IVC β layer of V1. The Koniocellular intercalated layers project to the blobs (layers 1-2) of V1.

Interspersed between the Magno and Parvo layers, are thin but dense layers called Koniocellular which receive from the K retinal ganglion cells.

23

The parallel channels established in the retina remain anatomically segregated even in LGN.

Parvocellular layers receive input from midget retinal ganglion cells, which are most numerous in the primate retina (~ 70%) and carry red-green contrast information.

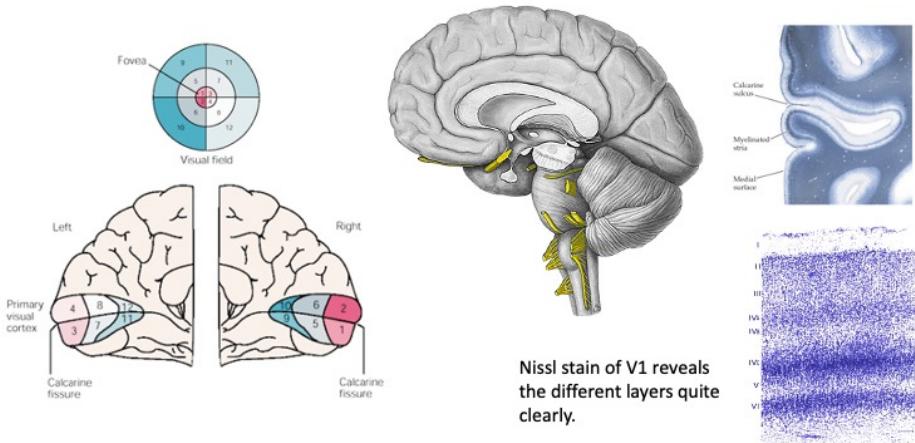
The magnocellular layers receive input from parasol retinal ganglion cells (~ 10%), which carry achromatic contrast information;

Finally, the Koniocellular layers receive input from retinal ganglion cells K (~ 10%), which carry blue-yellow opponent information;

Primary visual cortex (V1)

In humans, V1 (BA17) is located in the occipital portion of the brain along the calcarine fissure of the brain.

V1 constitutes the first level of cortical information processing.



The name striate cortex derives from the Gennari (1782) stripe, a distinctive line (visible to the naked eye) composed of terminations of myelinated axons from the LGN entering into to the IV cortical layer of V1.

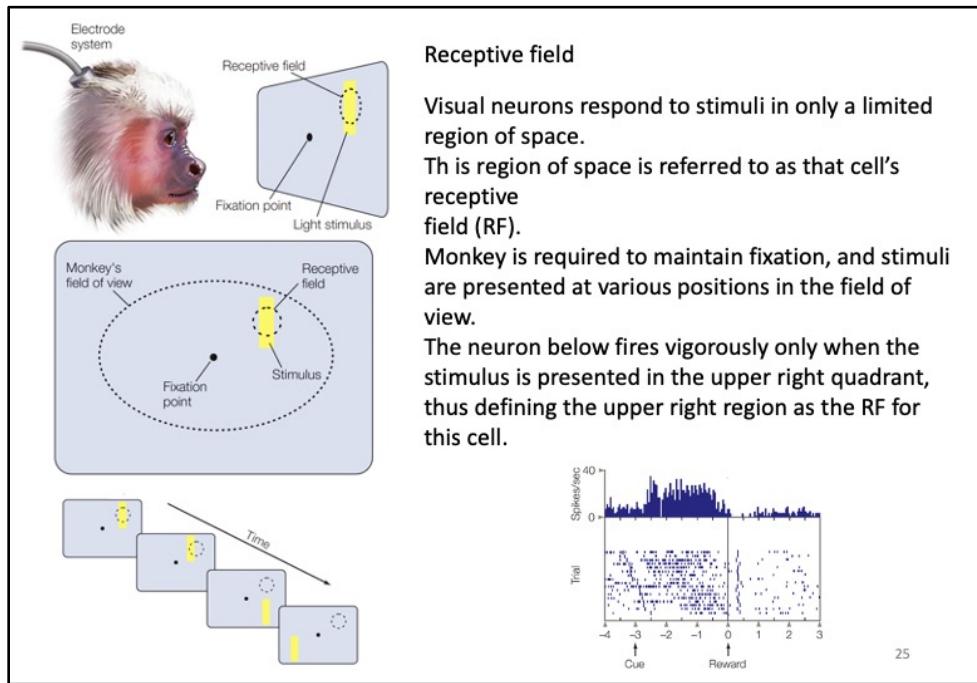
The primary visual cortex is a koniocortex (sensory cortex) divided into six distinct functional layers, numbered 1 to 6.

Layer 4, which receives most of the visual inputs from the LGN, is in turn divided into four layers, called 4A, 4B, 4C α , and 4C β .

Sublayer 4C α receives magnocellular inputs from the LGN, while 4C β receives parvocellular inputs.

Visual area V1 has a visuotopic map of the visual field. For example, the upper bank of the calcarine fissure responds to the lower half of the visual field, the lower bank responds to the upper half of the visual field.

In humans, V1 contains about approximately 140 million neurons per hemisphere (Wandell, 1995), i.e. about 40 V1 neurons per LGN neuron.



The concept of RF was introduced in 1906 by Charles Sherrington. The receptive field is a characteristic of all neurons and, in vision, indicates the region of the visual scene where the stimulus must fall to excite or inhibit the neuron being studied.

Single-cell recording

This technique allows recording signals (firing rate) from single neurons.

A fine-tipped, usually metal (platinum), electrode is inserted in the animal brain to record extracellularly change in electrical activity called action potential (AP, 1ms duration) or spike. Collected signals are appropriately amplified, filtered, viewed through an oscilloscope, and saved to a computer for offline analysis.

Since spikes are all-or-none highly stereotyped signals, most information is encoded in the brain as neuron firing rate, i.e., the number of AP in 1s.

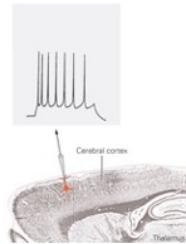
The primary goal of single-cell recording experiments is to determine what experimental manipulations produce a consistent change in the firing rate of an isolated neuron.

Disadvantages

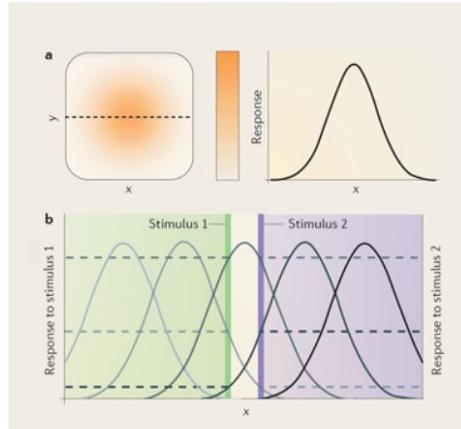
- invasive

Advantages

- high spatial and temporal resolution
- differentiation between excitation and inhibition



26



The receptive field (RF; Charles Sherrington, 1906) of a neuron is defined as the part of stimulus space within which a stimulus elicits a response from the neuron.

In the visual system, the neuron responds to a stimulus presented in a region of space in the visual field (that is, its RF) but not to the same stimulus when it is presented outside this region.

A given stimulus typically elicits the strongest response from the centre of the RF, with the response gradually declining as the stimulus is presented further away from the centre of the RF.

Thus, the RF can be well described by a two-dimensional Gaussian distribution.

27

Figure b, RF profiles of a set of five neurons with different RF locations. Horizontal dashed lines indicate the response of these five example neurons to two stimuli at nearby locations (vertical green and purple lines). Both stimuli fall into the same RF (middle grey curve), but they stimulate neurons with neighbouring RFs differently so that the population can resolve the two locations even though a single neuron cannot. In addition, the size of the RF determines the neuron's spatial frequency tuning: the smaller the RF, the higher the spatial frequency it can resolve.

Organization of receptive fields in the visual system

The receptive field size varies based on:

- position of the neuron along the visual pathway
- eccentricity, i.e., RF position relative to the fovea

At each eccentricity, receptive fields are relatively small at the first levels of visual processing and become progressively larger at later levels.

28

The amount of cortex devoted to one degree of viewing angle changes with eccentricity.

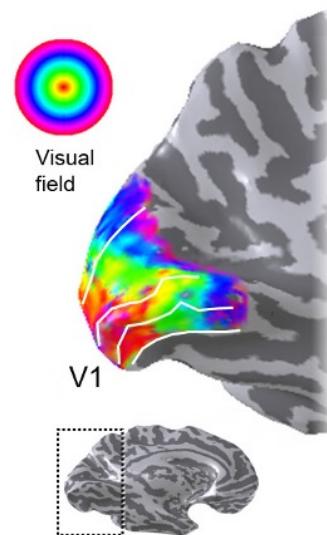
Accordingly, more cortical space is dedicated to the central part of the visual field, where the receptive fields are smaller and densely packed and the visual system has the highest spatial resolution.

Retinotopy

In early visual areas (e.g., V1 to V5), neuron RFs reveal an ordered organization, termed a retinotopic or visuotopic map.

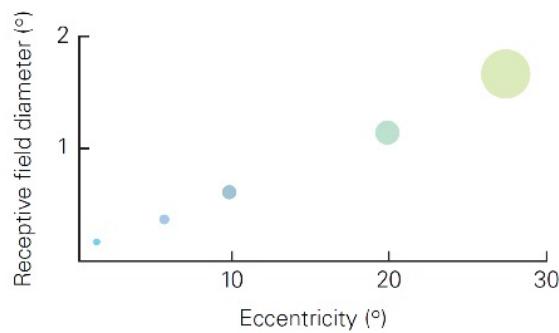
This refers to the existence of a non-random relationship between the position of neurons in the visual areas.

Neuron RFs form a 2D map of the visual field, such that neighbouring regions in the visual image (and therefore on the retinal surface) are represented by adjacent regions of the visual cortical area (i.e., orderly mapping of RF positions in retinotopic coordinates)



Eccentricity

The receptive fields of the retinal ganglion cells that monitor portions of the fovea subtend about 0.1° (equal to 6 min of arc), while those in the visual periphery reach up to 1° of visual angle or more.

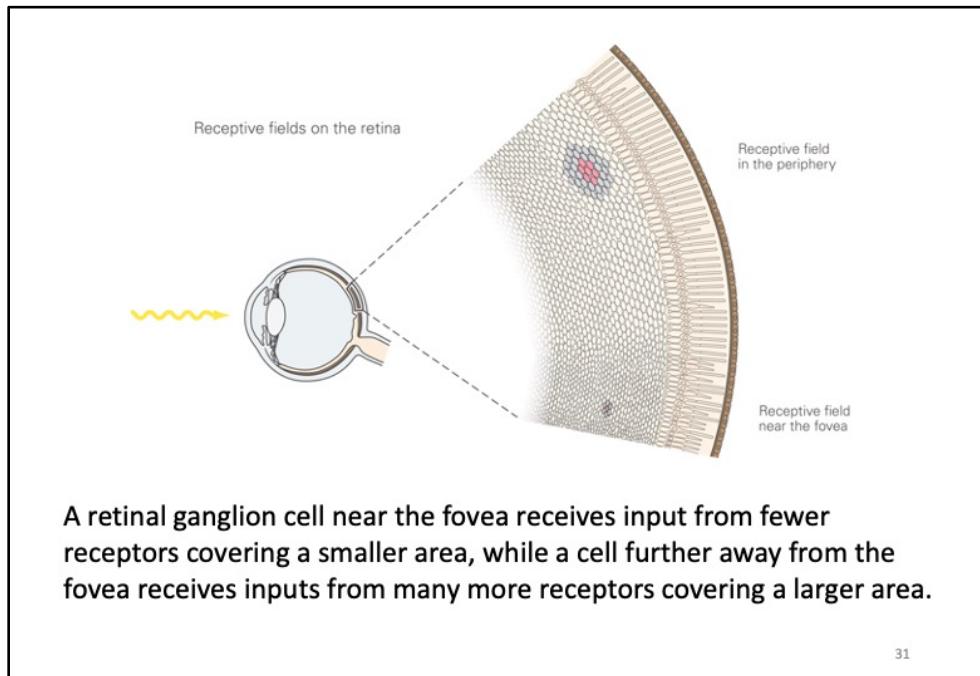


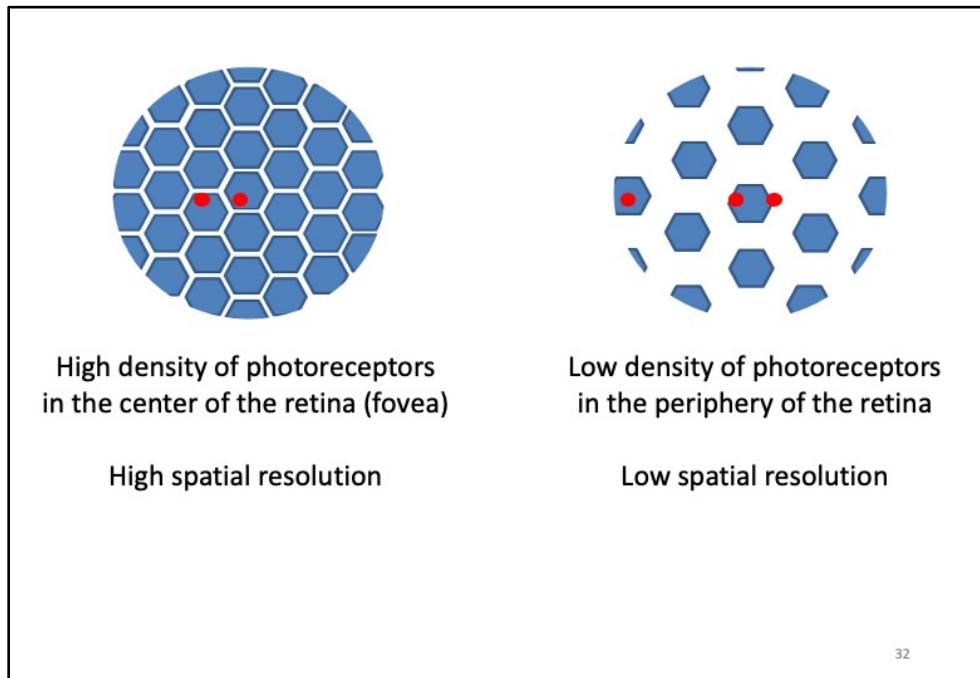
1 Arc min = $1/60$ degree

30

The amount of cortex devoted to one degree of viewing angle changes with eccentricity.

Accordingly, more cortical space is dedicated to the central part of the visual field, where the receptive fields are smaller and densely packed and the visual system has the highest spatial resolution.



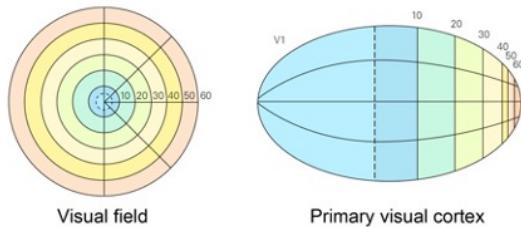


Spatial resolution: minimum distance at which it is possible to recognize as distinct two stimuli presented simultaneously in the visual field.

Cortical magnification

The amount of cortical area devoted to each degree of the visual field, known as the magnification factor, varies with eccentricity (i.e., the neural maps of the visual field are not isometric).

In fact, the central part of the visual field controls the largest area of the cortex. For example, in V1 more cortex is dedicated to the central 10° of the visual space than to everything else.



33

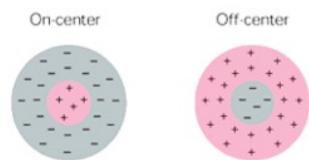
Receptive field properties

Properties change from relay to relay along a visual pathway.

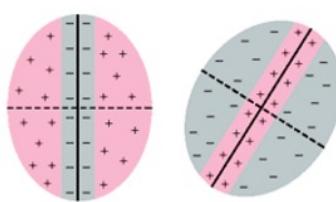
By determining these properties, one can assay the function of each relay nucleus and how visual information is progressively analysed.

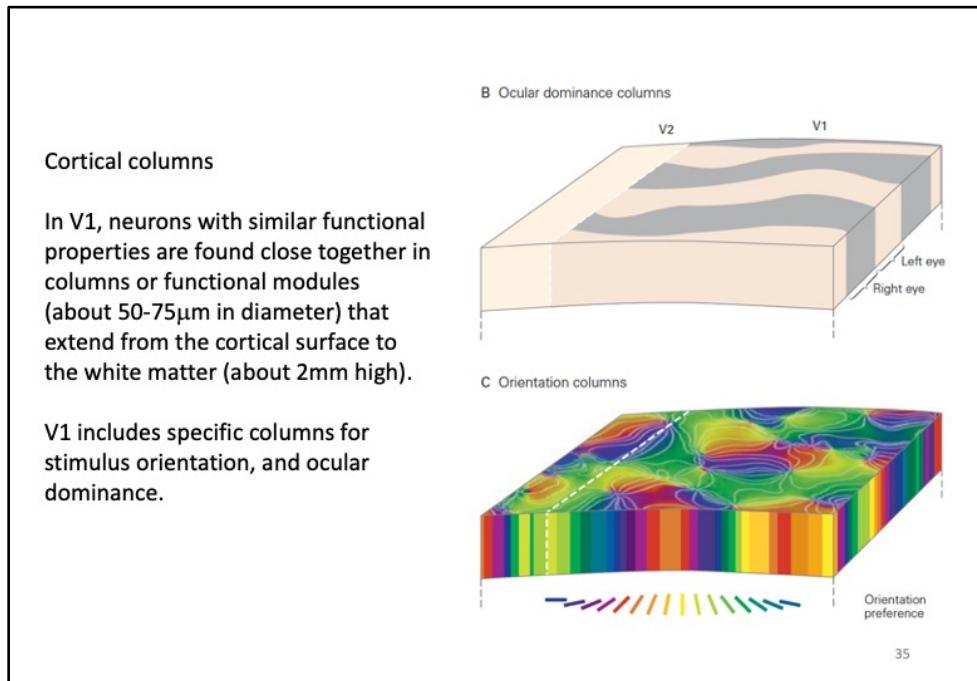
Whereas retinal ganglion cells and neurons in the LGN have concentric center-surround receptive fields, those in the primary visual cortex, although equally sensitive to contrast, analyse oriented contours (orientation selectivity).

A. RFs of retinal ganglion cells and LGN cells



B. RFs of primary visual cortex (V1) cells





There are also proposals for maps of direction selectivity (Weliky et al., 1996; Ohki et al., 2005), selectivity for spatial frequency (Issa et al., 2000) and for disparity (Kara and Boyd, 2009).

Orientation columns

Neurons with the same orientation preference (i.e., vertical) are grouped together into orientation columns. Each column contains a few hundred cells and is 50-75 μm wide;

Moving from one column to the adjacent one, orientation preference changes systematically by 10-15 °, both clockwise and counterclockwise, completing a 180 ° cycle (12 steps) every 750-1000 μm .

The set of columns corresponding to a complete sequence of orientations (a period) is called a hypercolumn.

There are approximately 3-4 thousand hypercolumns, each monitoring a position of the visual field, in accordance with the retinotopic topology.

36

Hubel & Wiesel (1962, 1968) observed that when they advanced a microelectrode along a penetration perpendicular to the cortical surface, successively recorded cells shared an identical orientation of their receptive field axis.

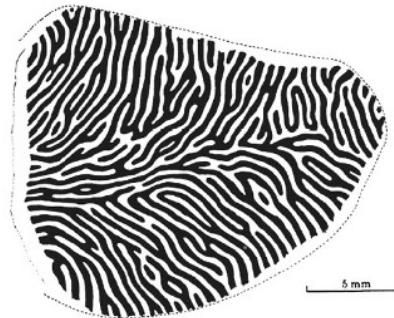
By contrast, when they advanced a microelectrode along a penetration tangential to the cortical surface, they observed that the preferred orientation of cells shifted in steps of 10-15° approximately every 65 μm .

Ocular dominance columns

The ocular dominance columns group neurons that respond more vigorously to stimuli presented to one of the two eyes. They are stripes with an average width of approximately 750 µm, running tangentially for various mm.

The ocular dominance columns reflect the segregation of inputs from different layers of the LGN, which receive inputs from retinal ganglion cells located in the ipsilateral or contralateral retina.

In tangential penetrations of V1, the dominance columns of the left and right eye have been found to alternate regularly with a periodicity of 750 to 1,000µm.



Ocular dominance columns in primary visual cortex (V1) of macaque monkey shown in tangential section. Regions receiving input from one eye are shaded black and regions receiving input from the other eye are unshaded. The dashed line signifies the border between areas V1 and V2 (taken from Hubel and Wiesel, 1977).

37

The function of ocular dominance columns remains an enigma.

One candidate function for ocular dominance columns has been stereopsis (binocular vision).

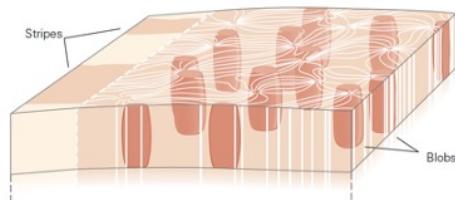
However, it has been reported that squirrel monkeys, which lack ocular dominance stripes have a stereoacuity comparable to that of human observers (Livingstone et al. 1995).

Blob e interblob

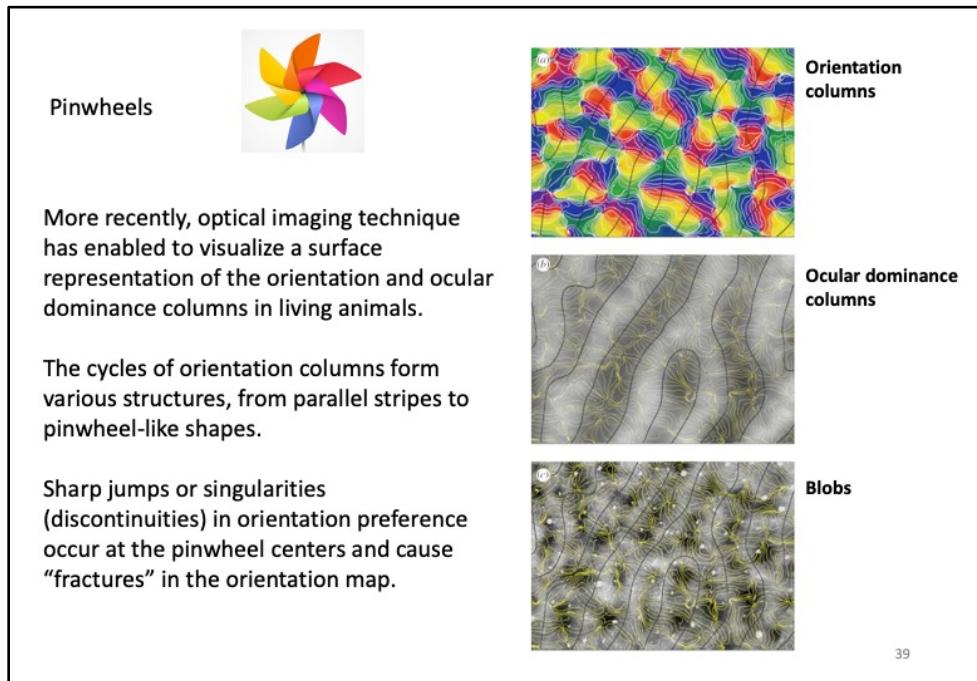
The columns of orientation and ocular dominance include groups of neurons that are poorly selective for orientation (they have circular receptors) but with strong preferences for the color of the stimulus.

These cell groups are located in the superficial layers (II and III) of V1. They are detectable by a specific marker for the cytochrome oxidase (CO) enzyme, which distributes in a regular pattern of regions defined as blobs (CO rich and color responsive) separated by interblob areas (CO poor and orientation responsive).

D Blobs, interblobs (V1), and stripes (V2)



38



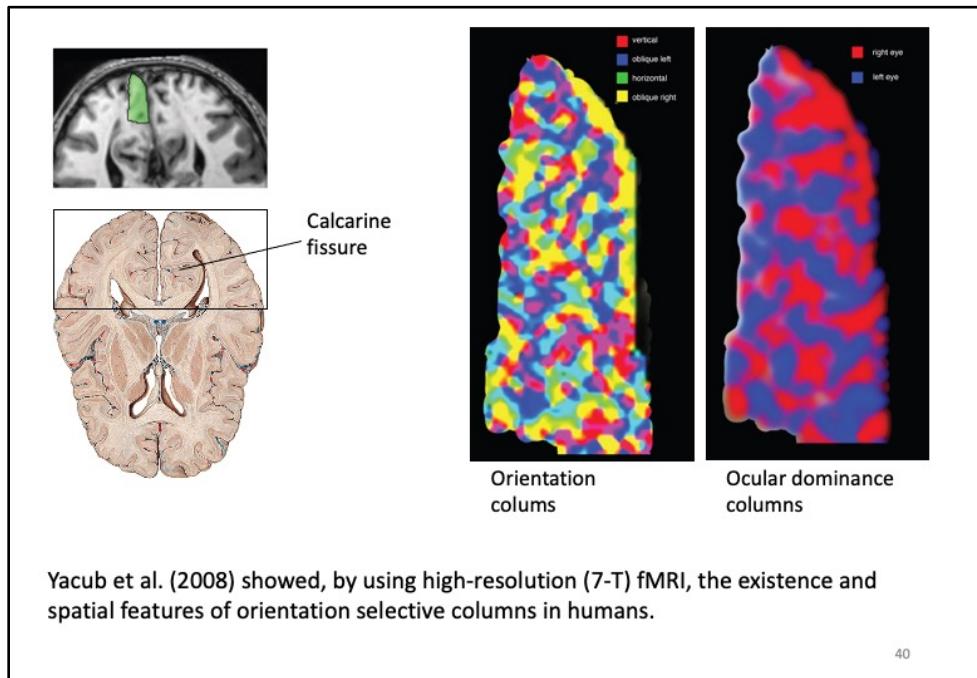
Developed for studies of cortical organization by Amiram Grinvald, this technique visualizes changes in surface reflectance associated with the metabolic requirements of active groups of neurons, known as intrinsic signal optical imaging, or changes in fluorescence of voltage-sensitive dyes.

(a) Optical imaging map of orientation preference in macaque striate cortex. The black lines represent the borders of ocular dominance columns.

(b) Iso-orientation contours from (a) superimposed on an optical imaging map of the ocular dominance columns.

Note that orientation singularities tend to be situated in the centres of ocular dominance columns.

(c) Comparison between iso-orientation contours and CO patches (blobs). In some instances, singularities and blobs coincide.
(from Horton and Adams, 2005).

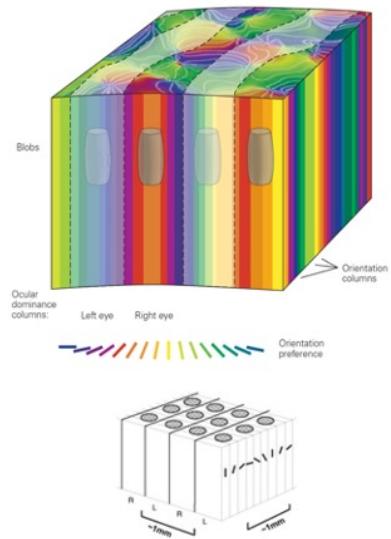


Yacub et al. (PNAS, 2008) demonstrated, by using highfield (7-T) fMRI, the existence and spatial features of orientation selective columns in humans.

Striking similarities were found with the known spatial features of these columns in monkeys.

In addition, it was found that a larger number of orientation columns are devoted to processing orientations around 90° (vertical stimuli with horizontal motion),

Ice cube model (Hubel e Wiesel, 1977)



A region of cortical tissue of about 1mm contains two orientation hypercolumns (a complete cycle of selective vertical columns for orientation), one for the left eye and one for the right that alternate regularly, blob and interblob. This computational module contains all the anatomical-functional types of V1 neurons, and would be repeated thousands of times to cover the entire surface of the visual field.

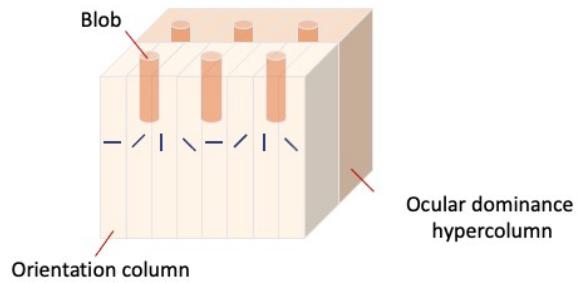
41

However, it remains unclear what advantage, if any, is conveyed by this form of columnar segregation.

One candidate function for cortical columns is the minimization of connection lengths and processing time, which could be evolutionarily important;

The functional organization of the primary visual cortex is therefore based on two systems running orthogonally to each other:

orientation system
ocular dominance system



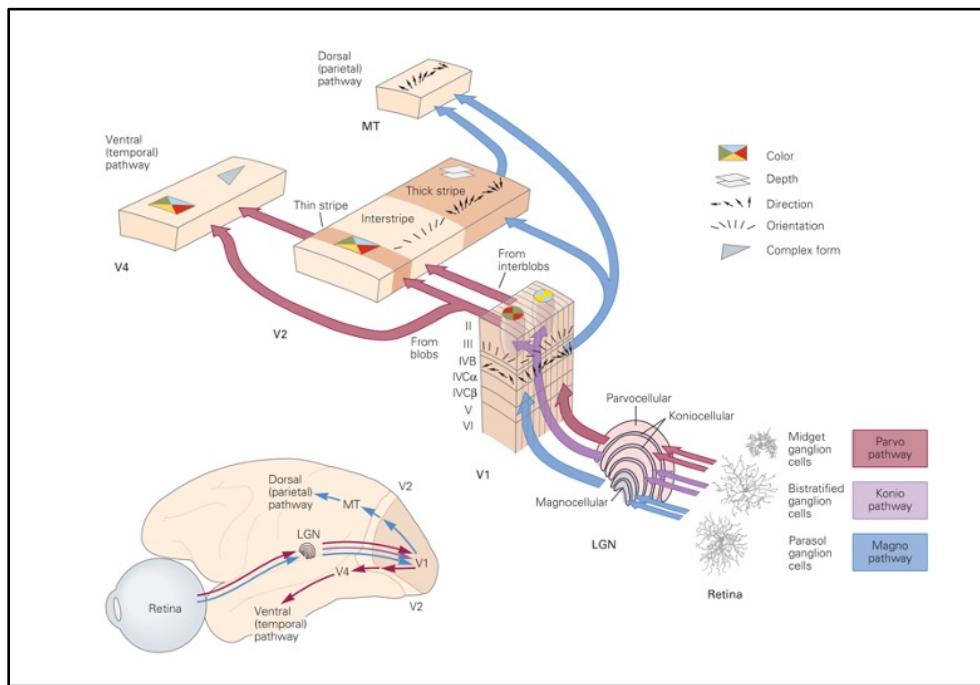
Area V2

In area V2, thick and thin dark stripes separated by pale stripes are evident with cytochrome oxidase labeling.

- The thick stripes contain neurons selective for direction of movement and for binocular disparity;
- The thin stripes contain cells specialized for color.
- The pale stripes contain orientation-selective neurons.

For every visual attribute to be analyzed at each position in the visual field, there must be adequate tiling, or coverage, of neurons with different functional properties.

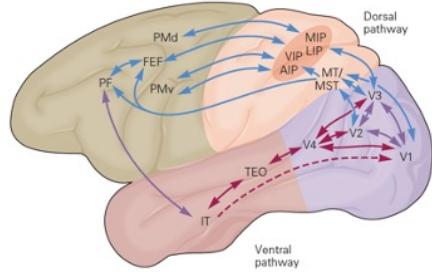
Any given position in the visual field can therefore be analysed adequately in terms of the orientation of contours, the color and direction of movement of objects, and stereoscopic depth by a single computational module.



Beyond V1 are the extrastriate visual areas (more than 30 areas in macaques), a set of higher-order visual areas organized as neural maps of the visual field.

Visual areas are organized in two hierarchical pathways, a ventral pathway involved in object recognition and a dorsal pathway dedicated to the use of visual information for guiding movements.

The ventral or object recognition pathway extends from V1 to the temporal lobe
The dorsal or movement-guidance pathway connects V1 with the parietal lobe and then with the frontal lobes.

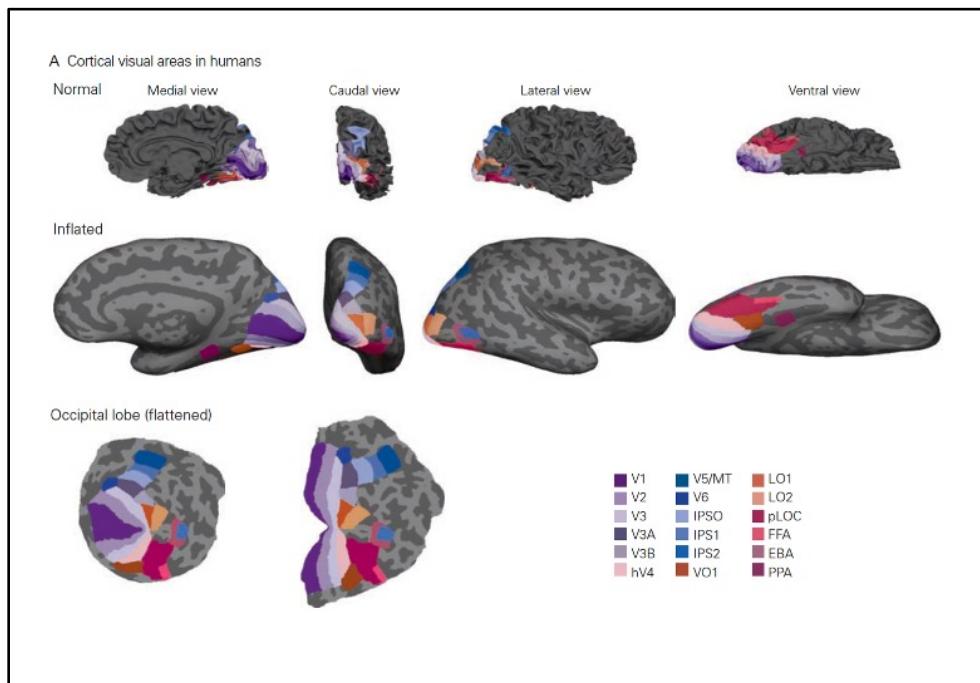


Ungerleider & Mishkin, Two cortical visual systems. 1982

The dorsal and ventral pathways are highly interconnected so that information is shared.

For example, stimulus movement information in the dorsal pathway (area V5) can contribute to object recognition through kinematic cues. Information about movements in space derived from areas in the dorsal pathway is therefore important for the perception of object shape and is fed into the ventral pathway.

Note: all connections between areas in the ventral and dorsal pathways are reciprocal: each area sends information to the areas from which it receives input.
Reciprocity is an important feature of connectivity between cortical areas





At the lowest level, visual attributes such as local contrast, orientation, color and motion are processed.

The intermediate level involves the separation of the visual image into surfaces and outlines (contours), and figures from background distinction.

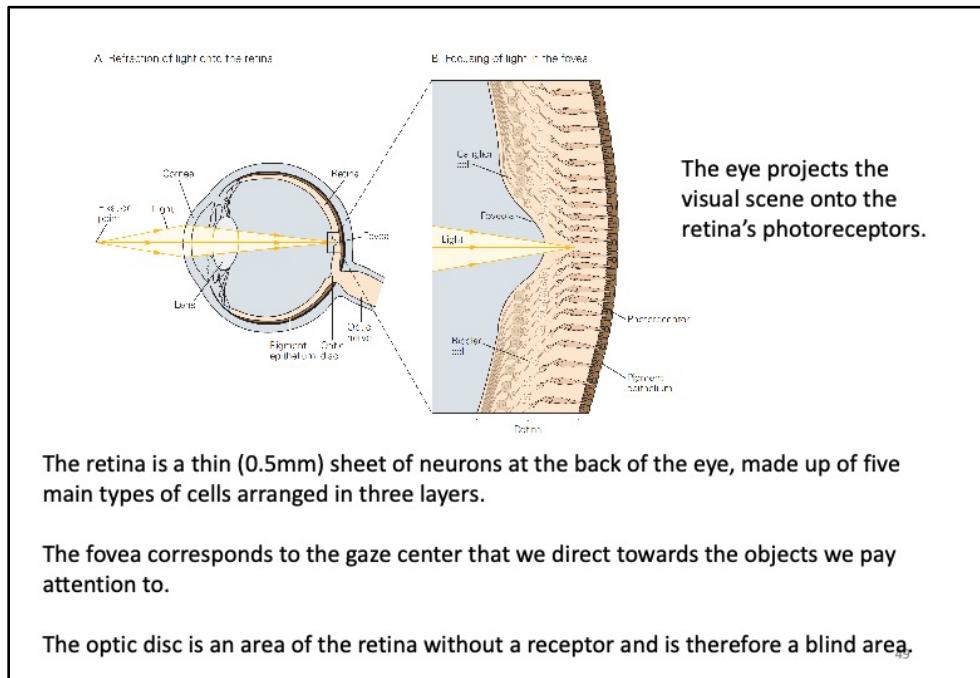
The highest level concerns object recognition.

Once a scene has been analyzed by the brain and the objects have been recognized, the objects can be associated with memories of shapes and their meanings.

Low-level visual processing involves:

retinal circuits
LGN
some (but not all) of the neurons of V1

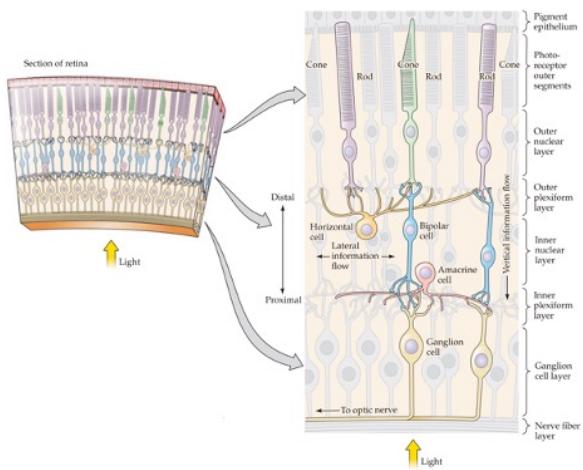
This level allows you to extract some elementary characteristics from the visual image.



A. Light from an object in the visual field is refracted by the cornea and lens and focused onto the retina.

B. In the foveola, corresponding to the very center of gaze, the proximal neurons of the retina are shifted aside so light has direct access to the photoreceptors

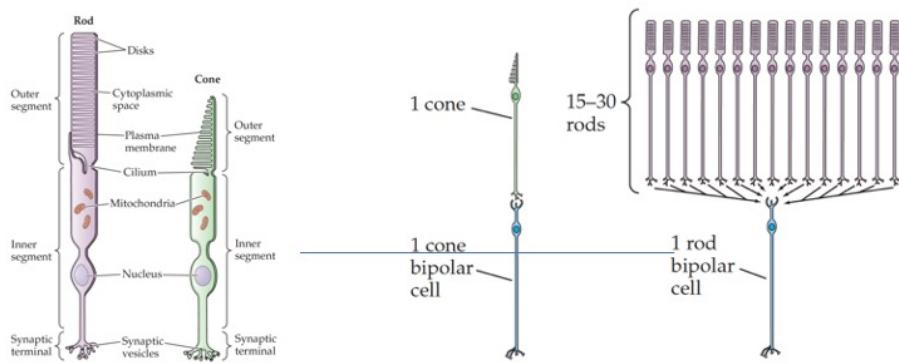
The fovea is the area of the retina where the density of photoreceptors, bipolar cells, and ganglion cells is greatest, and where vision is sharpest (higher two-point resolution).



Basic circuitry of the retina.

A three-neuron chain—photoreceptor, bipolar cell, and retinal ganglion cell—provides the most direct route for transmitting visual information to the brain.

Humans, like most vertebrates, possess two types of photoreceptors, rods and cones, differing in shape, function, connectivity and distribution in the retina.

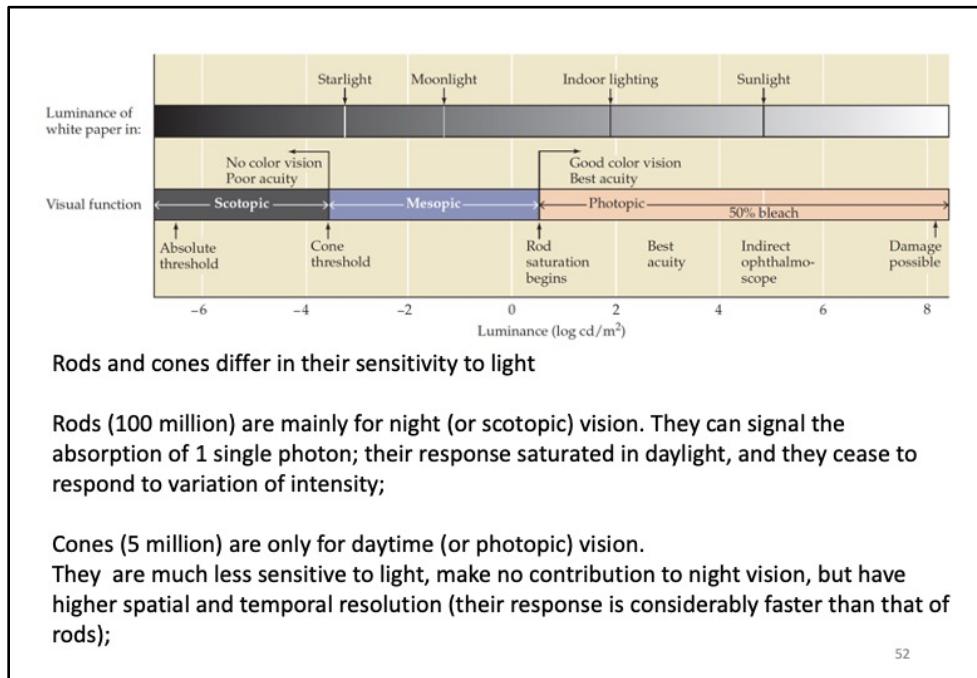


51

Difference in the amount of convergence for the rod and cone system.

In the rod system, the bipolar cell receives synapses from 15 to 30 rods.

In the cone system, in the center of the fovea in particular, each bipolar cell receives its input from a single cone and synapses with a single ganglion cell.



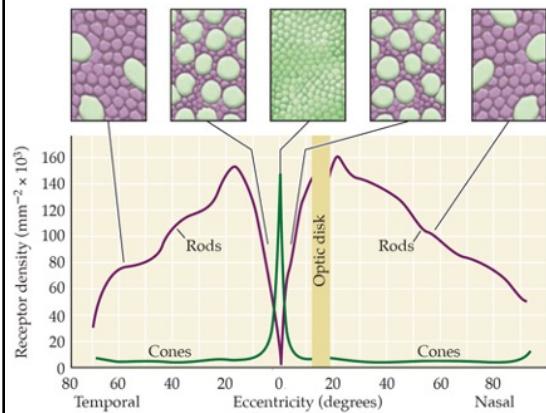
At very low light levels, vision is scotopic: light is detected by the rods of the retina. Rods are extremely sensitive to wavelengths close to 500 nm (blue-green) and play little, if any, role in color vision.

In brighter light conditions, such as daylight, vision is photopic: the light is detected by the cones responsible for color vision.

Cones are sensitive to a range of wavelengths, but are most sensitive to wavelengths close to 555nm (green-yellow).

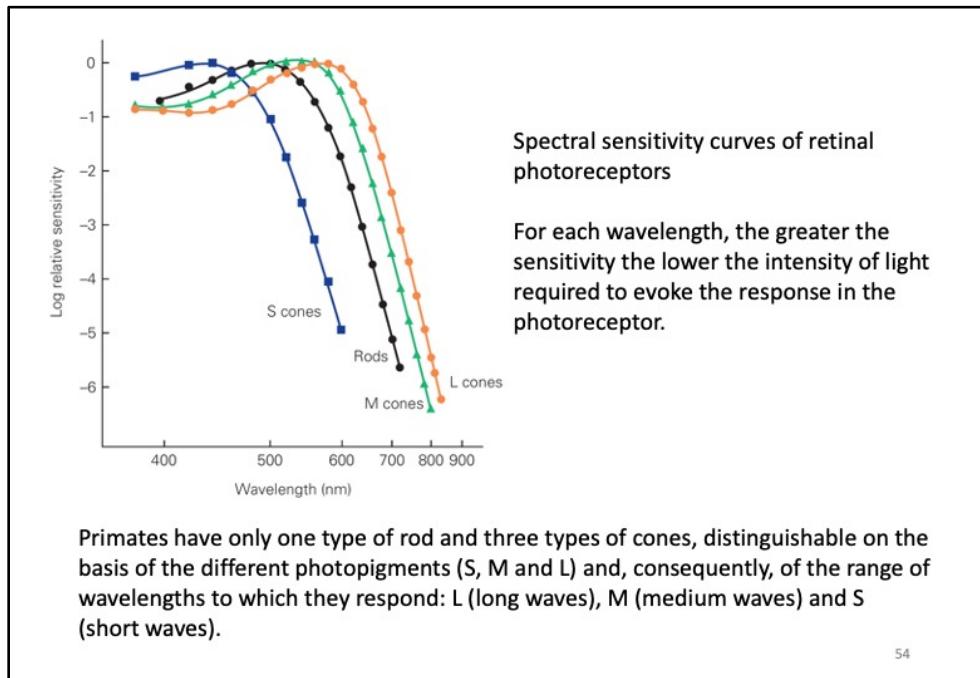
In mesopic view, both rods and cones provide signals to the retinal ganglion cells.

Distribution of photoreceptors in the human retina



Cones are present at a low density throughout the retina, with a sharp peak in the center of the fovea (the foveola).

Conversely, rods are present at high density throughout most of the retina, with a sharp decline in the fovea; rods are absent in the foveola.

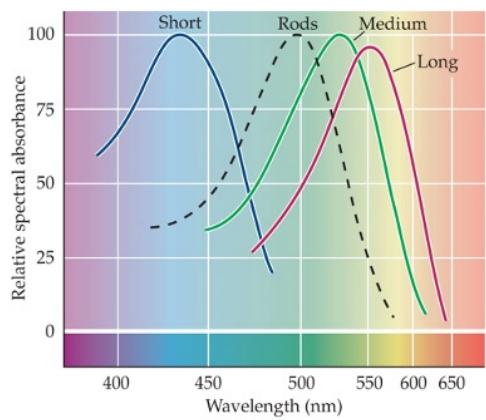


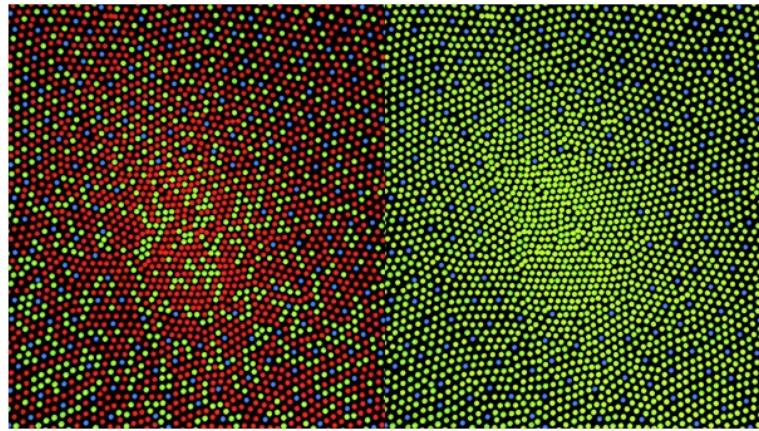
At each wavelength, the sensitivity is inversely proportional to the intensity of light required to elicit a criterion response in the sensory neuron.

Sensitivity varies over a large range and thus is shown on a logarithmic scale. The different classes of photoreceptors are sensitive to broad and overlapping ranges of wavelengths.

Because a single photoreceptor cannot distinguish between a change in the wavelength of light and a change in its intensity, the analysis of colour requires the comparison of signals from different types of cones

In night vision, when only rods are active, a green light consequently has exactly the same effect on the visual system as a red light of a greater intensity.

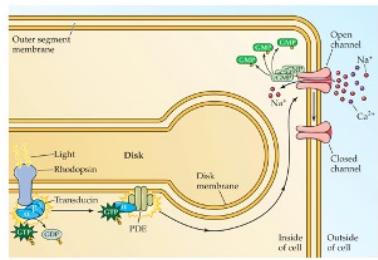
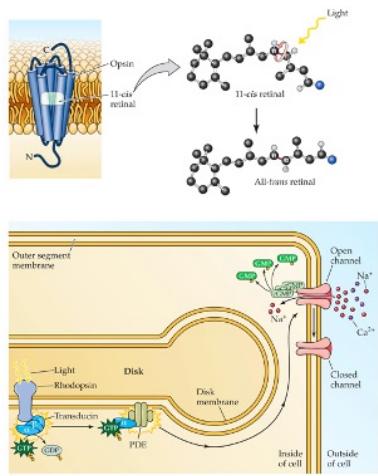




Distribution of cone cells in the fovea of an individual with normal color vision (left), and a color blind (protanopic) retina.
Note that the center of the fovea holds very few blue-sensitive S cones.

56

The S cones make up only 10% of all cones and are absent from the central fovea.

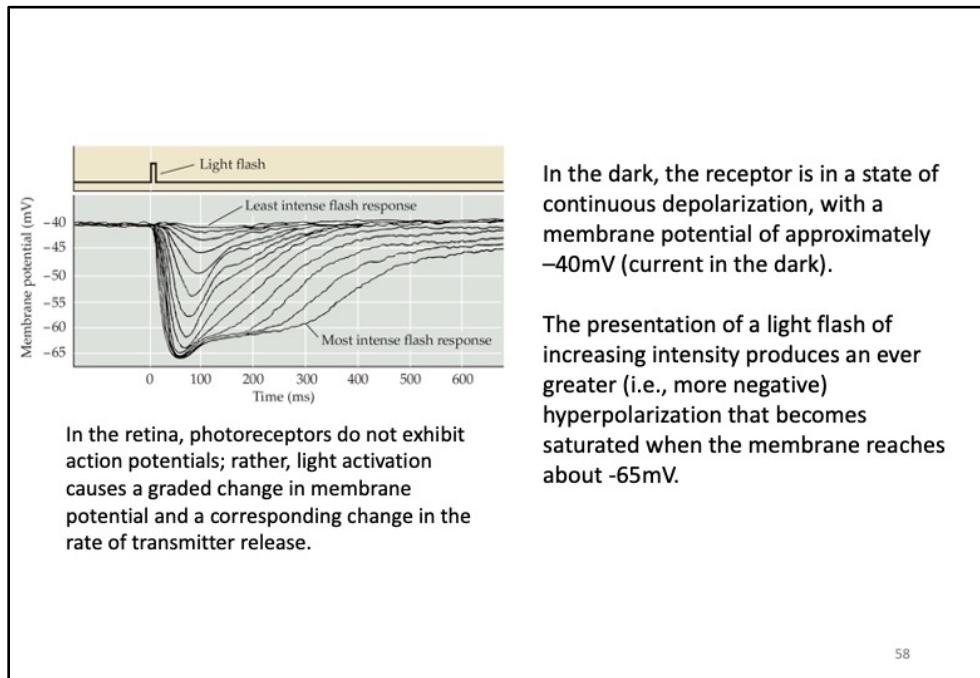


Phototransduction in rod photoreceptors.

Rhodopsin resides in the disk membrane of the photoreceptor. The opsin molecule encloses the light-sensitive retinal molecule.

Absorption of a photon of light by retinal leads to a change in configuration from the 11-cis to the all-trans isomer.

This in turn activates a phosphodiesterase (PDE) which then hydrolyzes cGMP, reducing its concentration and leading to the closure of channels in the outer segment membrane.



58

Dark current

In the dark, the cytoplasmic concentration of GMPc is high, the channels activated by the GMPc are open and allow a constant current of Na^+ to pass inside the cell, called the dark current.

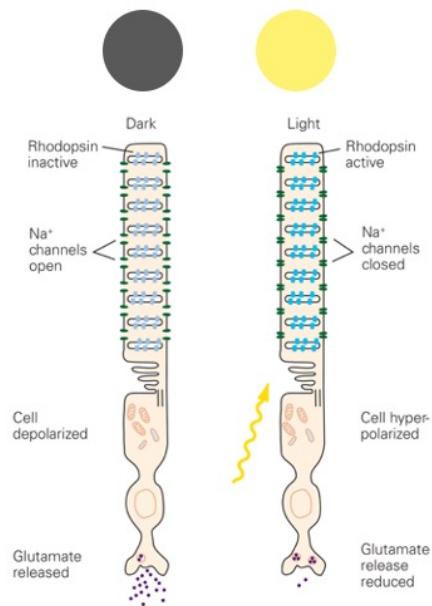
Therefore, in the dark, the membrane potential of the receptor is around -40 mV and is much more depolarized than that of most neurons;

When the light stimulus reduces the concentration of

GMPc by closing the channels controlled by the GMPc, the incoming current of Na⁺ decreases and the cell becomes hyperpolarized (-70 mV).

Photoreceptors produce a relatively simple neural representation of the visual scene:

Neurons in the bright regions are hyperpolarized, while those in the dark regions are depolarized.



RGCs respond to light with action potentials, unlike photoreceptors and bipolar cells.

There are several types of RGCs that differ in shape, connectivity and function:

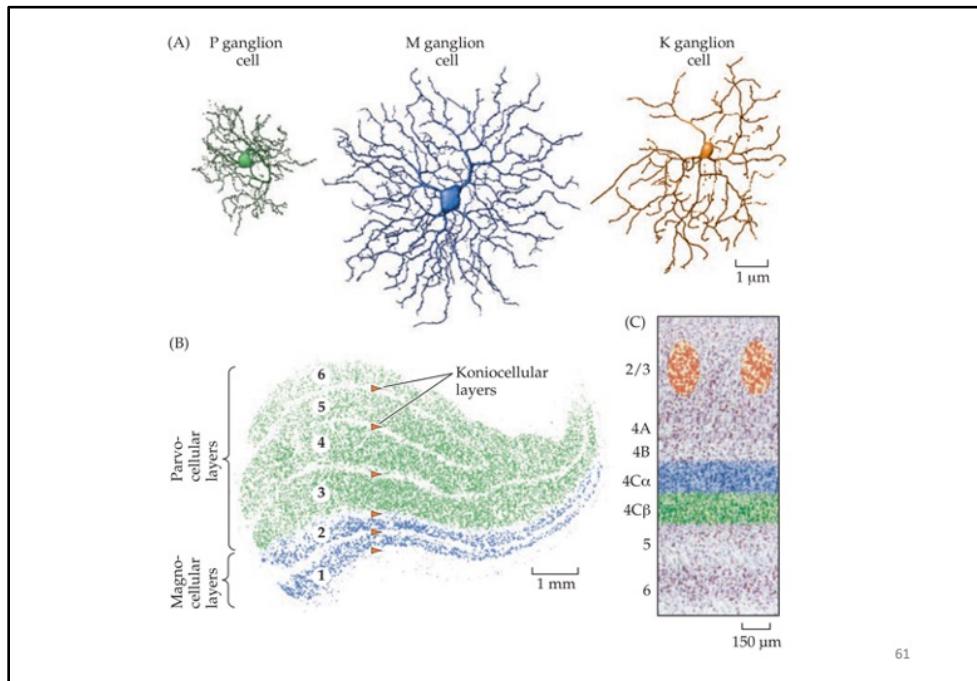
Midget or P (parvo) cells are the most numerous (70%), have smaller cell bodies and dendritic fields, and supply the parvocellular layers of LGN;

Parasol or M (magno) cells, about 10%, have large-diameter cell bodies and large dendritic fields supply the magnocellular layers of LGN;

K cells, about 10%, have small cell bodies and intermediate-sized dendritic fields. They supply the koniocellular layers of LGN;

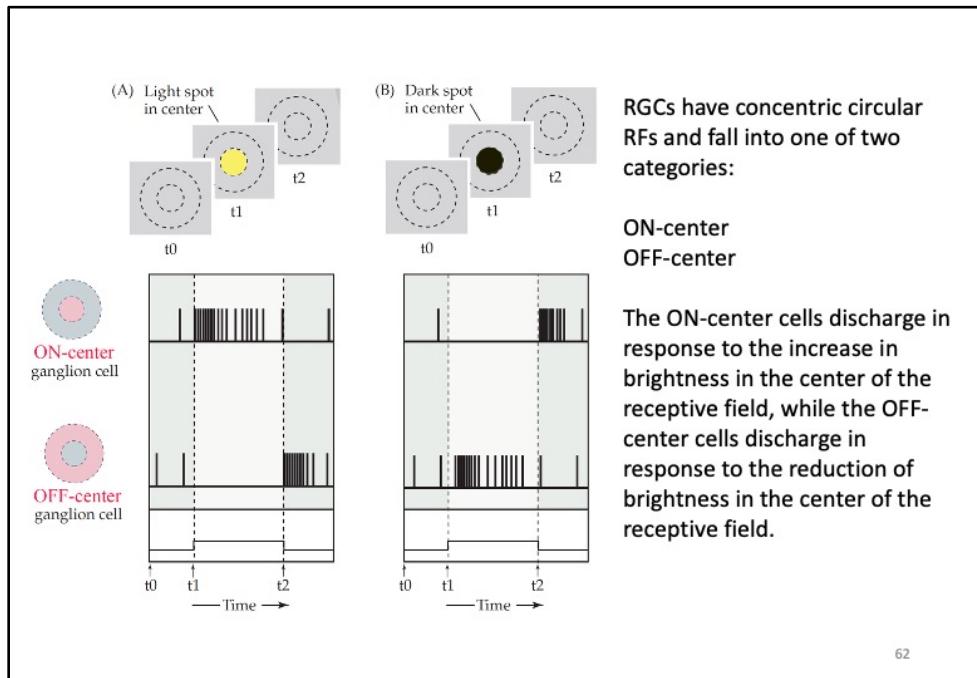
The mechanisms by which the brain deciphers the spatial patterns of light and dark that fall on the photoreceptors have been a vexing problem.

To understand what the complex neural circuits within the retina accomplish during this process, it is useful to begin by considering the responses of individual retinal ganglion cells to small spots of light.



RGCs constitute the last level of retinal processing and project information, through the optic nerve, directly to the LGN of the thalamus.

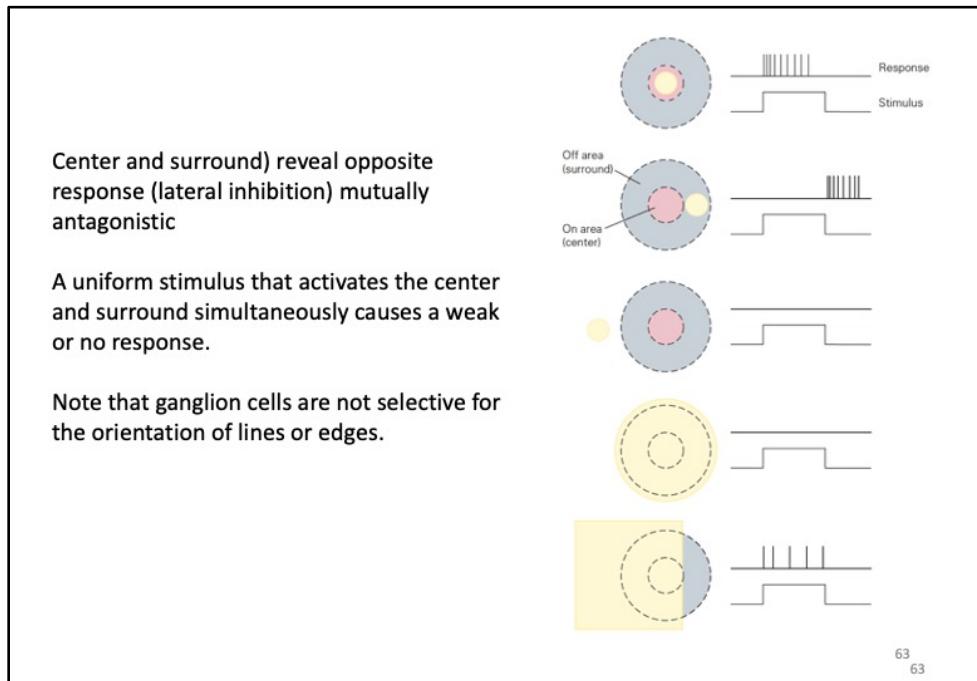
- A) Types of retinal ganglion cells
- B) LGN
- C) Termination of lateral geniculate axons in striate cortex. Magnocellular layers terminate in layer 4C α , parvocellular layers terminate in layer 4C β , and koniocellular layers terminate in layers 2 and 3.



ON-Center ganglion cells are excited by a light stimulus in the center of the receptive field;
OFF-Center ganglion cells are excited by a dark stimulus in the center of the receptive field;

Note that the firing rate of ON-center ganglion cells increases soon after the dark stimulus disappears;

Similarly, the discharge rate of OFF-center ganglion cells increases soon after the disappearance of the light stimulus;

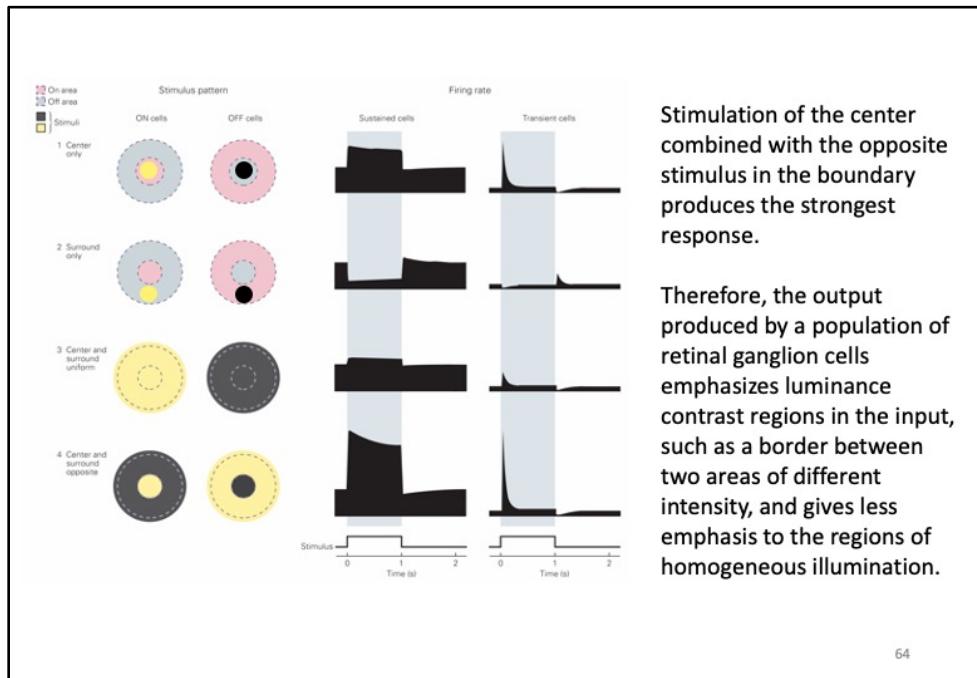


The retinal ganglion cells have an organization of the receptive field with two concentric circular areas with opposite and antagonistic response.

In ON-center cells, the illumination of the central part of the receptive field causes an excitatory response, i.e. an increase in the discharge of the cell, while the illumination of the surrounding part of the receptive field causes an inhibitory response (mechanism of lateral inhibition).

The OFF-center cells are instead organized in the opposite way: the illumination of the surrounding area causes an excitatory response, while the illumination of the central part of the receptive field causes an inhibitory response.

The simultaneous illumination (or darkness) of the center and surround does not evoke a variation in the discharge frequency.



Transient ganglion (M cells) produce a spike discharge only at the start of the stimulus;

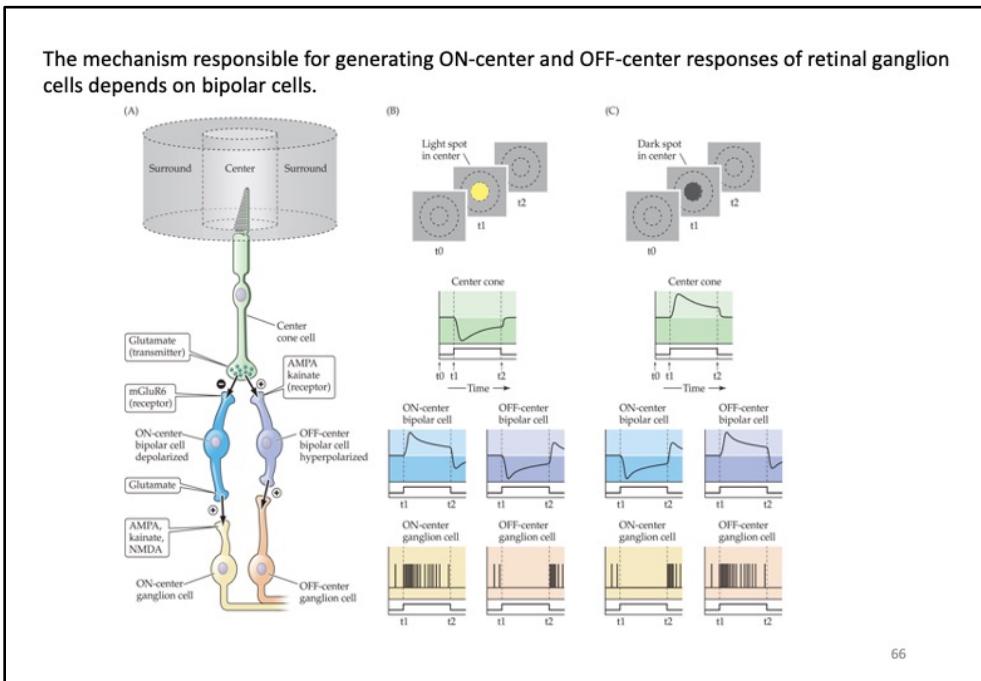
Sustained neurons (P cells) maintain the discharge almost constant for several seconds during light stimulation.

In general, however, ganglion cell output favors temporal changes in visual input over periods of constant light intensity.

The visual image that is stabilized on the retina with an eye-tracking device disappears from view in a few seconds.

However, this does not happen in normal vision; even when we keep our gaze fixed, since small automatic eye movements (microsaccades) continuously scan the image through the retina to prevent the disappearance of the visual image.

- RGCs respond only weakly to uniform stimulation.
- RGC mainly emphasize the contrasts of brightness present in the visual scene, and not the absolute intensity of the lighting itself.
- Most of the useful information in a visual scene lies in the distribution of brightness contrasts (edges).
- In fact, the absolute light values reflected by the different objects give very little information as they depend on the intensity of the lighting source.



66

There are two classes of bipolar cells: ON- and OFF-center bipolar cells.

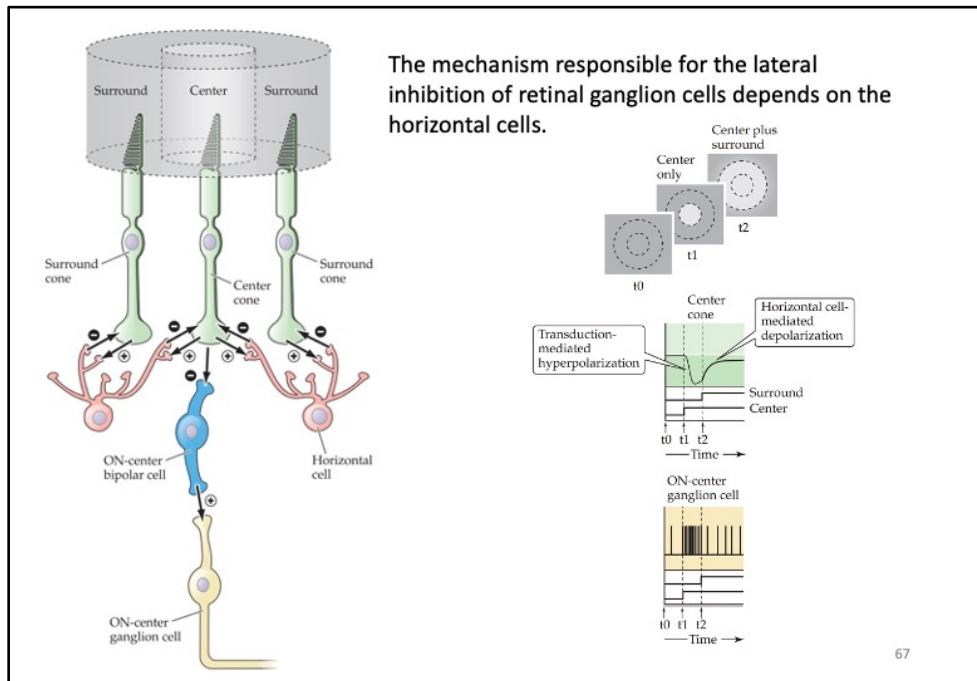
The selective response of ON- and OFF-center bipolar cells to light increments and decrements is explained by the fact that they express different types of glutamate receptors

ON-center bipolar cells express metabotropics receptor that cause the cells to hyperpolarize in response to glutamate (sign-inverting synapse).

OFF-center bipolar cells have ionotropic receptors that cause the cells to depolarize in response to glutamate (sign-conserving synapse).

Thus, glutamate has opposite effects on these two classes of cells, depolarizing OFF-center bipolar cells and hyperpolarizing ON-center cells.

The ON-center bipolar cells are activated by the light stimulus, while the OFF-center bipolar cells are inhibited by the light stimulus.



The center-surround antagonism (lateral inhibition) observed in RGCs is thought to be mediated by horizontal cells.

Horizontal cells receive synaptic inputs from photoreceptors. Although the mechanism of their action is not entirely clear, horizontal cells are thought to exert their influence on the photoreceptor terminals, by regulating the amount of transmitter that the photoreceptors release onto the bipolar cell.

The release of glutamate from the photoreceptor terminals has a depolarizing effect on the horizontal cells (synapse which preserves the sign), while the horizontal cells have a hyperpolarizing influence on the terminals of the photoreceptors (synapse which reverses the sign).

Consequently, the net effect of the horizontal cell is to oppose light-induced changes in the photoreceptor membrane potential (inhibitory effect).

The figure shows how these events lead to an inhibitory peripheral ring surrounding the receptive field center of the ON-center ganglion cell.

A light stimulus circumscribed and centered on a photoreceptor, which provides input to the center of the receptive field of the ON-center ganglion cells, produces a

hyperpolarization in the photoreceptor.

Under these conditions, the inhibitory effects produced by the horizontal cell are relatively small and the response of the photoreceptor to light remains very high.

With the addition of light in the peripheral ring, however, the inhibitory impact of the horizontal cells becomes significantly greater.

In fact, a large light stimulus involving photoreceptors that provide input to the center and periphery of the receptive field of the ON-center ganglion cells produces a strong hyperpolarization of the horizontal cells that converge on the photoreceptor terminal connected with the center of the receptive field.

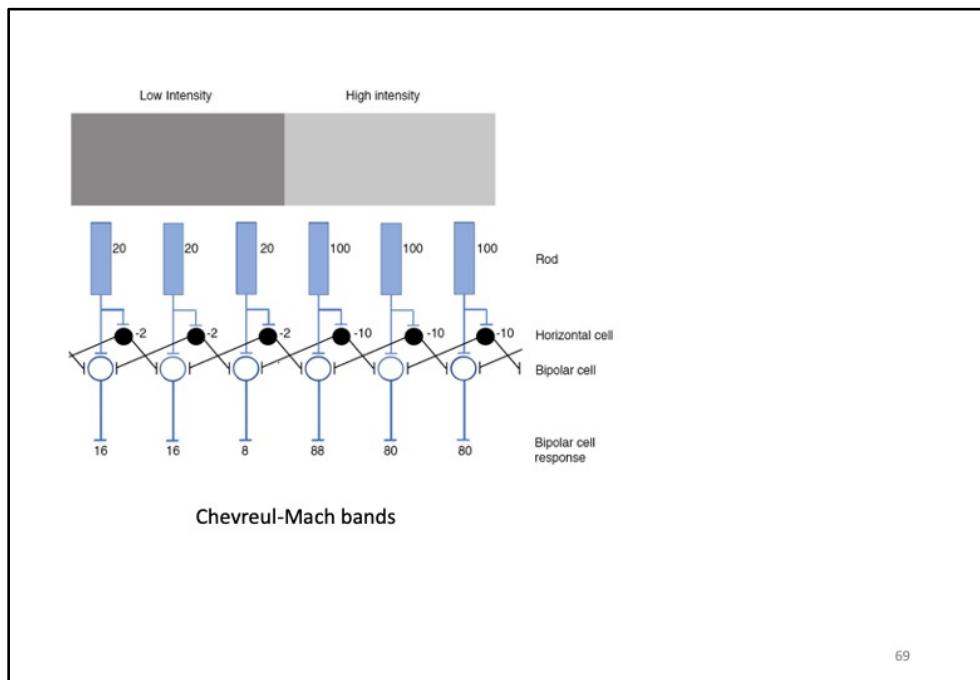
The reduction in activity (hyperpolarization) of horizontal cells has a depolarizing effect on the central photoreceptor, reducing the effect evoked by light and, ultimately, the discharge frequency of the ON-center ganglion cell.

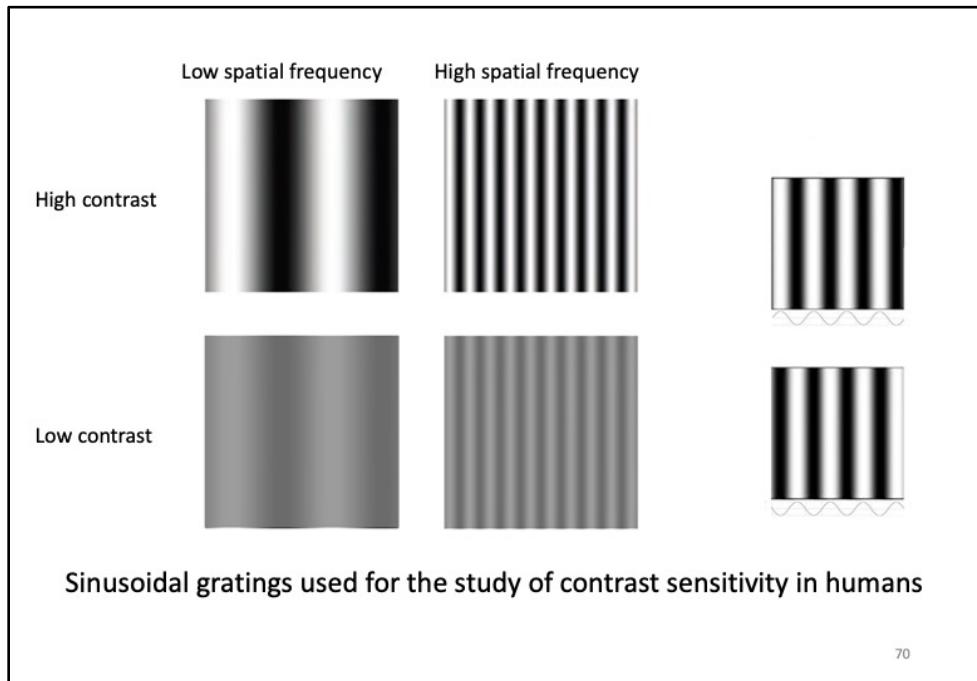
The lateral inhibition mechanism mediated by the horizontal cells enhances the contrast at each border, thus increasing the ability to see objects that contrast only weakly with the background, as in the image below.



Chevreul-Mach bands

Chevreul-Mach band phenomenon (1865) is an optical illusion that makes the band at the edge with the next lighter band appear darker (and the band at the margin with the previous darker band lighter), although each band is of a uniform gray color. It is due to the lateral inhibition mechanism that exaggerates the contrast between figure and background, highlighting the margin ('edge-detection' mechanism).



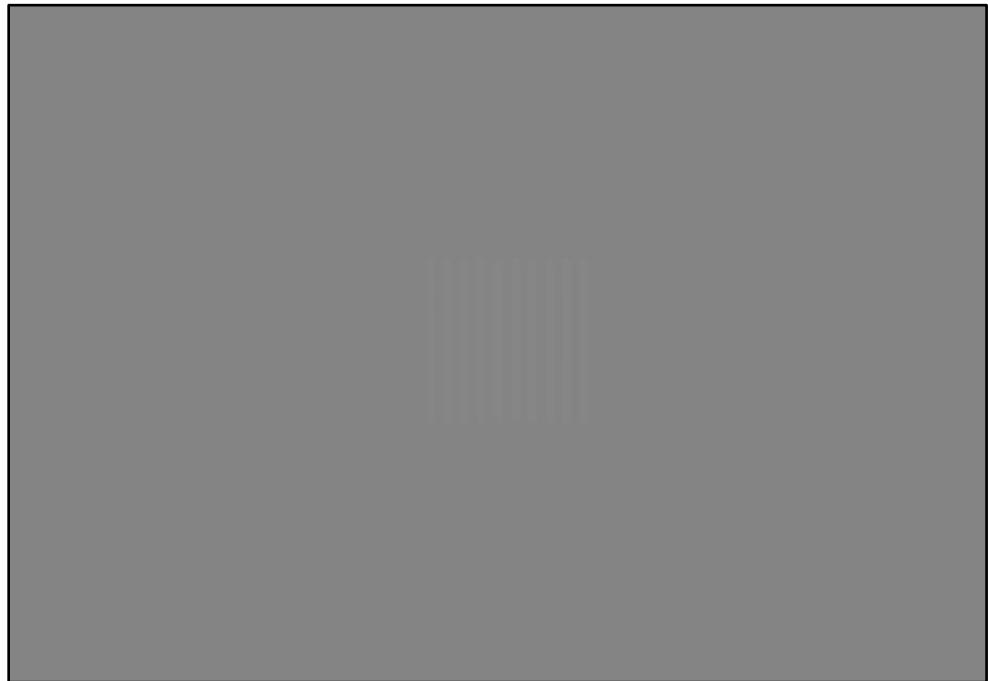


Sinusoidal gratings are stimuli commonly used for the study of the visual system in animals (properties of the receptive field of neurons) and in humans (psychophysics). In the gratings, the intensity (brightness) of the stimulus varies in a sinusoidal way, creating light and dark bands that are repeated in space.

For each grating it is possible to vary independently:

orientation of the bands (light and dark areas, arranged vertically, obliquely, etc.);
 contrast (the difference in brightness between light and dark bands);
 spatial frequency (the width of a pair of light and dark bands or cycle), is measured in cycles per degree of visual angle;
 phase (the position of the light and dark bands in the stimulus, i.e. whether the grating begins with a light or dark band) is measured in radians.

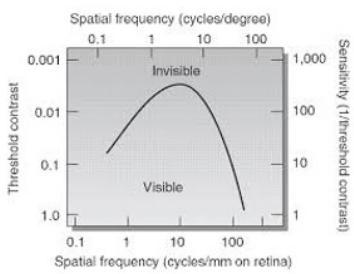
The two gratings on the right in the figure are identical in orientation, contrast and spatial frequency but have opposite phase,



To determine the sensitivity of a human observer to different spatial frequencies, the threshold is measured.

The contrast of the sinusoidal gratings is progressively reduced until the observer stops perceiving the stimulus.

Sensitivity is given by the inverse of the threshold ($1 / \text{threshold}$).



The contrast sensitivity (CSF) function describes an observer's sensitivity to sinusoidal gratings as a function of their spatial frequency.

This is measured using a contrast detection experiment in which the minimum (threshold) contrast required to detect sinusoidal gratings of various spatial frequencies is determined.

Sensitivity is defined as $1 / (\text{threshold contrast})$ (so if the threshold is low, the sensitivity is high).

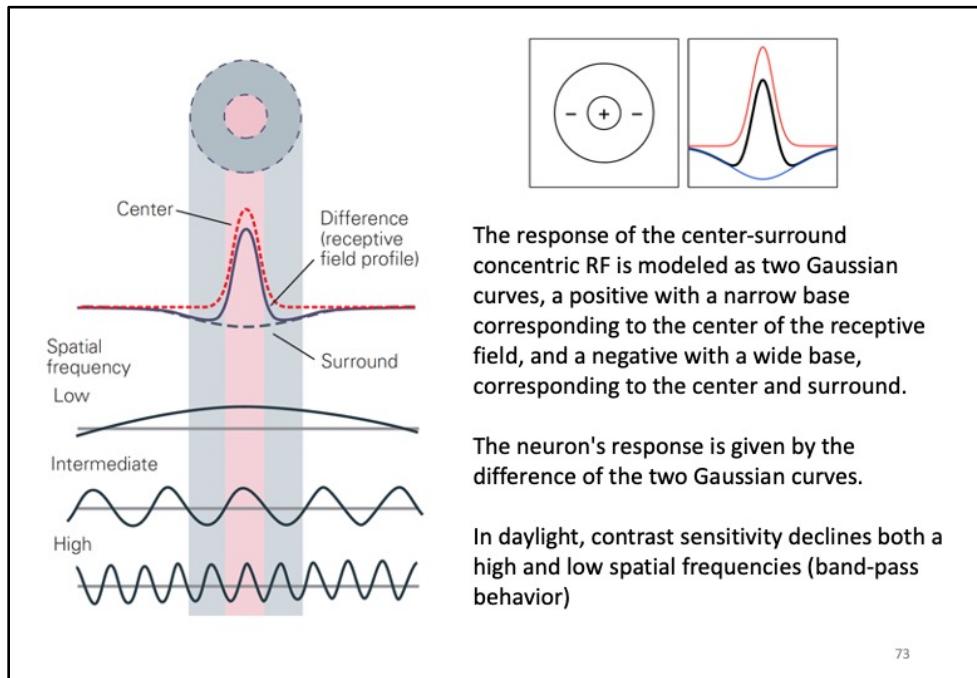
72

Humans are more sensitive to an intermediate range of spatial frequencies (about 4-6 cycles / degree) and less sensitive to both lower and higher space frequencies.

Gratings with a frequency of about 5 cycles per degree are the most visible. The visual system is said to have bandpass behavior because it rejects everything but a narrow band of spatial frequencies.

In the figure above, the stimulus contrast increases from top to bottom, while the spatial frequency increases from left to right.

The central bars in the figure (medium spatial frequency) are visible even at low contrast, while the wide bars and narrow bars are visible only at high contrast.



In humans, if sinusoidal gratings are used, sensitivity is greater for spatial frequencies around 5-8 cycles / visual degree, and is attenuated both for higher frequencies (up to acuity around 30-50 cycles / degree) and for frequencies less than 1 cycle / degree.

Multiplying the profile of the grating stimulus (intensity vs position) with the profile of the receptive field (sensitivity vs position) and integrating over all space calculates the stimulus strength delivered by a particular grating.

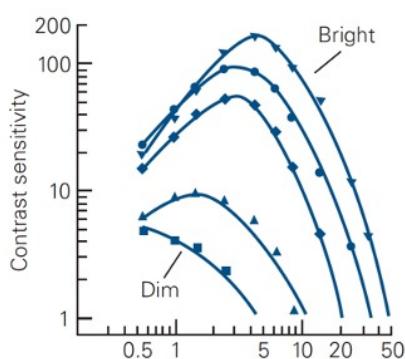
In day light, contrast sensitivity declines sharply at high spatial frequencies, with an absolute threshold at approximately 50 cycles per degree.

Interestingly, sensitivity also declines at low spatial frequencies. The attenuation at low frequencies reflects the inhibitory and antagonistic action of the periphery (surround) of the receptive fields of the retina, geniculate and cortex.

Patterns with a frequency of approximately 5 cycles per degree are most visible.

The visual system is said to have band-pass behavior because it rejects all but a band of spatial frequencies.

Human subject



In dim light, the visual system's contrast sensitivity declines, but more so at high than at low spatial frequencies.

Thus, the peak sensitivity shifts to lower spatial frequencies.

In this state, the visual system has so-called low-pass behavior, for it preferentially encodes stimuli of low spatial frequency.

The fact that in dim light the receptive fields of ganglion cells lose their antagonistic surrounds explains the transition from band-pass to low-pass spatial filtering

P cells

dense in central retina (fovea)
narrow dendritic trees and thin axons
cone inputs
small receptive fields
slow and sustained responses
color opponency
low contrast sensitivity
respond to high spatial frequencies from small areas of the visual field

M cells

more common in peripheral retina
large dendritic trees and thick axon
rod input
wide receptive fields
fast and transitory responses
no color opponency
high contrast sensitivity
respond to low spatial frequencies from large areas of the visual field

Magnocellular	Parvocellular
Large RF	Small RF
ON- and OFF-center	ON- and OFF-center
High contrast sensitivity	Low contrast sensitivity
Color blind	Color sensitive
Transient	Sustained
Low spatial resolution	High spatial resolution
High temporal resolution	Low temporal resolution
Global aspect	Local aspect
Fast and raw	Slow and detailed

76

Selective P cell lesion
complete loss of color perception
deficits high spatial frequencies discrimination (reduced visual acuity)

Selective M cell lesion
reduced perception of movement
Deficits high temporal frequencies discrimination;

The contribution to the visual perception of the magno- and parvocellular systems was experimentally tested by examining the visual abilities of monkeys after selective lesion of the magno- or parvocellular layers of LGN.

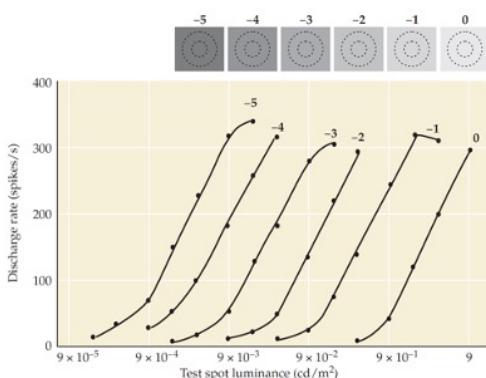
The lesion of the magnocellular layers has little effect on visual acuity or color vision but drastically reduces the ability to perceive the rapid change of stimuli.

On the contrary, the lesion of the parvocellular layers has no effect on the perception of movement but seriously compromises visual acuity and color perception.

These observations suggest that the visual information conveyed by the parvocellular system is particularly important for high spatial resolution - the detailed analysis of an object's shape, size and color.

The magnocellular system, on the other hand, appears critical for activities that require high temporal resolution, such as assessing the position, speed and direction of a rapidly moving object.

Retinal ganglion cells have adjustable operating range



Sigmoid curves illustrating the discharge rate of a single ON-center ganglion cell to the onset of a small test spot of light in the center of its receptive field.

Each curve represents the discharge rate evoked by spots of varying intensity at a constant background level.

The response rate is proportional to stimulus intensity over a range of 1 log unit, but the operating range shifts to the right as the background level of illumination increases.

Sakmann and Creutzfeldt, Pflugers Arch, 1969

78

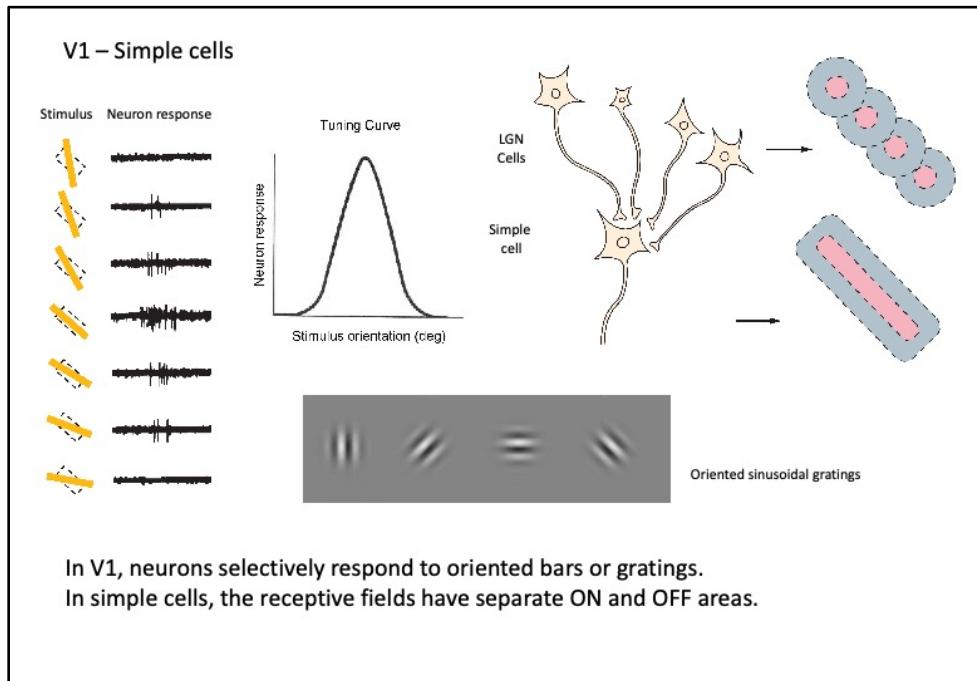
By scaling the ganglion cell's response to prevailing levels of illumination (adjusting the gain), the entire dynamic range of a ganglion cell's firing rate can be used to encode information about intensity differences over the range of luminance values relevant for visual scene.

The primary visual cortex (V1) has several functional categories of cells:

simple cells

complex cells

end-stopped cells



Neurons in area V1 are classically divided into two types: simple and complex (Hubel and Wiesel, 1959).

Neurons have elongated RFs and respond to a narrow range of orientations.

Different neurons respond optimally to distinct orientations (orientation tuning curve).

Example of a neuron in area V1 that selectively responds to lines that adapt to the orientation of its receptive field.

This selectivity is the first step in the brain's analysis of the shape of an object.

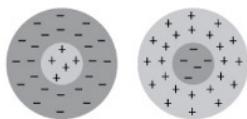
The orientation of the receptive field is thought to result from the alignment of the center-surround circular receptive fields of different LGN cells.

In the monkey, the neurons of the LGN have non-oriented circular receptive fields.

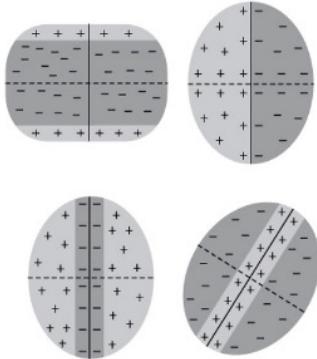
However, projections of adjacent LGN cells onto a simple cell create a receptive field with a specific orientation.

Simple cells respond well to sinusoidal gratings (Gabor patches) of specific spatial frequencies and phases.

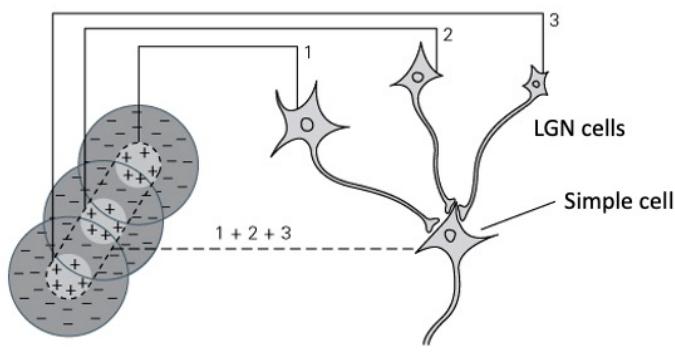
Lateral geniculate nucleus



Primary visual cortex



The receptive fields of the simple cells of the primary visual cortex are different and less homogeneous than those of the ganglion cells of the retina and the LGN



The linear receptive fields of simple cells of the primary visual cortex arise from the convergence of multiple cells of the lateral geniculate nucleus on a single simple cell.

Complex cells

Have rectangular receptive fields, larger than those of simple cells;

Respond to linear stimuli with specific orientation;

The position of the stimulus within the receptive field is not critical as the demarcation between on and off zones is not so clear;

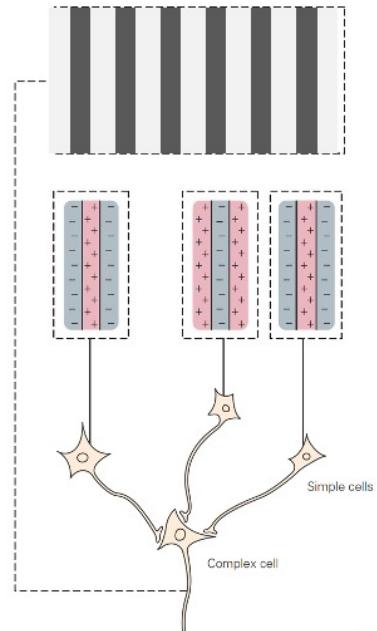
Movement of the stimulus in the receptive field is particularly effective in activating the cells;

Complex cells selectively respond to stimuli that move in particular directions;

V1 – Complex cells

In complex cells, the ON and OFF regions are superimposed, i.e. each position in the receptive field responds to both white and black bars, and the cells respond when a line or edge crosses the receptive field along an axis perpendicular to the orientation of the receptive field.

This constancy in the response to variations of stimulus location in the RF is commonly called position invariance.



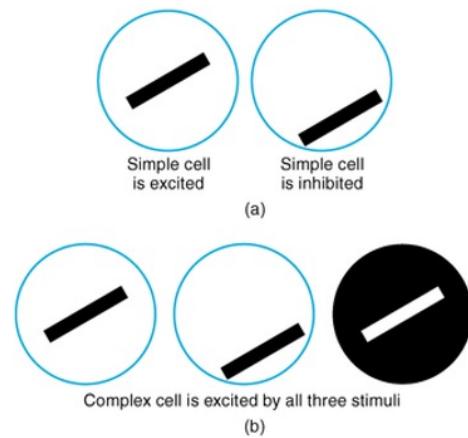
84

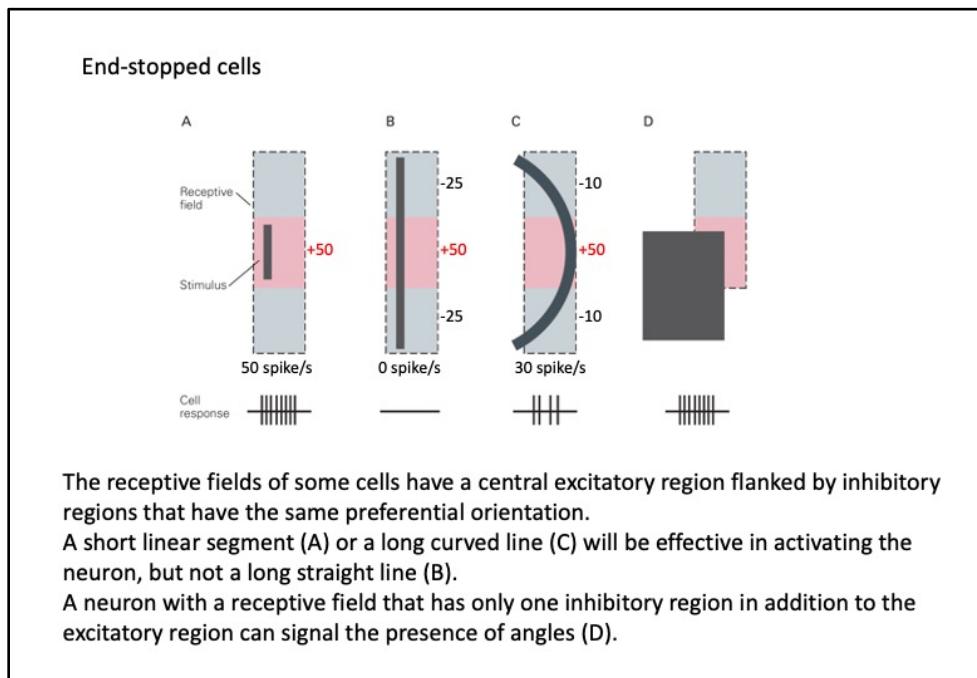
Complex cells are less selective for the position of the stimulus in the receptive field

The receptive field has no defined ON and OFF regions and responds similarly to light (on a dark background) or dark (on a light background) stimuli in all positions of the receptive field.

They are activated as a linear oriented stimulus crosses their receptive fields in one direction.

► Response Characteristics of Neurons to Orientation
in the Primary Visual Cortex





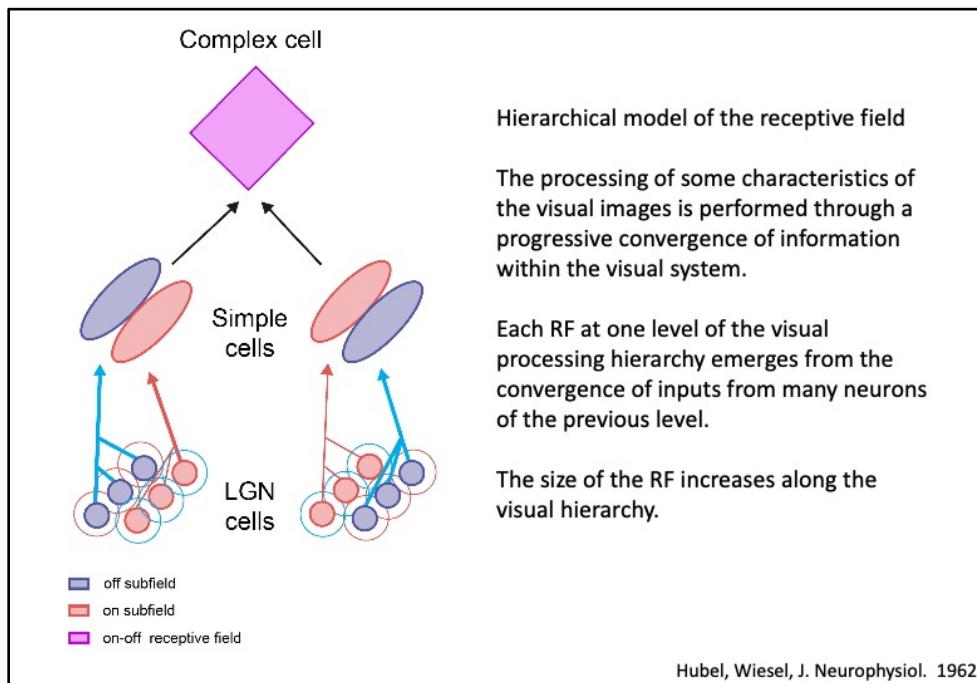
Respond better to linear stimuli of a certain length, or that have an end that does not extend beyond a specific portion of the cell's receptive field.
End-stopped may serve to detect angles ("angle-detectors") or curved lines of visual images.

ON and OFF regions of the RF have the same preferred orientation (vertical, in the neuron illustrated in the figure).

Therefore, the inhibitory effect is greater if the same oriented contour is presented both in the ON and OFF regions.

A short linear segment (A), or a long curved line (C) will be effective in activating the neuron, because excitation will be greater than inhibition.

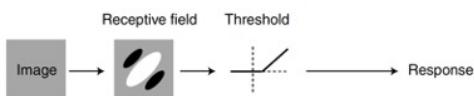
On the contrary, a long straight line (B) will not be effective, because excitation will be canceled by the inhibitory effect.



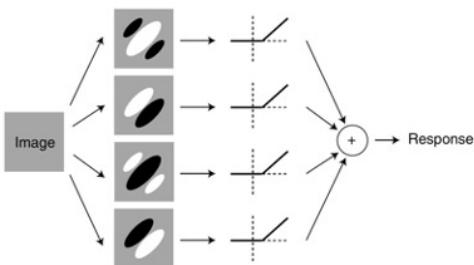
According to the hierarchical model (Hubel and Wiesel, 1962), simple cell receptive fields are constructed from the convergence of geniculate inputs with receptive fields aligned in the visual space.

In turn, complex receptive fields arise from the convergence of simple cells with similar orientation preferences.

A Simple cell



B Complex cell



The models of simple and complex cells proposed by Movshon, Thompson and Tolhurst (Movshon et al. 1978)

A, simple cells. The first stage is linear filtering, i. e. a weighted sum of the image intensities, with weights given by the receptive field. The second stage is rectification: only the part of the responses that is larger than a threshold is seen in the firing rate response.

B, complex cells. The first stage is linear filtering by a number of receptive fields such as those of simple cells (here we show four of them with spatial phases offset by 90 deg). The subsequent stages involve rectification, and then summation.

In addition to stimulus position, V1 neurons are selective for a number of attributes:

Orientation. This selectivity must arise from computations that take place within cortex, because LGN responses are not selective for orientation.

Spatial frequency. V1 neurons are typically sharply selective for the spatial frequency of a stimulus. Spatial frequency is best defined for a grating pattern, where frequency is the inverse of the distance between bars. This selectivity arises naturally from the shape of the receptive fields that have multiple ON and OFF regions.

Direction. Cells in area V1 are commonly selective for direction of stimulus motion.

Temporal frequency. Is the inverse of the period between temporal oscillations between dark and light. V1 neurons typically prefer lower temporal frequencies than those that can drive LGN neurons.

Disparity. In animals with front-facing eyes (such as carnivores and primates), much of the visual field is covered jointly by both eyes. This poses a challenge as signals need to be integrated, but also an opportunity for computing binocular depth (stereoscopy). The signals from corresponding regions in the two eyes are kept separate in the LGN, and are combined in V1.

Color. Retinal ganglion cells respond along one of three "cardinal directions", known informally as red-green, blue-yellow, and black-white. V1 neurons are also organized along "cardinal directions".

3D vision

Psychophysical studies indicate that 3D vision is based on:

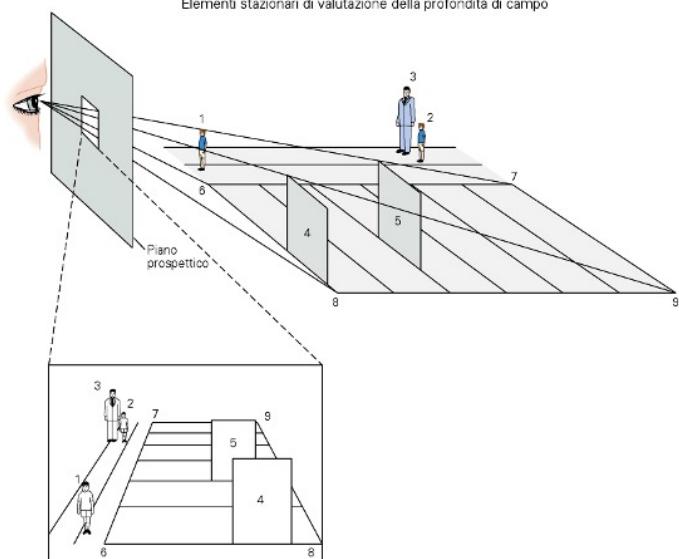
- monocular elements
- stereoscopic elements

Monocular elements are able to create the sense of depth for distances greater than 30 meters:

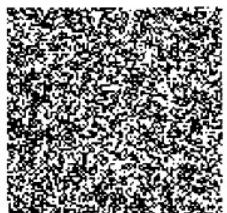
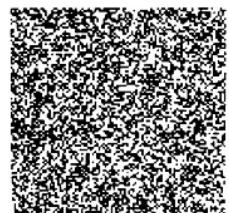
- familiarity with the object
- interposition
- linear perspective
- size of objects
- distribution of shadows and lighting
- parallax movement



Elementi stazionari di valutazione della profondità di campo

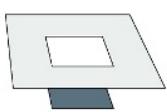
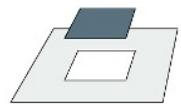


Schema del piano prospettico



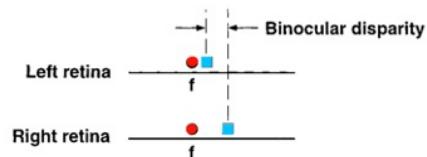
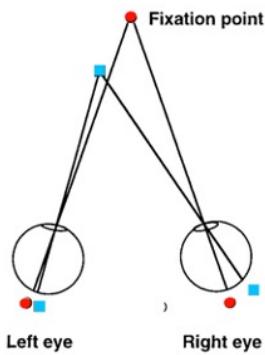
1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	W	A	A	B	B	0	1
1	1	1	Z	B	A	B	A	0	1
0	0	1	Z	A	A	B	A	1	0
1	1	1	W	B	B	A	B	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

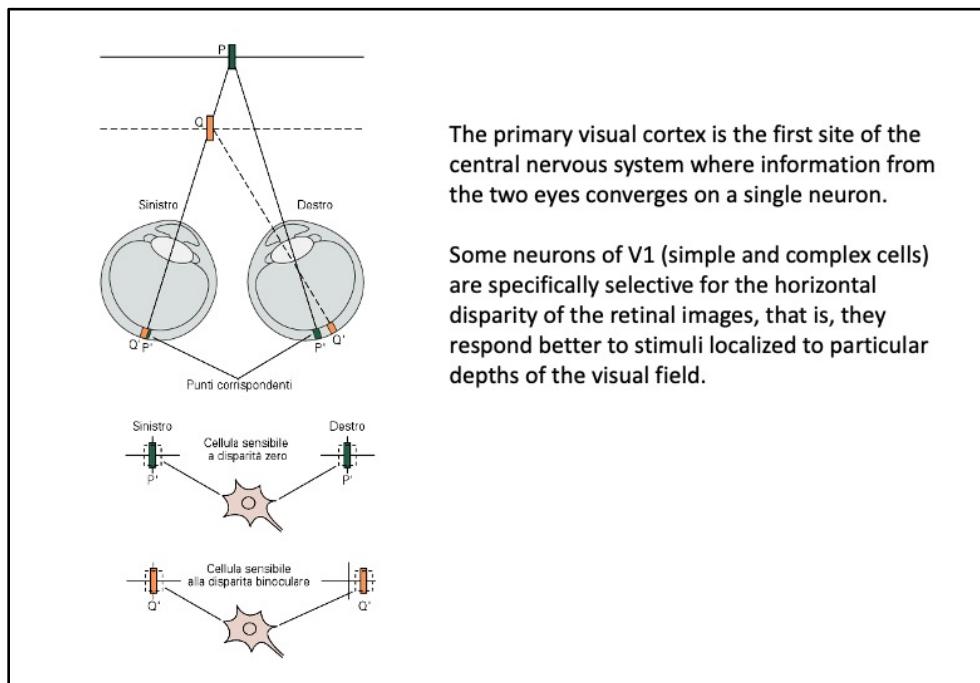
1	0	1	0	1	0	0	1	0	1
1	0	0	1	0	1	0	1	0	0
0	0	1	1	0	1	1	0	1	0
0	1	0	A	A	B	B	Z	0	1
1	1	1	B	A	B	A	W	0	1
0	0	1	A	A	B	A	W	1	0
1	1	1	B	B	A	B	Z	0	1
1	0	0	1	1	0	1	1	0	1
1	1	0	0	1	1	0	1	1	1
0	1	0	0	0	1	1	1	1	0

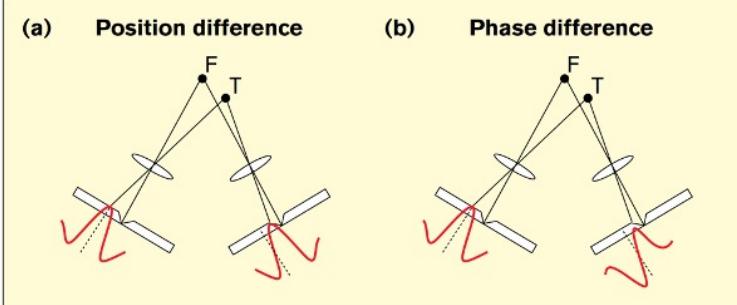


The experiments carried out by Bela Julesz with random dot stereograms show that stereopsis is a visual faculty separate from the perception of forms.

When we fix our eyes on a point, the convergence of the eyes causes the fixed point to fall on identical areas of both retinas. Points lying outside the fixation plane stimulate slightly different parts of each eye's retina creating binocular disparity.







The dotted line marks the equivalent retinal location in the two eyes.

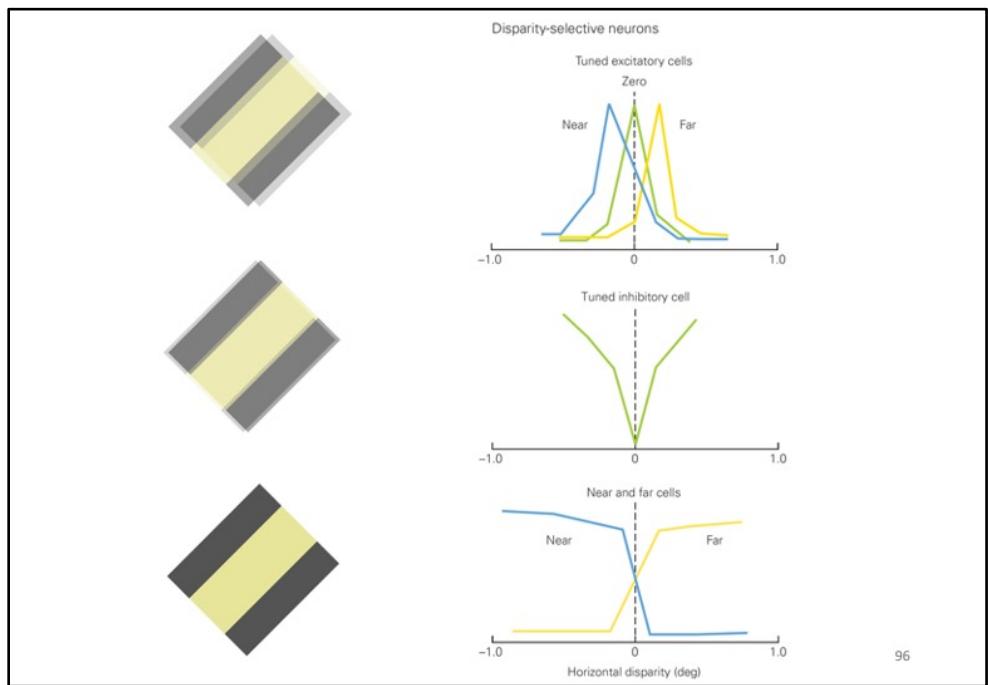
Neurons with receptive fields in both eyes could detect this disparity in two ways.

(a) Position difference: the right eye receptive field is an exact copy of the left eye receptive field, but in a different retinal location.

(b) Phase difference: the envelope enclosing the right receptive field profile sits in the same position as for the left receptive field, but within the envelope, the right receptive field has a different structure, responding best to white light on the right hand side.

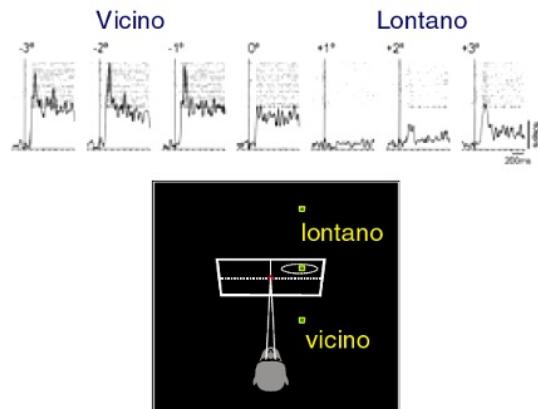
95

The fovea of each eye fixates point F; because object T is closer to the observer than F, the image of T falls at different retinal locations in the two eyes.



In addition to V1, neurons sensitive to retinal disparity are also observed in V2, V3 and particularly in MT and MST.

MST neurons appear sensitive to stimuli that move in particular direction at specific depths of field.



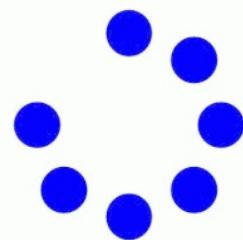
Visual motion

- **Contributes to the recognition of objects (background figure segregation);**
- **It is used to establish the depth or distance of an object (parallax);**
- **Helps in navigation and interaction with the outside world;**
- **It serves to direct attention.**

The movement of visual images is essentially analyzed from the dorsal visual pathway.

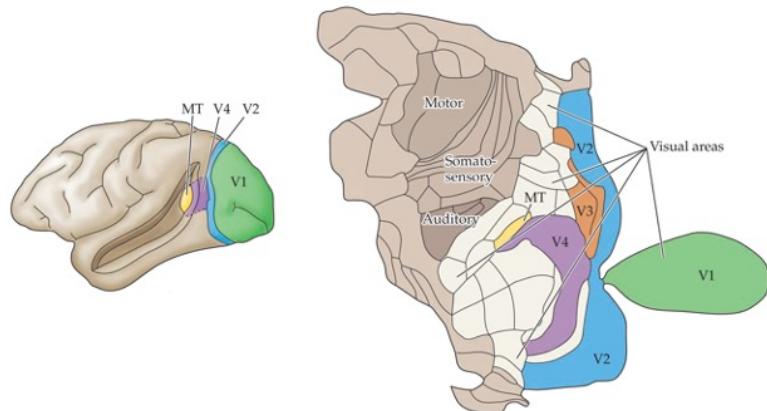
The (illusory) perception of apparent movement constitutes proof of the existence of specific mechanisms for the analysis of visual movement.

The neurons of the IVB layer of V1 are particularly sensitive to the direction of movement of a stimulus within the receptive field;



phi phenomenon

A key role in the perception of movement is played by the MT area, (Middle Temporal) a small region of the dorsal path, located in the posterior part of the superior temporal sulcus, and first described in 1971 independently by J. Kaas and S. Zeki .

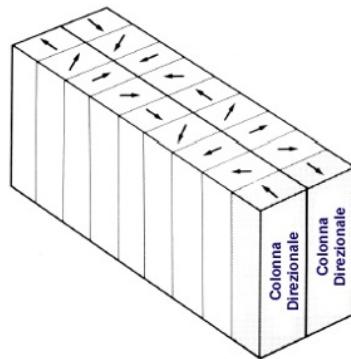


About 95% of MT neurons selectively respond to stimuli that move in a specific direction (preferred direction).

The MT area contains a topographic, retinotopic representation of the visual field

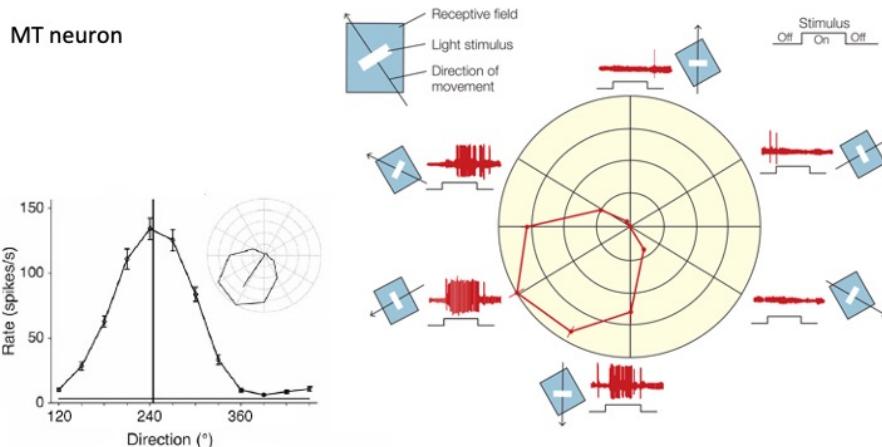
MT shows a columnar cortical organization (analogous to V1):

the neurons present in each column selectively respond to stimuli in the same region of the visual field (same RF) and with the same direction of movement. The preferred direction changes gradually from one column to the contiguous one.



In addition to the direction of movement, MT cells are selective for the speed of the stimulus. Conversely, they do not selectively respond to shape and colors.

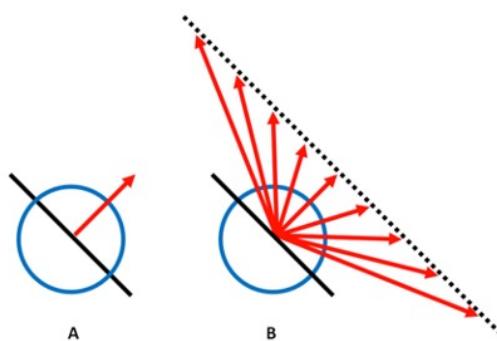
Example of MT neuron responding maximally to downward and leftward directions of movement (225° = preferred direction). Note that the neuron also responds, but less, to directions contiguous to the preferred one.



If a moving contour is seen through a small aperture, only the velocity component perpendicular to the contour can be detected through the opening (Wallach 1935).

The velocity component parallel to the orientation of the contour can only be observed at the extremes.

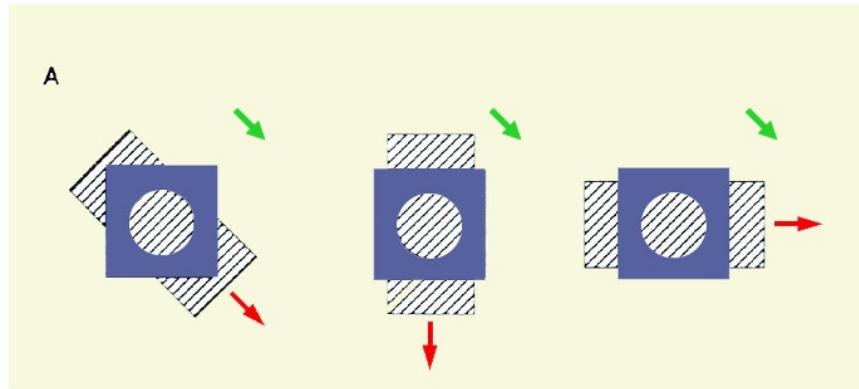
Since neurons in the primary visual cortex (V1) have small receptive fields, this "aperture problem" (Marr, 1982) must be overcome if the visual system is to function properly.



103

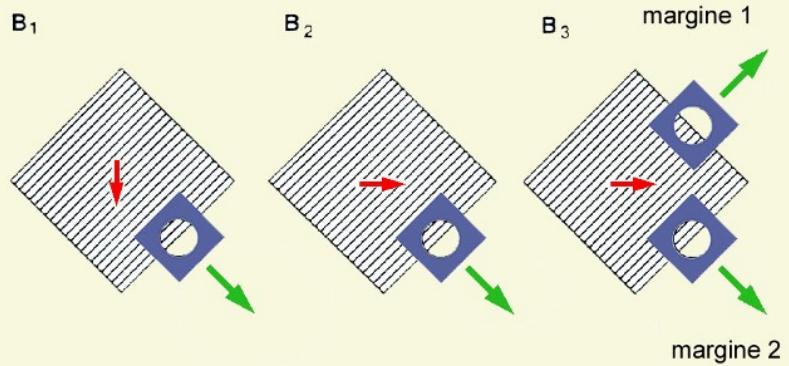
<https://www.coursera.org/lecture/visual-perception/perceived-direction-aperture-effects-ggqjr>

The aperture problem



Due to the aperture problem, movements of a stimulus in three different directions (red arrows) are always interpreted by the neuron as movement in a single direction (green arrows), perpendicular to the orientation of the contour.

The aperture problem



The responses of the neurons in B₁ and B₂ are highly ambiguous, and fail to detect the true direction of the figure's movement.

In B₃, the aperture problem is solved by integrating the response of two or more neurons whose receptive fields are placed in different points of the figure.

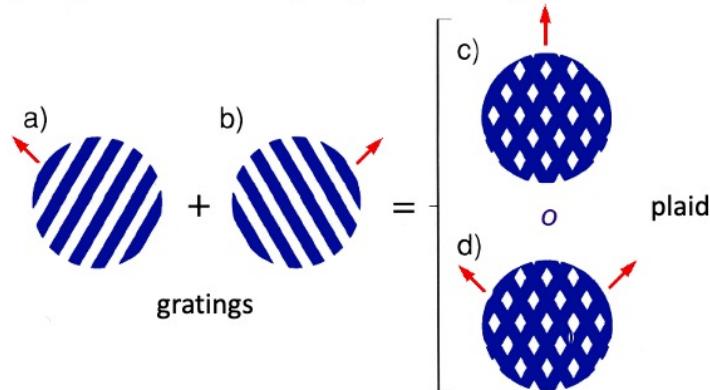
Visual motion processing occurs in two stages

To solve the aperture problem, it has been hypothesized by some researchers that the visual processing of moving stimuli occurs through two successive cortical stages:

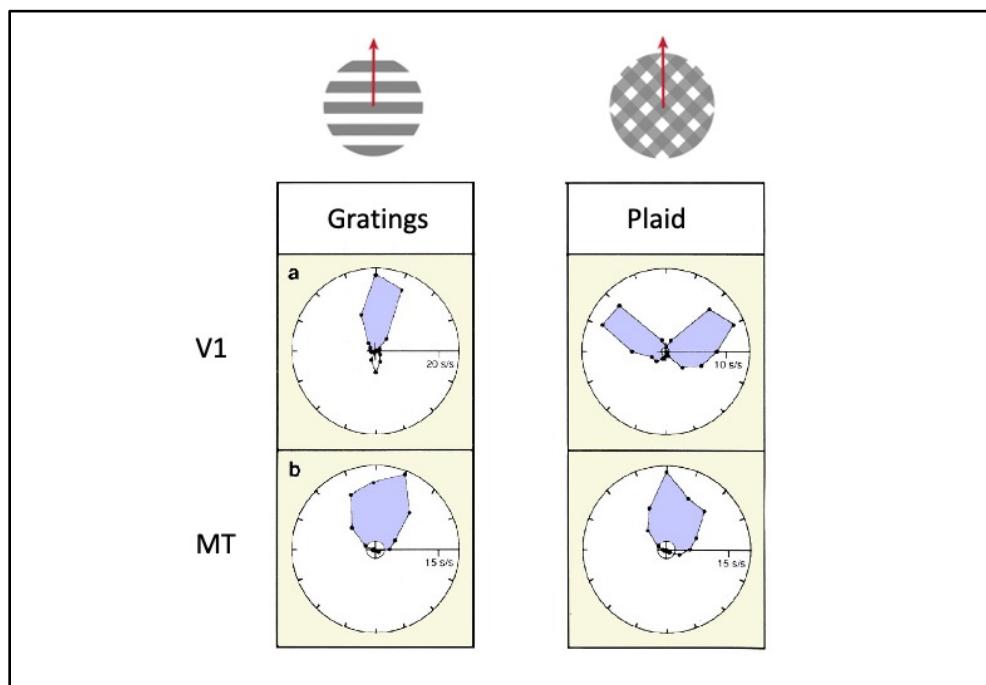
- 1) In the first stage, the individual components of the stimuli are analyzed, through the responses of neurons capable of signaling only the movement of the local components, perpendicular to the movement of the contour.
- 2) In the second stage, higher order neurons would integrate the different local components of the stimulus, analyzed by the neurons of the previous stage.

The integrated signal produced by higher-order neurons corresponds to the real direction of motion of the object and to the perception that an observer has of it.

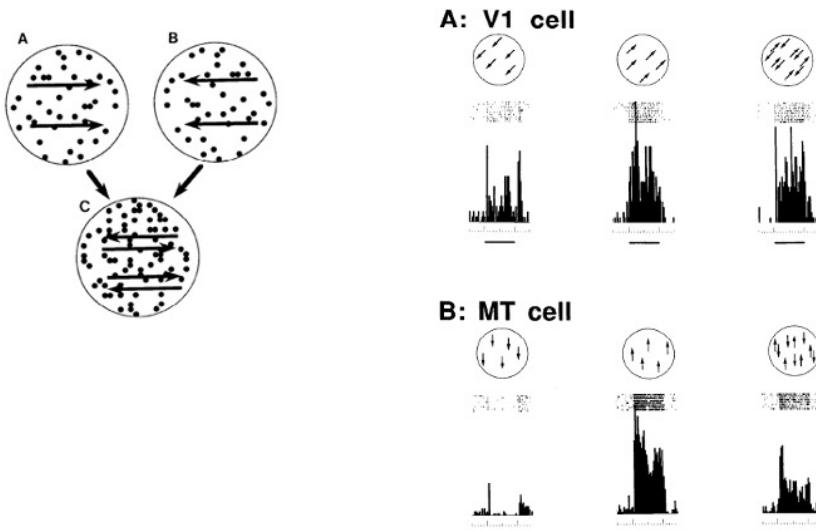
This hypothesis was evaluated by Movshon and colleagues (1982) in the MT area of a monkey, through the use of complex stimuli ('PLAID') obtained from the superimposition of sinusoidal gratings directed in opposite directions.



By superimposing two gratings (a and b) that move in the opposite direction (top-right and top-left) we obtain a gratings or PLAID that appears to move clearly in only one direction (top, resulting from the sum of the directions of the two gratings, as in c).



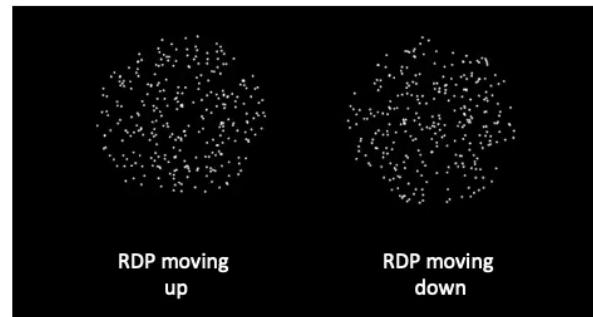
The response of area MT and V1 neurons to transparent motion



109

Selective visual attention can be directed toward a specific region of space (i.e., space-based attention),

Middle temporal (MT) area is crucial for the perception of visual motion.
MT neurons show direction tuning curves (bell-shaped response profiles) depending on the direction of motion of the stimulus (random dot pattern, RDP).



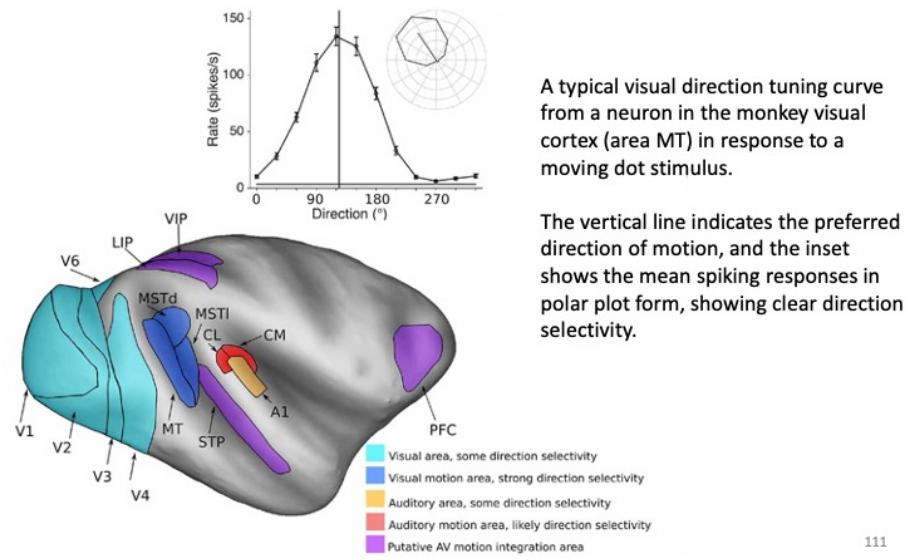
RDP moving
up

RDP moving
down

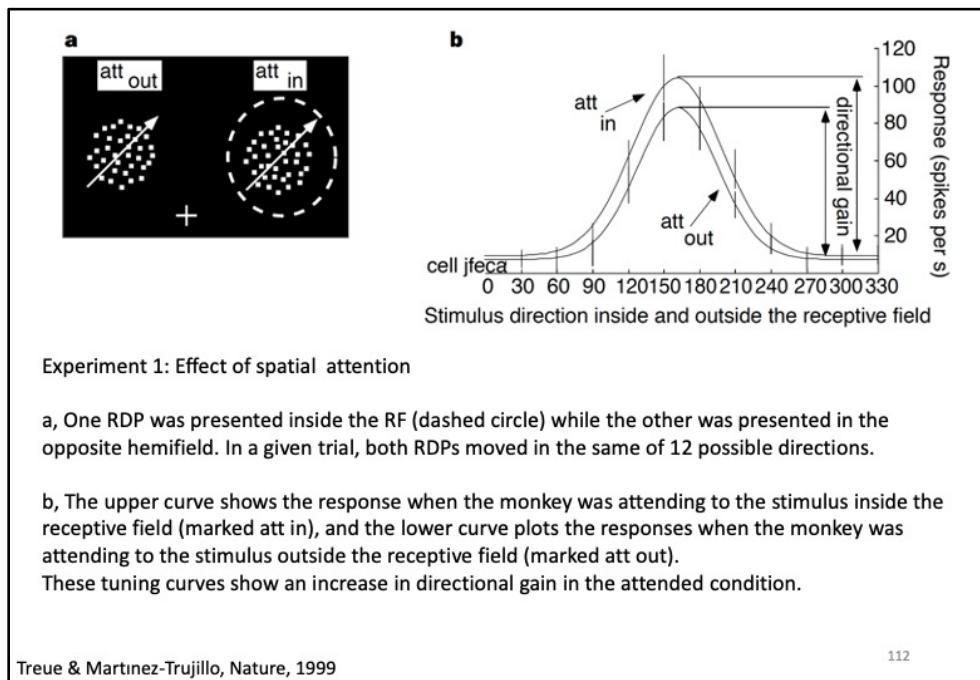
Treue & Martinez-Trujillo, Nature, 1999

110

Spatial attention modulates the responses of neurons selective to direction of motion in area MT and MST.



111



Attending to a given direction enhances the responses of neurons whose preferred direction aligns with the attended direction and reduces the responses of those neurons preferring the opposite direction.

These findings shows that spatial and feature-based attention represent separate (and summable) processes that have a multiplicative effect on the responses of neurons.

Treue & Martínez-Trujillo, Nature, 1999

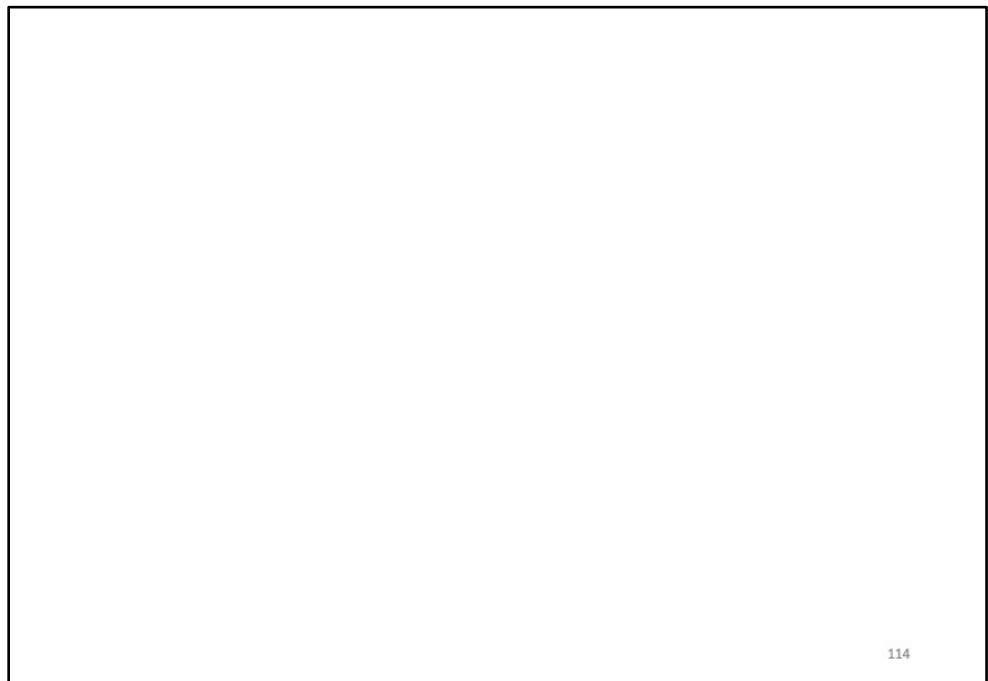
113

These results demonstrate a physiological correlate of non-spatial, feature-based attention by showing neuronal response modulations in the absence of spatial shifts of attention.

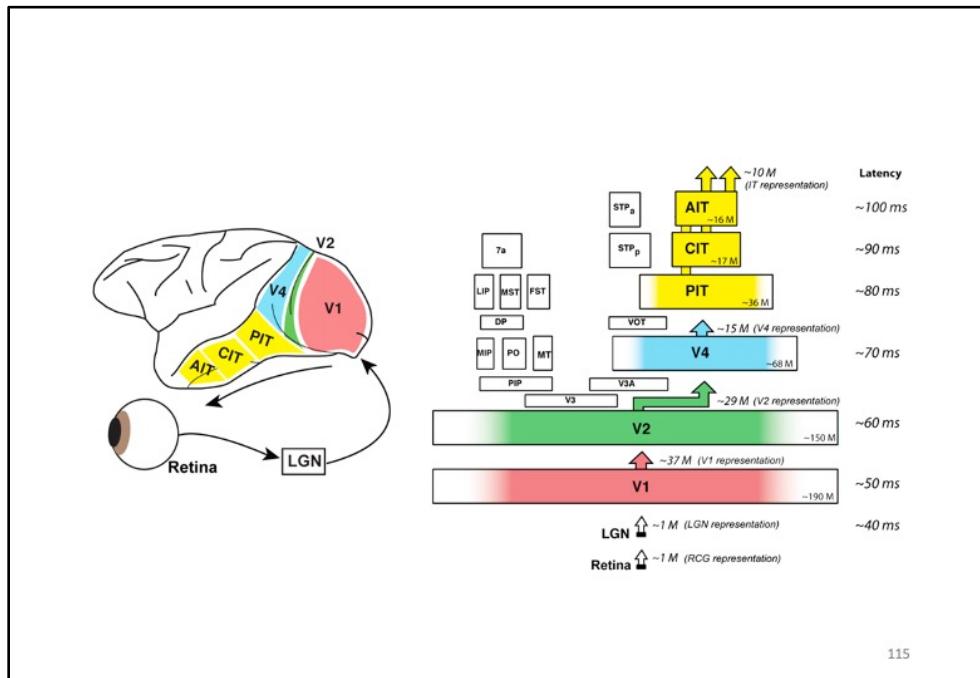
Thus, attending to a given direction enhances the responses of neurons whose preferred direction aligns with the attended direction and reduces the responses of those neurons preferring the opposite direction.

These findings shows that spatial and feature-based attention represent separate (and summable) processes that have a multiplicative effect on the responses of neurons.

Such attentional modulations resemble changes to a neuron's sensory gain and thus can be mimicked by sensory effects, such as reducing the luminance contrast of a stimulus, which similarly does not change the tuning width of direction-selective neurons, suggesting that response modulation based on attentional and sensory aspects employ common mechanisms.



114



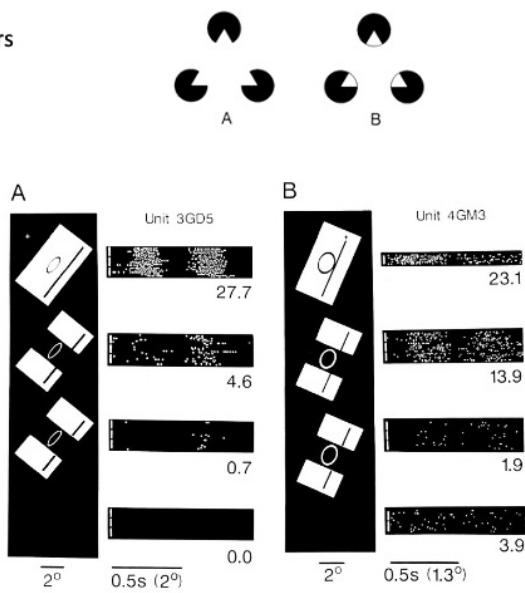
115

Neural responses to illusory contours

Figures in which humans perceive illusory contours evoke responses in cells of area 18 (V2) in the monkey visual cortex.

Modifications that weakens the perception of contours also reduces the neuronal responses.

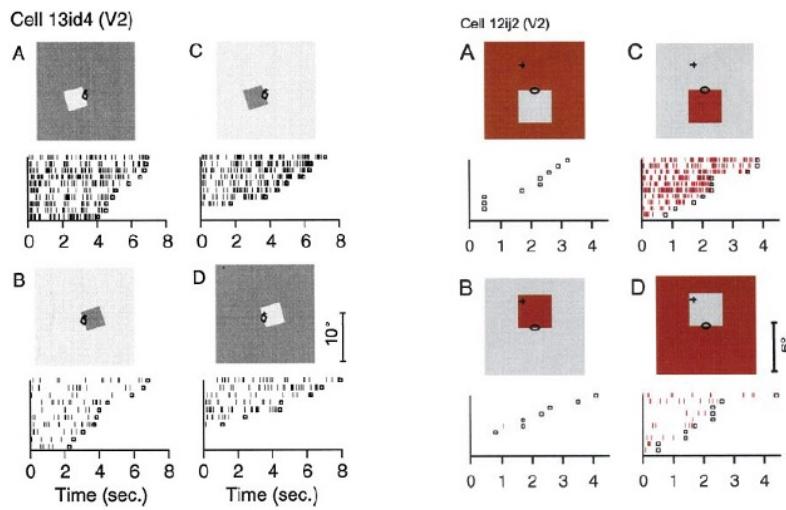
In contrast, cells in V1 (area 17) were apparently unable to see these contours.



Von der Heydt et al., Science, 1984

116

Coding of border ownership in monkey area V2



Zhou et al., J Neurosci, 2000

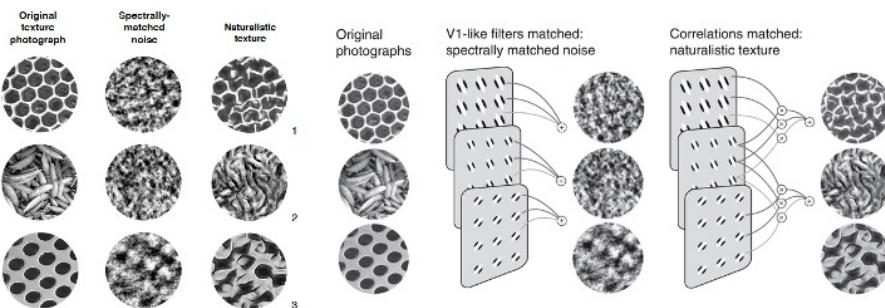
117

The influence of visual stimulation far from the receptive field center indicates mechanisms of global context integration.

The short latencies and incomplete cue invariance suggest that the border-ownership effect is generated within the visual cortex rather than projected down from higher levels.

There is no generally accepted account of the function of V2, partly because no simple response properties robustly distinguish V2 neurons from those in V1.

Stimuli replicating the higher-order statistical dependencies found in natural texture images were used to stimulate macaque V1 and V2 neurons.



Freedman et al., Nat. Neurosci., 2013

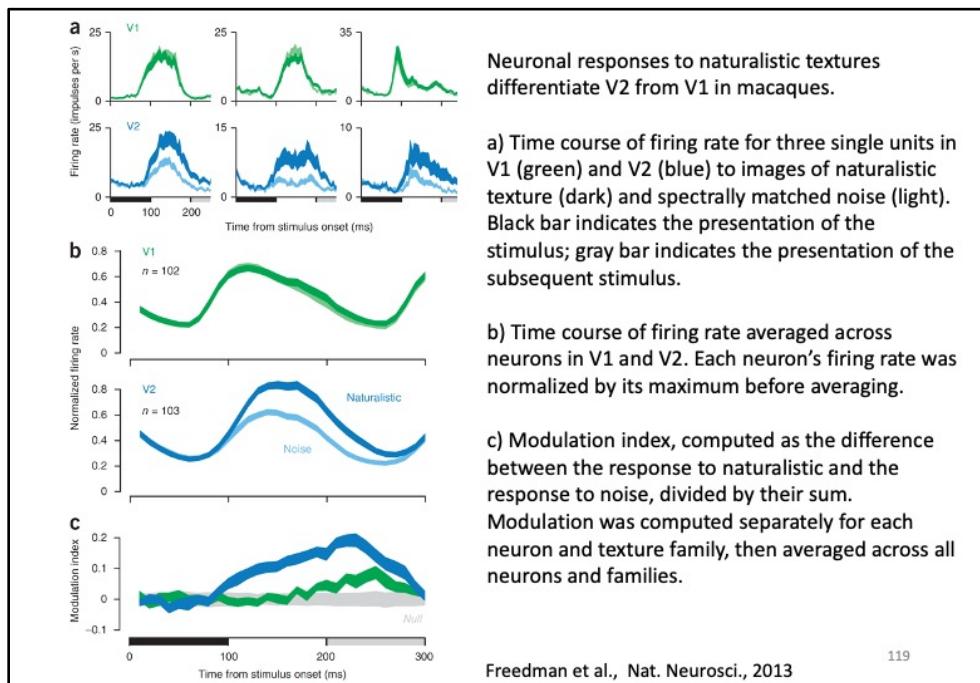
118

Spectrally matched noise images.

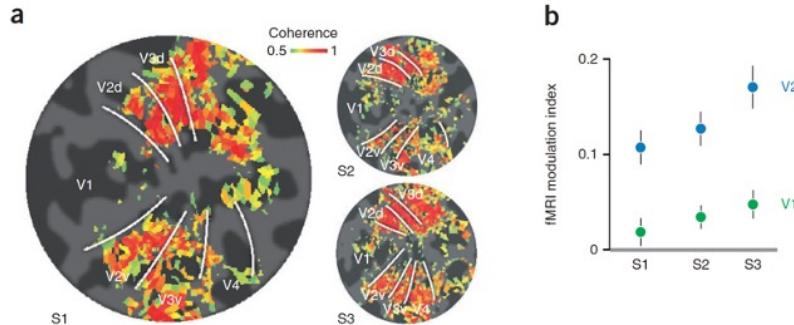
The original texture is analyzed with linear filters and energy filters (akin to V1 simple and complex cells, respectively) tuned to different orientations, spatial frequencies and spatial positions.

Noise images contain the same spatially averaged orientation and frequency structure as the original but lack many of the more complex features.

Naturalistic texture images. Correlations are computed by taking products of linear and energy filter responses across different orientations, spatial frequencies and positions. Images are synthesized to match both the spatially averaged filter responses and the spatially averaged correlations between filter responses. The resulting texture images contain many more of the naturalistic features of the original.



Naturalistic texture stimuli modulate the responses of neurons in area V2, while having only a minimal effect on neurons in area V1.
These modulations were similar and substantial in both macaques and humans.



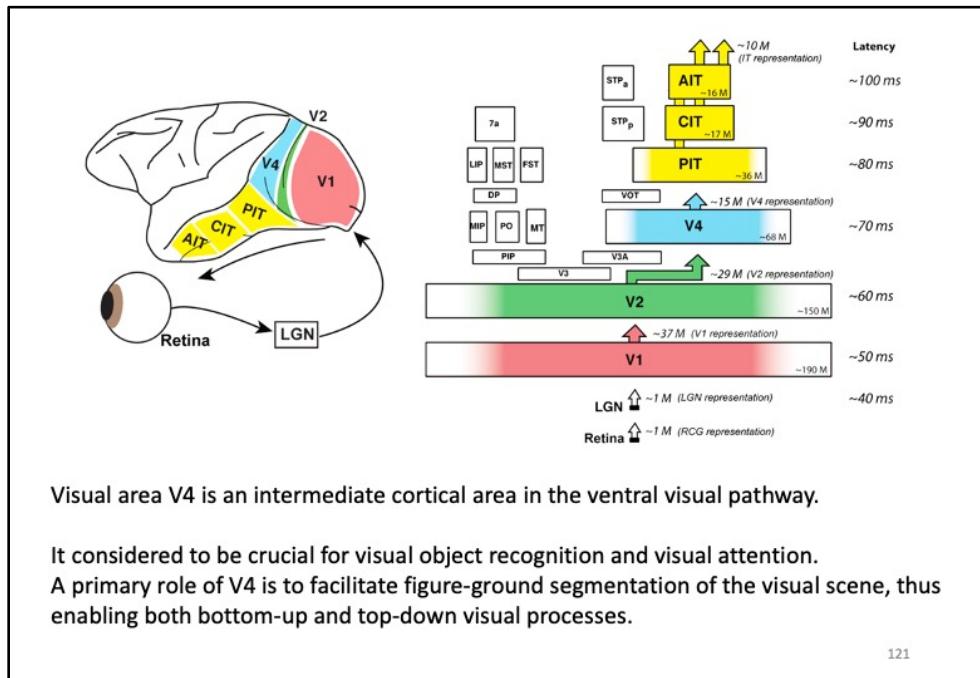
fMRI responses to naturalistic textures differentiate V2 from V1 in humans.

Freedman et al., Nat. Neurosci., 2013

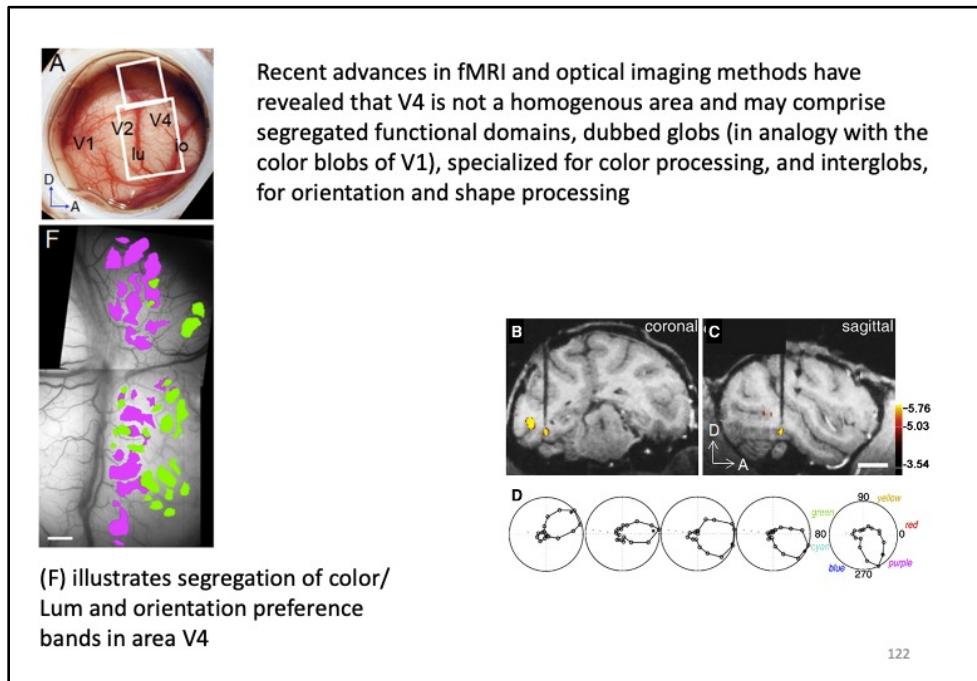
120

a) Responses in subjects 1–3 (S1–S3) to alternating blocks of naturalistic texture images and spectrally matched noise shown on a flattened representation of the occipital pole. Color indicates coherence, which captures the extent to which the fMRI responses to naturalistic and noise stimuli differ, computed voxel by voxel after averaging responses to all texture families. White lines indicate boundaries between visual areas, identified in an independent retinotopic mapping experiment.

b) fMRI modulation averaged across voxels and texture families in V1 and V2 for the three subjects. Error bars indicate s.e.m. across texture families.



In the macaque monkey, V4 is located on the prelunate gyrus and in the depths of the lunate and superior temporal sulci and extends to the surface of the temporal-occipital gyrus.

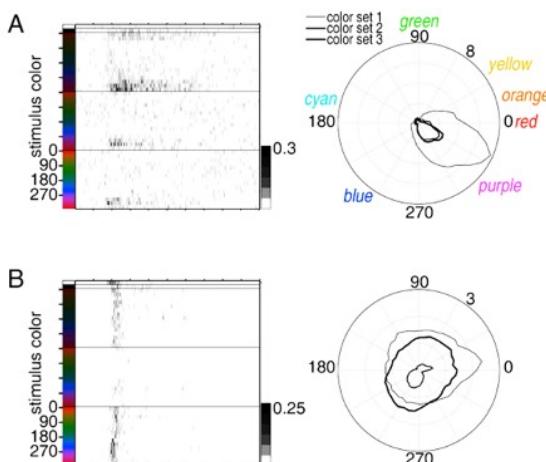


Several color globs identified with fMRI seen in coronal (B) and sagittal (C) views.

Color voxels are better activated by equiluminant color than black and white gratings.

Electrode is seen targeting a color glob. (D) Shift in color preference of neurons along length of an electrode penetration through a glob.

Figure shows color-response measurements for a typical cell in a glob (A), and a cell in the interglob (B) region. Left panels show time histograms to an optimally shaped bar of various colors; right panels show the color tuning in polar coordinates.



Most glob cells were excited by a specific hue, as shown by hue tuning in the polar plots.

This hue selectivity was luminance invariant. That is, hue tuning does not change with change in luminance between stimulus and background.

Conway et al., Neuron, 2007

123

These measurements were conducted by presenting an optimally configured bar at the center of the response region for each cell, and in each trial changing the color of the bar.

The top two rows in each histogram show the responses to white and black.

The rest of the histogram plot is divided into three sections.

The top section shows responses to a set of colors that were equiluminant with each other but lower luminance than the background (top), equiluminant with background (middle), and of higher luminance than the background (bottom).

The glob cell shown in figure was excited by bluish-red, shown by the maximal response density at the bottom of each section of the histogram.

The hue responses can be compressed into polar plots. Despite differences in overall response magnitude to the three color sets, the peak hue response within each color set was the same: each of the three curves in right panels point to 330.

This shows that the hue selectivity was luminance invariant.

Luminance invariance does not imply that luminance does not modulate the responses, just that a change in luminance does not drastically shift or cancel out the hue tuning.

Luminance invariance can be quantified by determining the degree to which the patterns of responses to the different color sets are correlated.

The correlation coefficient of the response to color set 1 and color set 2 for was 0.94; between color set 1 and 3, 0.87; and between color set 2 and 3, 0.96; average, 0.92.

The interglob cell (B), on the other hand, did not show strong hue tuning, to any color set;

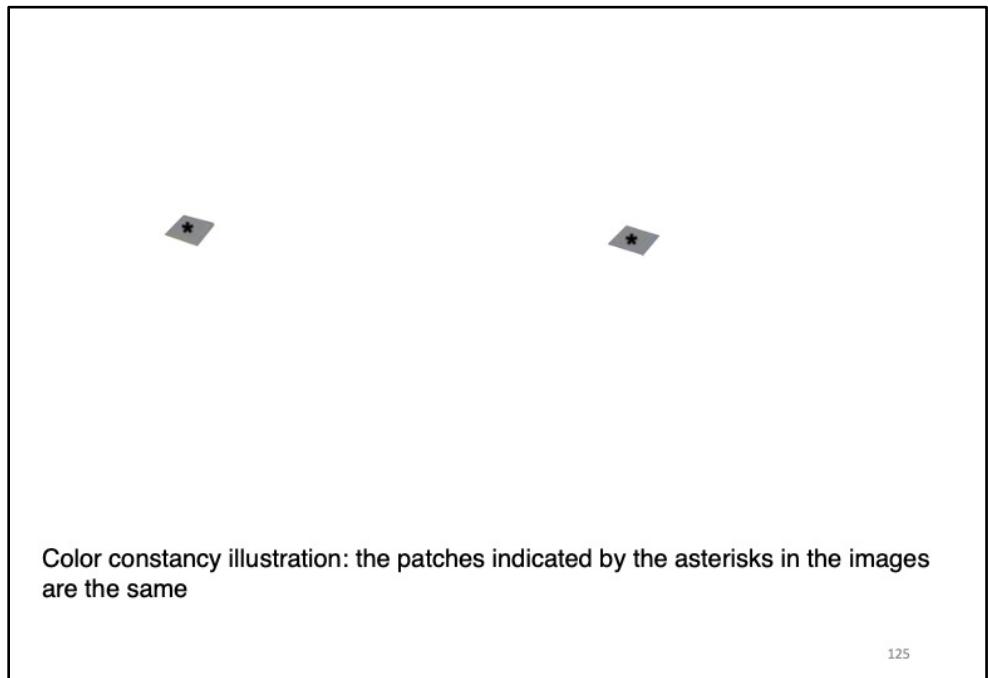
Color constancy

It refers to the effect whereby the perceived or apparent color of a surface remains constant despite changes in the intensity and spectral composition of the illumination.

The following shows a striking example in which the same physical patch appears yellow when the scene appears to be illuminated by blue light but appears blue when the scene appears to be illuminated by yellow light.

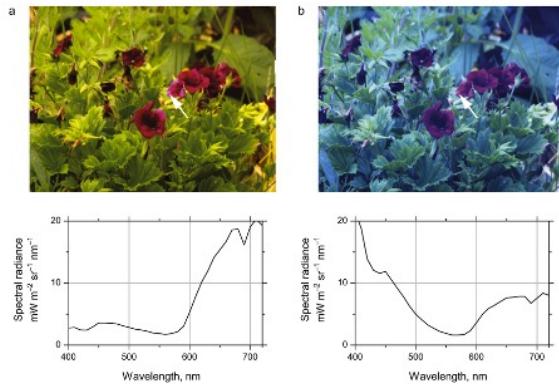
Surrounded by white, the patch appears gray.

This example illustrates the potency of color contrast in generating color perception.



Color constancy illustration: the patches indicated by the asterisks in the images are the same

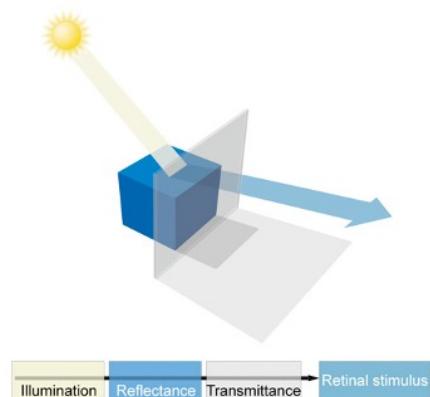
125



Images of a pelargonium under sunlight and skylight with respective correlated color temperatures (a) 4000 K and (b) 25,000 K and the corresponding radiance spectra reflected from the arrowed region of a petal.

Constancy is achieved because the brain determines an object's color from a local comparison of the light reflected from the object with the light reflected from adjacent regions of the scene.

Retinal luminance in visual stimuli (the inverse optics problem)



Retinal luminance, the spectral light reflected from an object is a product of two variables, Illumination (the spectral content of the illuminant, i.e., the light source) and reflectance (the absorptive properties of the object, as well as a variety of other factors (e.g., atmospheric transmittance).

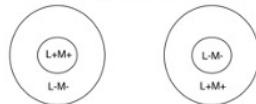
These physical determinants of retinal luminance values are conflated in visual stimuli and cannot be disentangled by any algorithmic process.

Retina and LGN: Color Opponency

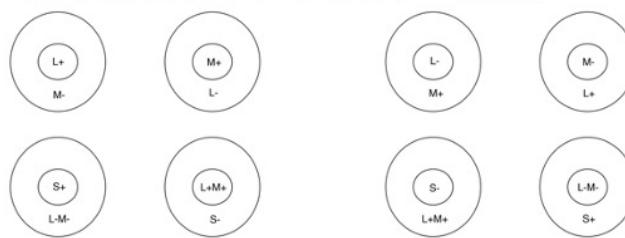
The principle dimensions in the space of cone excitations produced by natural objects are:

- 1) a luminance axis where the L- and M-cone signals are added;
- 2) a red-green opponent axis where the difference of L- and M-cone signals is taken;
- 3) a blue-yellow color opponent axis where the S-cone signal is differenced with the sum of the L- and M-cone signals;

M-cells Magnocellular (luminance cells)



P-cells Parvocellular (color-luminance cells)



Receptive fields of single-opponent cells in the retina and LGN are not capable of color contrast.

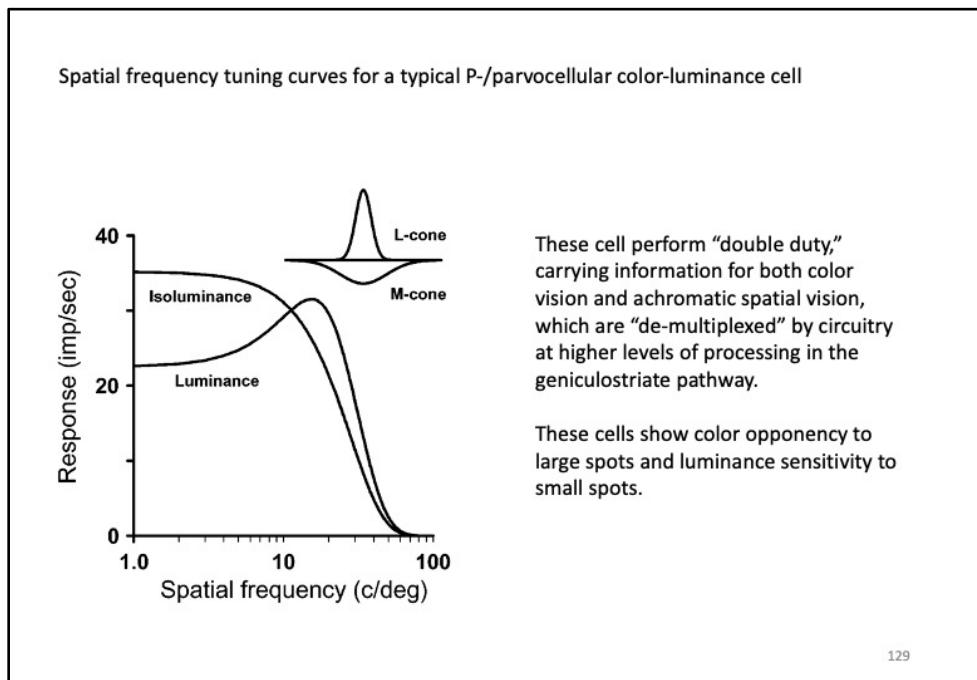
128

Psychophysical experiments have revealed that there are three independent channels in color vision, or cardinal directions of color space. (Krauskopf et al., 1982)

'Single-opponent' type receptive field.

'L+' denotes excitatory L cone input; 'M-' denotes inhibitory M cone input.

The cell thus prefers 'red' in the centre — maximally stimulating the L cone — and a lack of 'green' in the surround.



The inset in the top right corner shows a schematic representation of the receptive field's sensitivity profile.

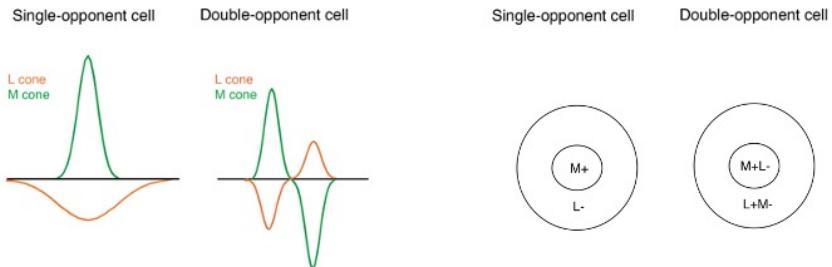
The cell receives excitatory inputs from L-cones in the center and inhibitory inputs from M-cones in the whole receptive field.

The sign difference between the cone inputs gives this neuron its color-opponency.

The difference in spatial extent between the inputs produces its spatial opponency.

Thus, these cells show both low-pass color tuning and high-pass luminance tuning, suggesting they are capable of subserving both color vision, which has relatively low spatial resolution, and form vision, which has relatively high spatial resolution and is largely determined by luminance edges.

Primary visual cortex: calculating color contrast



Single-opponent cells cannot resolve color contrast.

Double-opponent cells may represent the neural basis for color contrast and color constancy.

The optimal stimulus for a Green-on double-opponent cell is a green spot on a red background. Because of their specialized receptive-field structure, double-opponent cells are candidates for the neural basis for color contrast and color constancy.

130

Single-opponent

These cells increase their firing rate above some baseline rate in response to activation of one cone class and decrease their firing rate when a different cone class is activated. Thus, they cannot resolve color contrast.

Double-opponent

These cells are so named because their receptive fields are both chromatically and spatially opponent

Object identification and categorization

The visual experience of the world is fundamentally centered on objects.

By visual object we mean a set of visual characteristics (e.g., visual features) grouped or joined perceptually in discrete units on the basis of the organizational principles of the Gestalt, such as proximity, similarity, closure, good continuation, good form, connection, etc.

By visual recognition we mean the ability to assign a verbal label (e.g., a name) to objects in the visual scene.

There are at least two possible object recognition tasks, distinguished by level of specificity: identification and categorization.

An object can be recognized at an individual level (e.g., a Siamese cat), or at a more general categorical level, as an object belonging to a given class (a cat, a mammal, an animal, and so on).

131

It is quite simple for computer vision techniques to identify (rather than categorize) objects.

On the contrary, for the human vision the task of identification (compared to categorization) is more difficult.

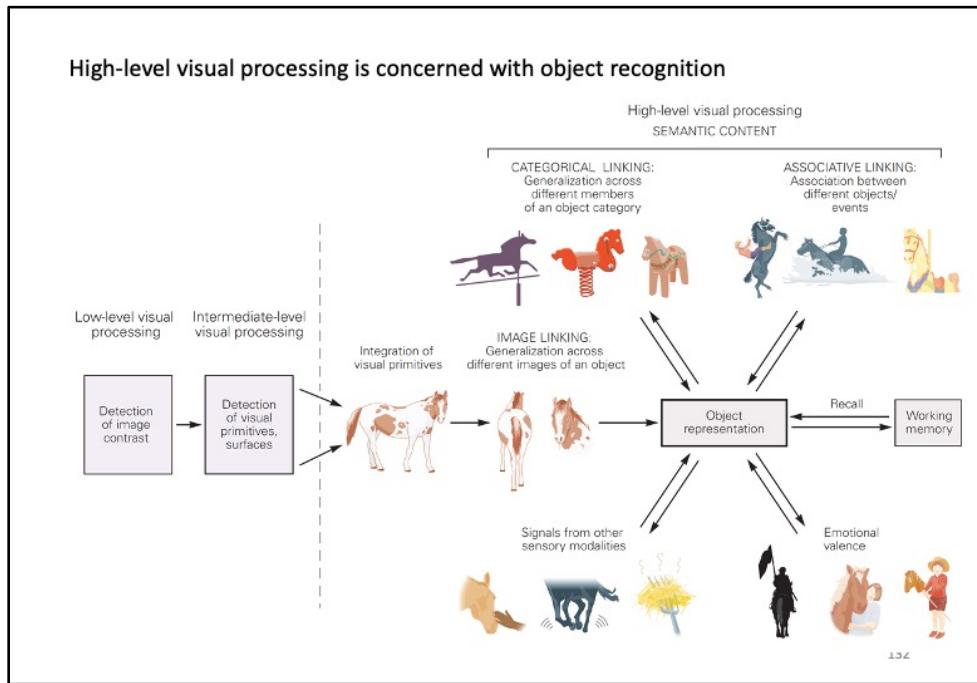
Categorization

A category exists whenever two or more distinct objects or events are treated equivalently.

For example, when distinct objects or events are labeled with the same name, or when the same action is performed on different objects.

Although the stimuli are distinct, organisms do not treat them uniquely; but they respond on the basis of past experience and categorization.

In this sense, categorization can be considered one of the most basic functions of living beings (Mervis and Rosch, 1981)



We effortlessly and rapidly (100-200ms) detect and classify objects from among tens of thousands of possibilities despite the tremendous variation in appearance that each object produces on our eyes.

Our daily activities (e.g., finding food, social interaction, selecting tools, reading, etc.), and thus our survival, depend on our accurate and rapid extraction of object identity from the patterns of photons on our retinae.

The fact that half of the nonhuman primate neocortex is devoted to visual processing speaks to the computational complexity of object recognition.

Object recognition involves integration of visual features extracted at earlier stages in the visual pathways.

This integration requires generalization across different retinal images of an object, as well as generalization across different members of an object category.

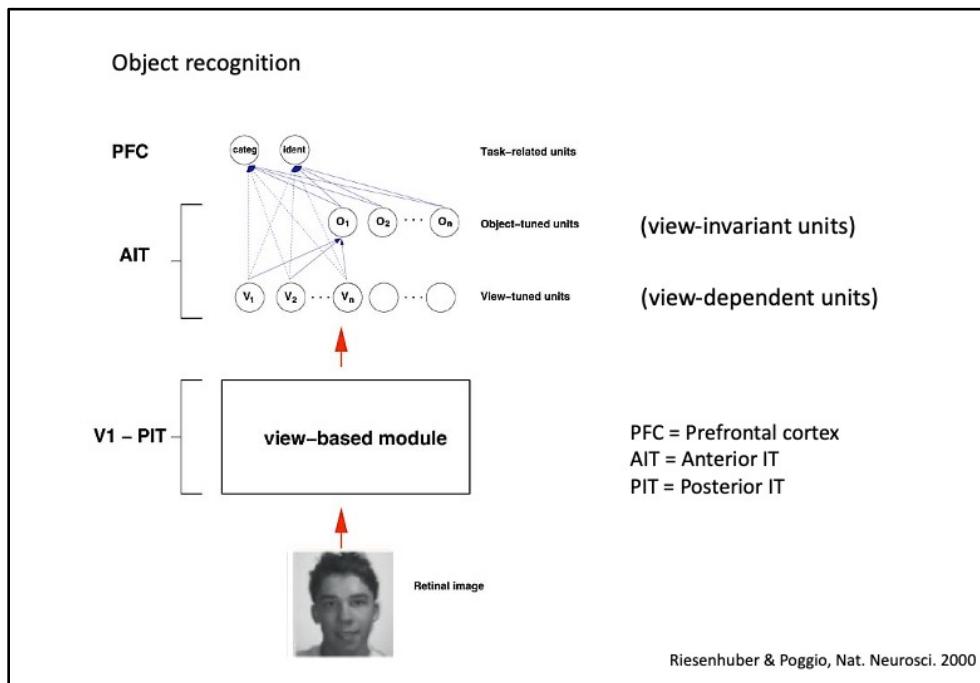
The representation also incorporates information from other sensory modalities, attaches emotional valence, and associates the object with the memory of other objects or events.

Selectivity and object constancy (or invariance)

A computational difficulty of object recognition is that it requires both:

selectivity (different responses to distinct objects, such as one face with respect to another face);
and invariance with respect to image transformations (similar responses to, for example, rotations or translations of the same face);

In fact, we are able to recognize the same object even when the image it projects on the retina varies considerably.

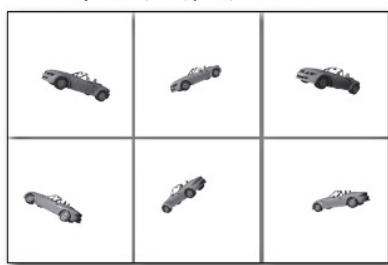


The phenomenon of perceptual constancy, or "object invariance problem", is the crucial point of object recognition, and also the main obstacle for computer vision recognition systems.

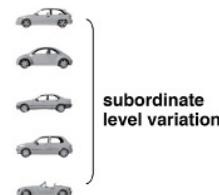


"Identity preserving image variation"

View: position, size, pose, illumination



Clutter, occlusion



135

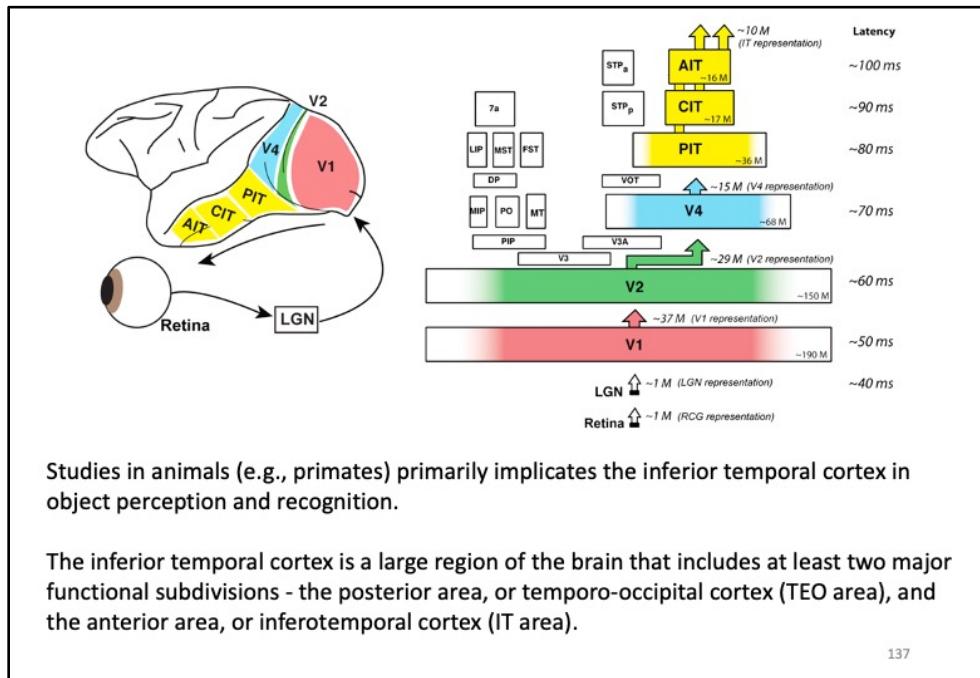
"car":

Not "car":

Core object recognition is the ability to rapidly (<200 ms viewing duration) discriminate a given visual object (e.g., a car, top row) from all other possible visual objects (e.g., bottom row).

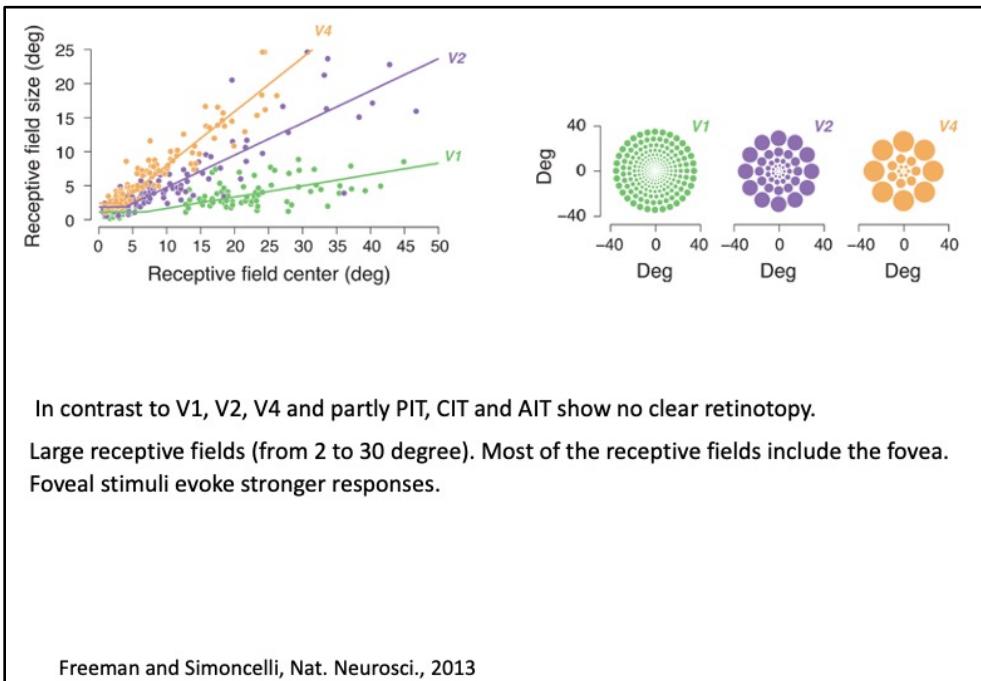
Primates perform this task remarkably well, even in the face of identity-preserving transformations (e.g., changes in object position, size, viewpoint, and visual context).

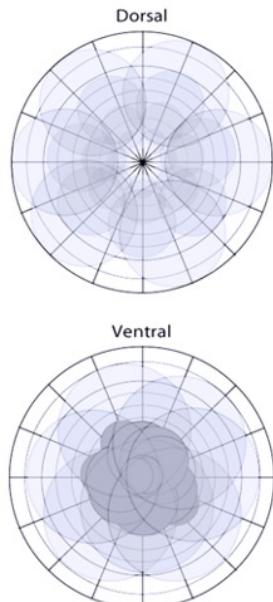
136



Area V1,V2 and V4 are located in the occipital lobe;

Area TEO (TEmporal-Occipital) and IT (InferoTemporal) are located in the temporal lobe;





Representation of the visual field in the dorsal and ventral visual pathway

The dorsal visual pathway is characterized by a more complete sampling of the visual field with wide receptive fields, which include the visual periphery, and, in some cases, sparing of foveal regions.

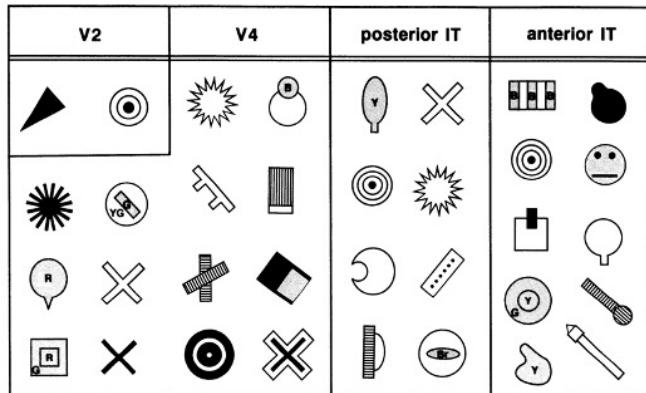
In contrast, the areas of the ventral visual pathway show greater focalization, which always involve coverage of the foveal regions.

This difference would indicate a crucial role of the dorsal pathway in the exploration of salient stimuli of the visual field, while the ventral pathway would be involved in object recognition and use (exploration vs. exploitation).

Sheth and Young, *Front. Integrat. Neurosci.*, 2016

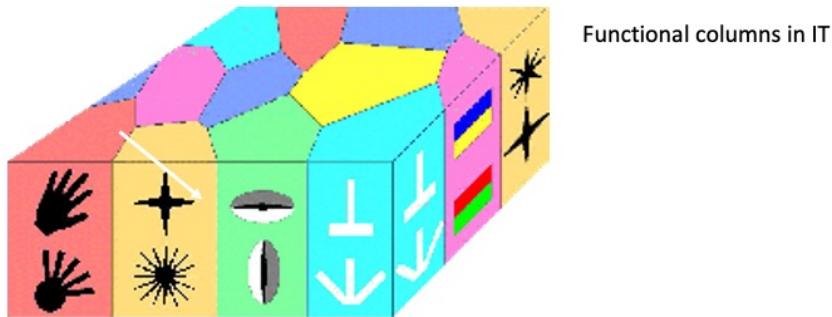
139

Increased complexity of effective stimuli along the ventral visual path



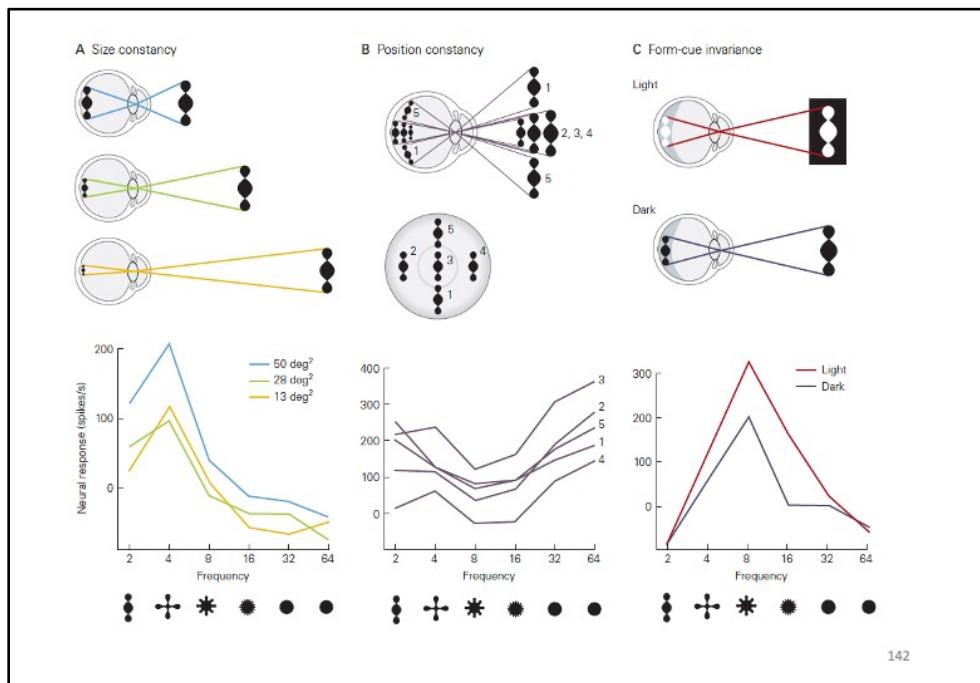
Kobatake and Tanaka, J. Neurophysiol., 1994

Tanaka (1996) suggested that responses to effective stimuli are organized in topographical order, so that two adjacent neurons are much more likely to respond similarly to a set of stimuli than neurons that are more distant.

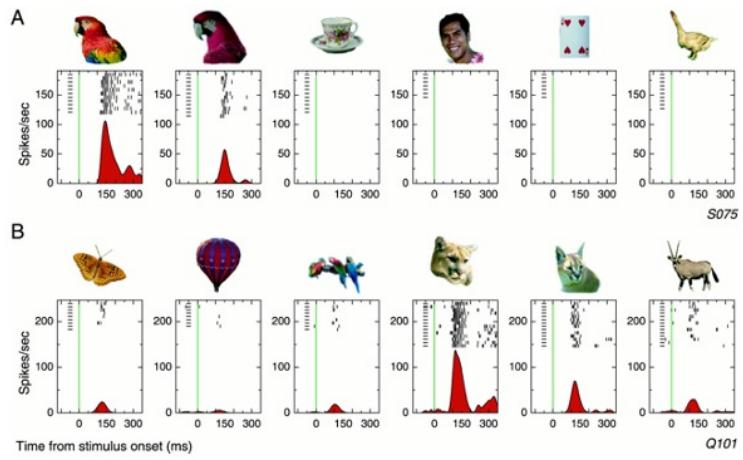


The neurons that respond to the different parts of an object are not randomly arranged in the IT area.

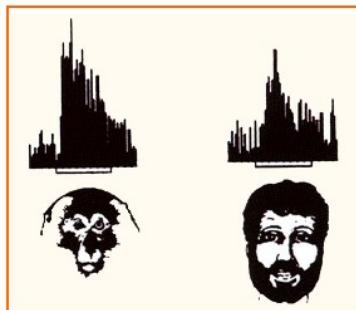
Neurons that respond to same elements are grouped in a cortical column, analogous to the orientation columns found in V1. Contiguous columns encode elements related to each other.



Neurons in the IT respond to relatively complex stimuli, often to biologically relevant objects such as human and other animal faces, hands and other parts of the body.

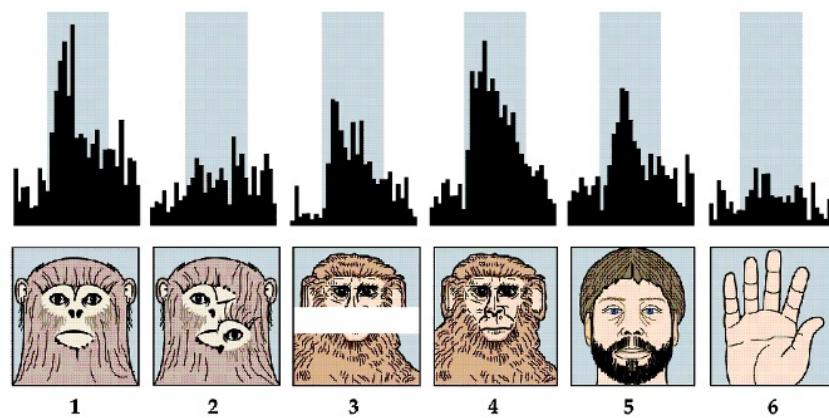


In the early 1980s, several researchers (Bruce et al., Perrett et al.) Identified in the monkey a group of IT neurons that responded selectively to faces.



Question: Are there in IT, as for faces, selective cells for the different types of objects that can be encountered in the outside world (neurons for chairs, for flowers, for cars, etc ...)?

Neuron that responds to faces: The neuron responds to faces of different species (1, 4, 5). The discharge is reduced if the elements of the face are mixed (2) or occluded (3). The neuron does not respond to other biologically relevant stimuli (5).



Desimone, Albright, Gross and Bruce, J. Neurosci, 1984

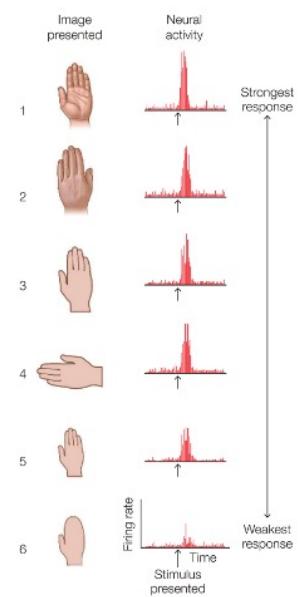
Recording of a single neuron from monkey IT cortex (Desimone et al., 1984)

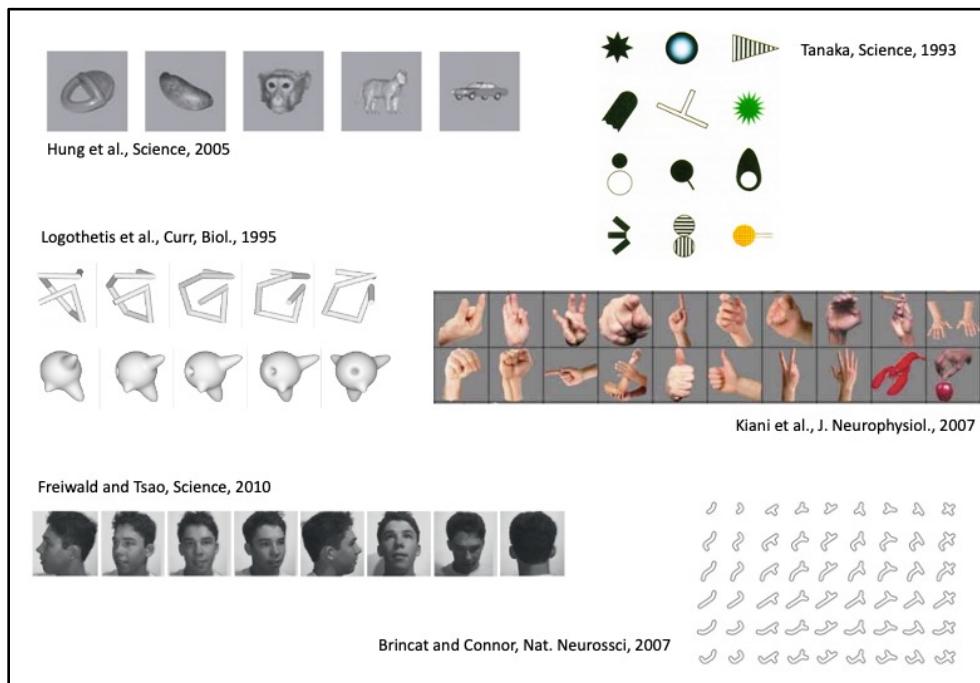
This cell is activated by the vision of the human hand.

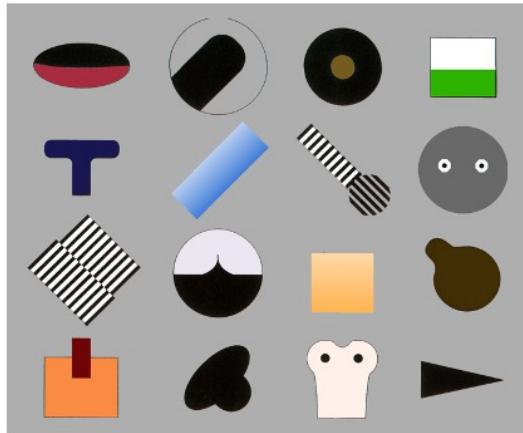
The first five images in the figure show the cell's response to various perspectives of a hand.

Activity is high regardless of hand orientation and only decreases slightly when the hand is noticeably smaller.

The sixth image shows that the response decreases if the stimulus has the same shape, but does not have well-defined fingers.





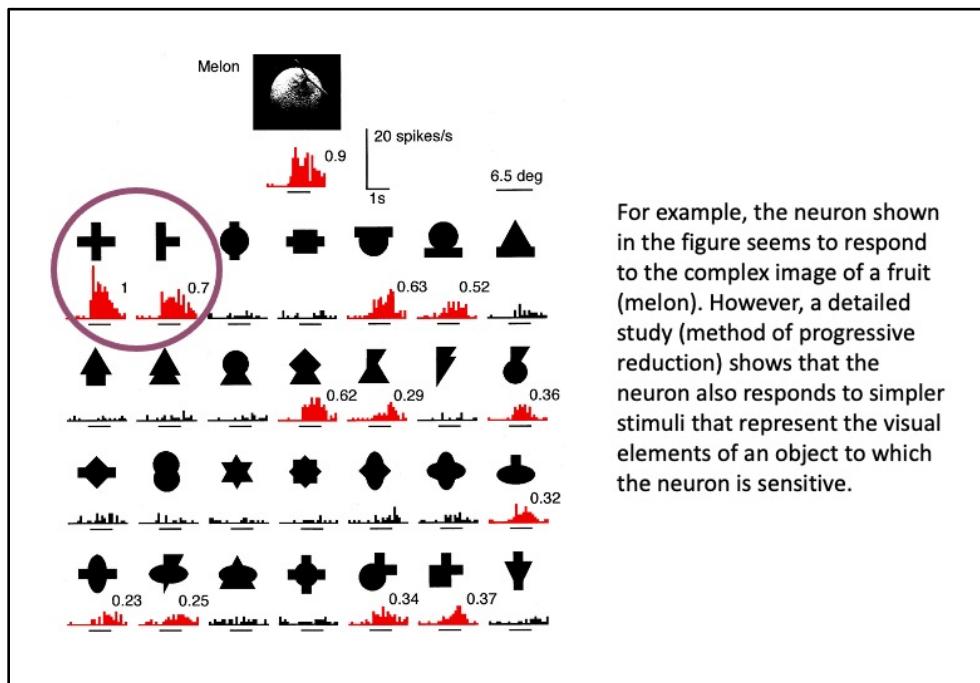


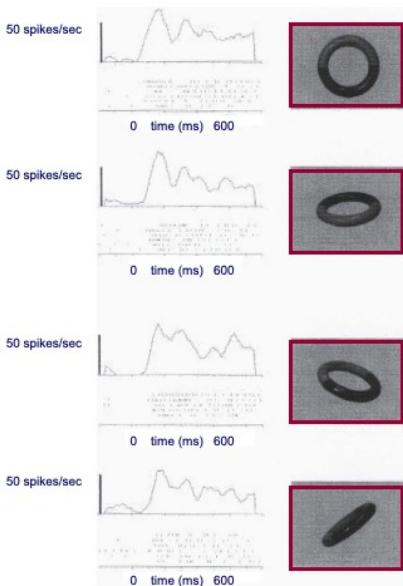
Tanaka hypothesized that the representation of each object occurs not through the activation of a single neuron but of a population of IT neurons (distributed representation), each of which encodes only one element or a particular (which can be present in several objects) and not the whole object.

Schematic representation of the different elements to which IT neurons respond (from Tanaka, 1996).

Studies by Tanaka have shown that IT neurons that seem selective for a specific object, actually respond to a part or component of the object, and not to the object as a whole.

These parts of objects that activate neurons can be found in a variety of different stimuli, and constitute a sort of visual alphabet for the recognition of objects.



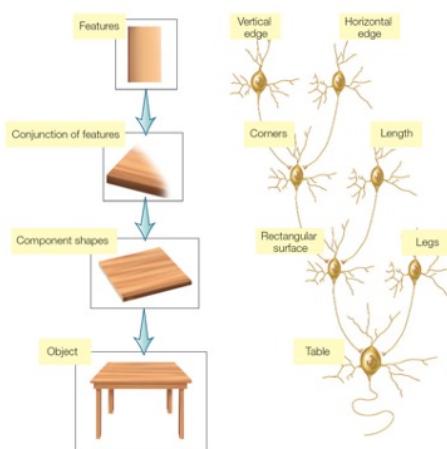


The majority of IT neurons respond to a stimulus only when it is presented from specific points of view (view-dependent responses).

Some neurons (10%) selectively respond to familiar stimuli regardless of their position with respect to the observer (view-independent responses).

These responses, although rare, indicate that IT is capable of forming a (relatively abstract) representation of the object, rather than responding to one of the different forms that the object can take when its position with respect to the observer changes.

Hierarchical model of the object recognition



The finding that IT cells selectively respond to more complex stimuli than V1, V2 and V4 is consistent with a hierarchical model of object perception.

According to this model, each subsequent level encodes more complex combinations from the inputs of the previous level.

The type of neuron that can recognize a complex object has been called the gnostic unit, referring to the idea that the cell (or cells) signals the presence of a complex, highly specific, and significant stimulus: that is, a known object, place or animal that has been encountered in the past.

Local or distributed coding?

It is tempting to conclude that the cell represented by the activity of IT cells signal the presence of an object (a hand or face), independent of the point of view.

In this regard, the researchers coined the term 'grandmother cell' to convey the idea that people's brains may have a gnostic unity that is activated only when the grandmother comes into view.

Other Gnostic units would specialize in recognizing, for example, a blue Volkswagen or the Golden Gate Bridge.

Distributed code hypothesis

An alternative to the Grandmother cell hypothesis is that object recognition is the result of a distributed activation pattern on the population of IT neurons.

According to this hypothesis, recognition is due not to one unit but to the collective activation of many units.

Distributed code theories easily explain why we can recognize similarities between objects (say, a tiger and a lion) and make mistakes between visually similar objects - both objects activate many of the same neurons.

Losing some units may degrade our ability to recognize an object, but the remaining units may be enough.

Distributed code theories also explain our ability to recognize new objects. New objects have a resemblance to familiar things, and our perceptions result from activating units that represent their characteristics.

153

The results of the studies on single neurons of the temporal lobe are in agreement with the theories of the distributed code of object recognition.

Although it is surprising that some cells are selective for complex objects, the selectivity is almost always relative, not absolute.

- IT neurons respond only to visual stimuli.
- The receptive fields always include the fovea, that is the part of the retina most involved in the fine recognition of a visual stimulus.
- The receptive fields tend to be large, providing the opportunity to generalize the stimulus within the receptive field, and often extend along the midline in both visual hemifields, thus joining the two halves of the space for the first time. This property depends on the interhemispheric connections through the splenium of the corpus callosum and the anterior commissure.
- IT neurons encode complex characteristics of the stimulus (not simple features, such as color, form orientation, depth).

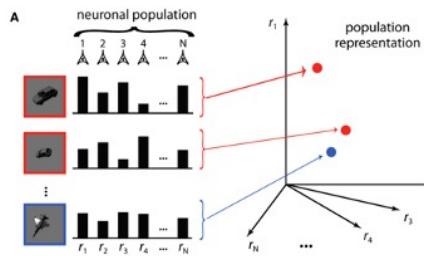
IT neuron selectivity often appears somewhat arbitrary.

A single IT neuron could, for example, respond vigorously to a crescent of a particular color and texture.

Cells with such selectivity likely provide inputs to higher-order neurons that respond to specific objects.

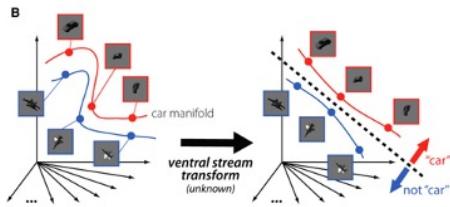
- A small percentage of IT units are selective for faces. Some are sensitive to emotional expression and others to the direction of gaze. Hand-selective cells are also found.
- Faces and probably other shapes appear to be encoded by a pattern of activity distributed over a set of cells, rather than by gnostic units (grandmother's cell,) that is, a cell that responds to complex and highly specific visual stimuli such as one's grandmother.
- The selectivity of IT cells is usually invariant with respect to changes in stimulus size, contrast, color, and exact location on the retina.
- There appears to be a vertical organization for the selectivity of the stimulus of IT neurons.
- The activity of IT neurons can be modulated by the animal's attention.
- IT cells can exhibit both short- and long-term memory effects for visual stimuli, and their selectivity can be changed by experience.

Ventral visual pathway gradually “untangles” information about object identity



Response of a population of neurons to a particular view of one object can be represented by a response vector in a space whose dimensionality is defined by the number of neurons in the population.

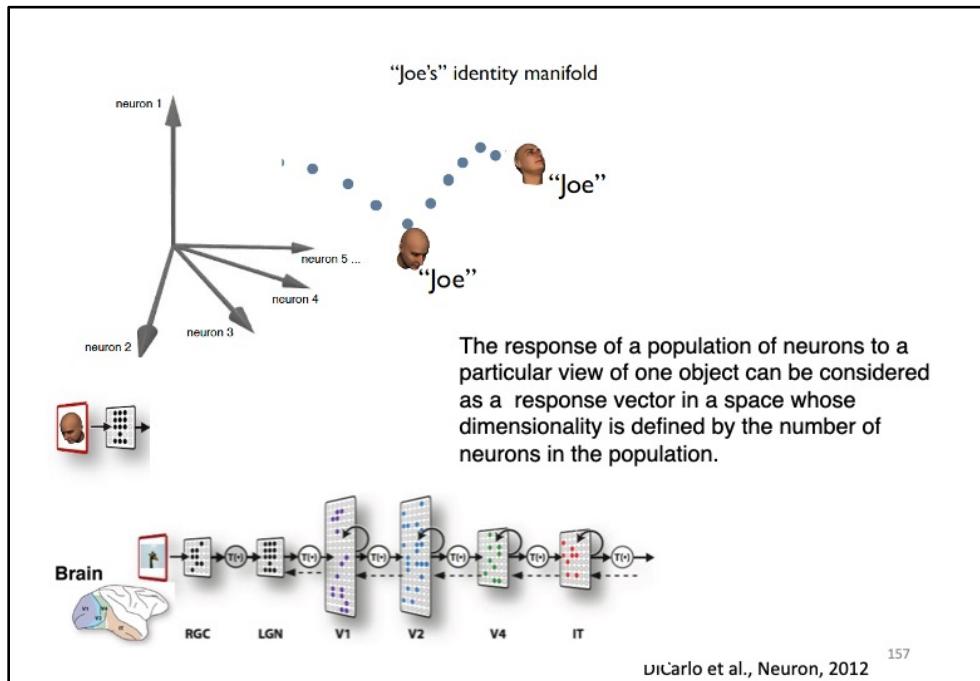
When an object undergoes an identity-preserving transformation, it produces a different pattern of population activity, which corresponds to a different response vector.



Together, the response vectors corresponding to all possible identity preserving transformations define a low-dimensional surface in this high-dimensional space—an object identity manifold.

DiCarlo et al., Neuron, 2012

156



Object recognition is the ability to separate representation that contain one particular object from representation that do not.

Thus, object manifolds are thought to be gradually untangled through nonlinear selectivity and invariance computations applied at each stage of the ventral pathway.

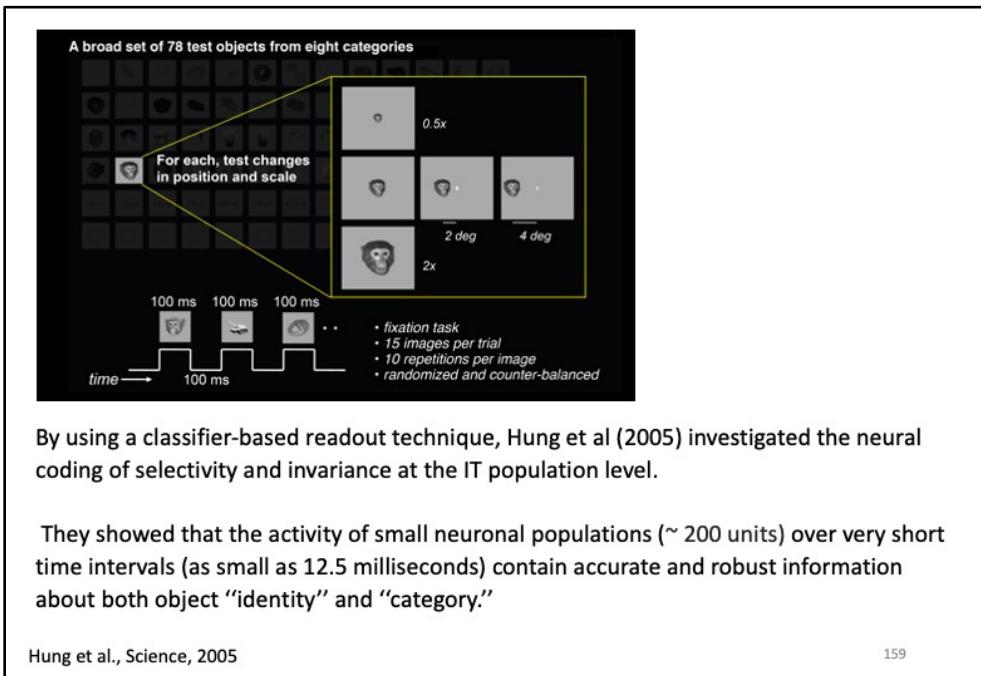
At higher stages of visual processing, neurons tend to maintain their selectivity for objects across changes in view; this translates to manifolds that are more flat and separated (more “untangled”).

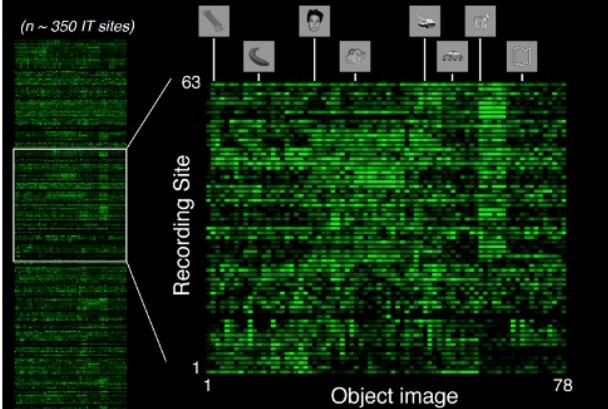
DiCarlo et al., Neuron, 2012

158

For neurons with small receptive fields that are activated by simple light patterns, such as retinal ganglion cells and V1, each object manifold will be highly curved.

Moreover, the manifolds corresponding to different objects will be “tangled” together,

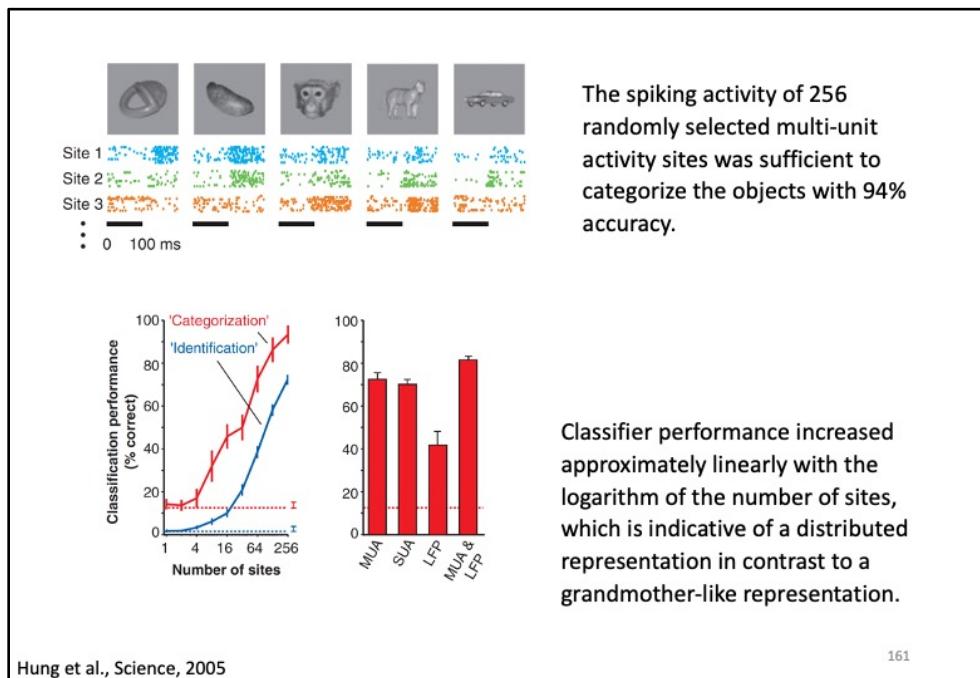


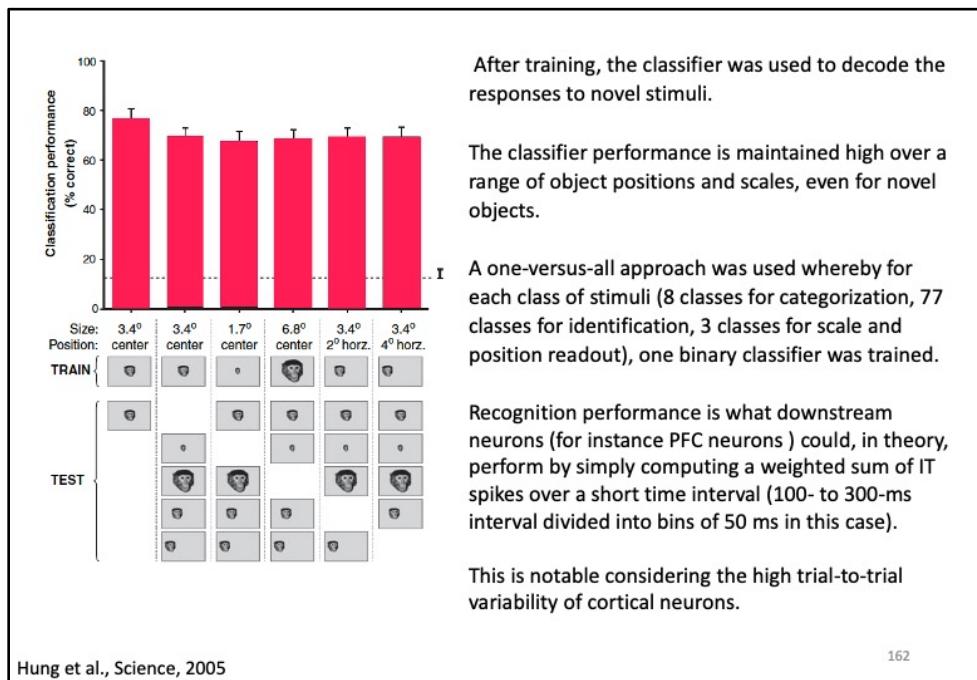


The readout technique consists of training a regularization classifier to learn the map from neuronal responses (from the independently recorded neurons) to each object label.

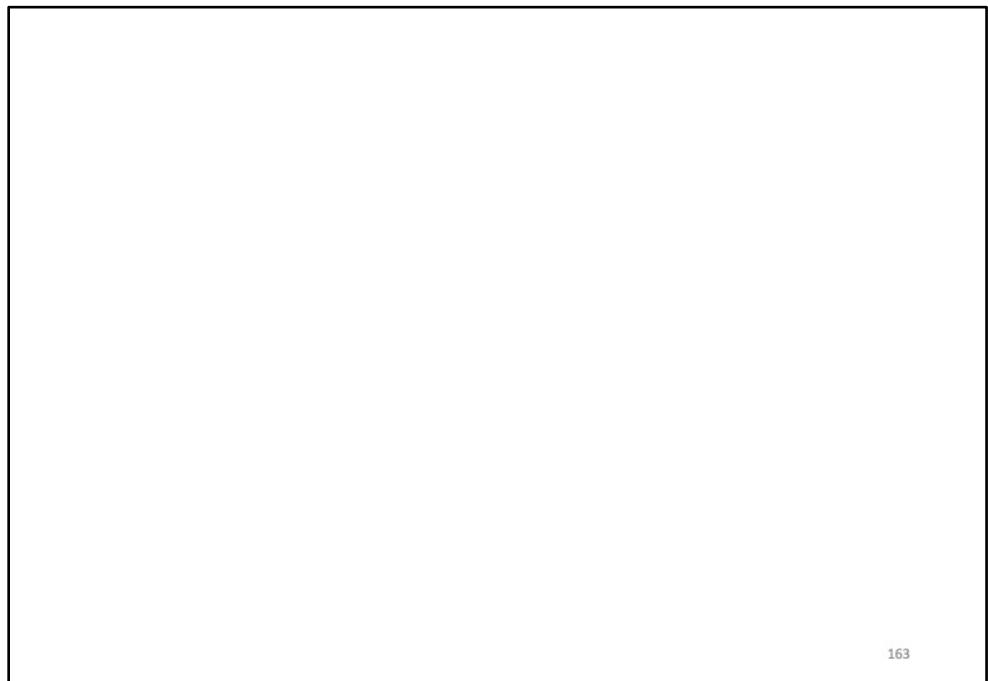
The input consists of the neuronal responses

The activity of ~ 250 randomly selected multi- and single-unit activity in response to 78 images of different objects were recorded at 350 IT sites in two monkeys





Objects could be reliably categorized and identified (with less than 10% reduction in performance) even when transformed (spatially shifted or scaled), although the classifier only saw each object at one particular scale and position during training.



163