**Visual Object Recognition**
**Computational Models and Neurophysiological Mechanisms**
Neuro 130/230. Harvard College/GSAS 78454

While we wait for everyone to join,
a brief math problem:

How many possible images are there of size 100 x 100 pixels, with 256 shades of gray per pixel?

# Visual Object Recognition
# Computational Models and Neurophysiological Mechanisms
Neurobiology 130/230. Harvard College/GSAS 78454

Note: no class on 09/04/2023 (Labor Day)

Class 1 [09/11/2023]. Introduction to Vision

Class 2 [09/18/2023]. The Phenomenology of Vision

**Class 3 [09/25/2023]. Natural image statistics and the retina**

Class 4 [10/02/2023]. Learning from Lesions

Note: no class on 10/09/2023 (Indigenous Day)

Class 5 [10/16/2023]. Primary Visual Cortex

Class 6 [10/23/2023]. Adventures into *terra incognita*

Class 7 [10/30/2023]. From the Highest Echelons of Visual Processing to Cognition

Class 8 [11/06/2023]. First Steps into in silico vision

Class 9 [11/13/2023]. Teaching Computers how to see

Class 10 [11/20/2023]. Computer Vision

Class 11 [11/27/2023]. Connecting Vision to the rest of Cognition [Dr. Will Xiao]

Class 12 [12/06/2023]. Visual Consciousness

**FINAL EXAM, PAPER DUE 12/11/2023. No extensions.**

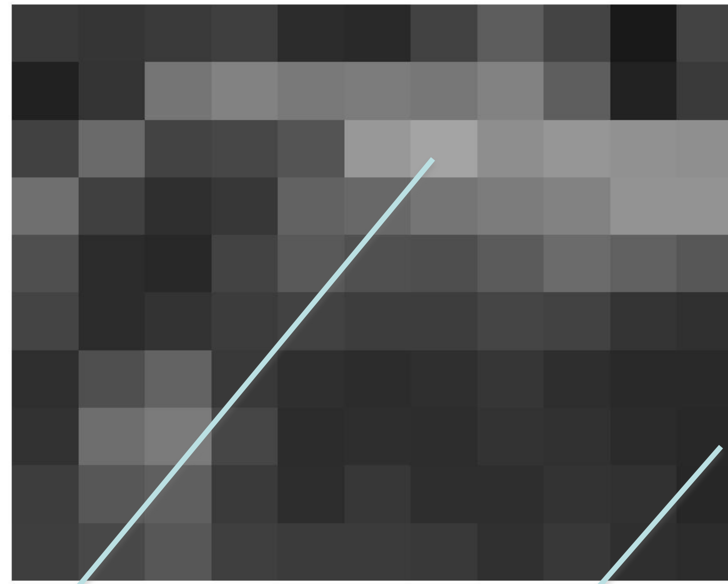
# Starting from the very beginning

- Let there be light, and there was light.
- Objects reflect light
- Light photons impinge on the retina (Latin: small net)
- The retina conveys visual information to the brain

An oversimplified first-order description:

*The eye functions as a very sophisticated and spectacular digital camera*

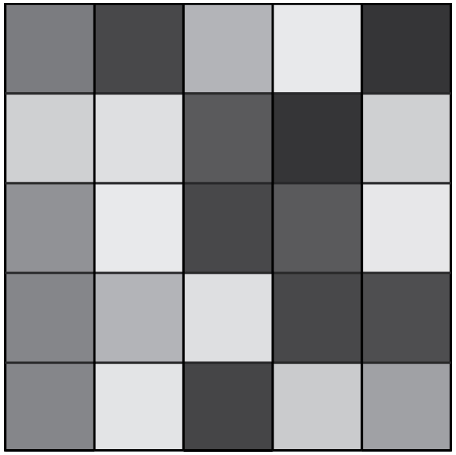
# An image as a collection of pixels



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 57 | 53 | 58 | 63 | 44 | 41 | 66 | 93 | 68 | 25 | 67 |
| 33 | 52 | 117 | 130 | 121 | 124 | 119 | 130 | 94 | 34 | 58 |
| 65 | 106 | 67 | 71 | 84 | 152 | 164 | 142 | 150 | 145 | 143 |
| 111 | 64 | 47 | 55 | 98 | 104 | 117 | 124 | 130 | 147 | 147 |
| 79 | 44 | 40 | 67 | 89 | 80 | 78 | 91 | 107 | 97 | 87 |
| 68 | 44 | 51 | 60 | 66 | 61 | 61 | 69 | 66 | 52 | 48 |
| 47 | 79 | 99 | 57 | 47 | 44 | 47 | 54 | 46 | 41 | 41 |
| 50 | 110 | 123 | 70 | 44 | 46 | 45 | 51 | 49 | 43 | 40 |
| 61 | 87 | 95 | 58 | 45 | 55 | 46 | 46 | 51 | 49 | 39 |
| 62 | 72 | 87 | 63 | 59 | 59 | 57 | 48 | 56 | 47 | 44 |
| 49 | 51 | 52 | 52 | 52 | 48 | 48 | 51 | 52 | 55 | 56 |

# Natural images are special

## We only encounter a small subset of the space of possible images



Consider an image of size 100 x 100 pixels
Assume a pixel can have 256 shades of gray
How many such images are possible?

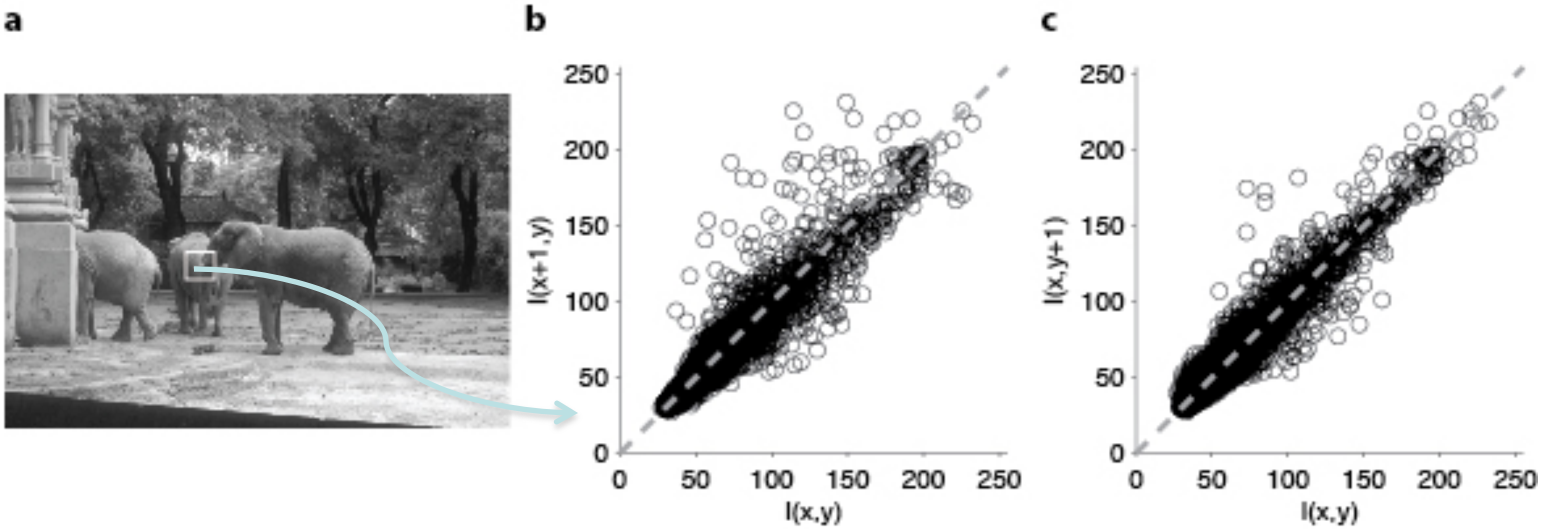
Answer
For a size of 1x1 pixel, there are 256 possible images
For a size of 1x2 pixels, there are $256^2$ possible images
For a size of 100x100 pixels, there are $256^{10000}$ possibilities*

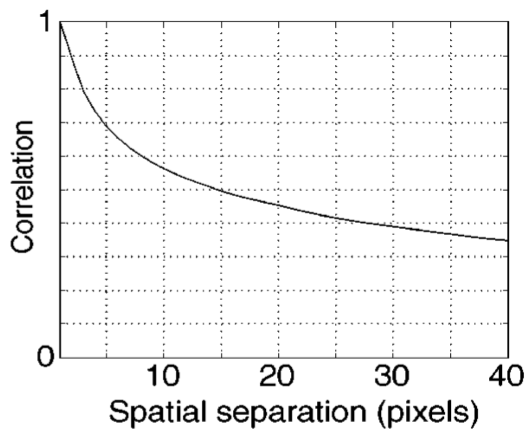
*Yet, we only encounter a small fraction of these possibilities in natural images*

*Some of those are "related" by translation, rotation or inversion, etc

# Spatial regularities in natural scenes

## The properties of nearby points are correlated

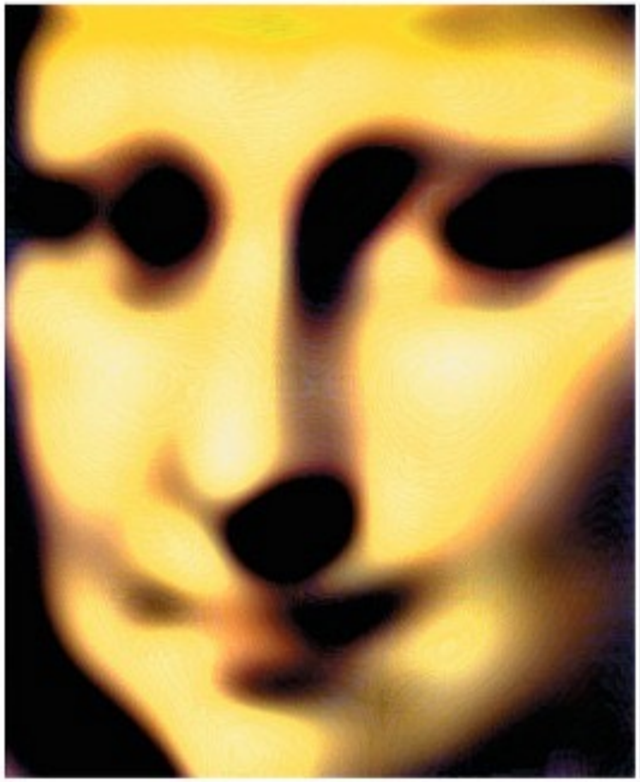



Simoncelli and Olshausen 2001

# Images contain different spatial frequencies

## Some are more prevalent than others



Low frequencies

Middle frequencies

High frequencies

Livingstone, M. (2002). Vision and Art: The Biology of Seeing. Harry N. Abrams.

# Natural image statistics

## Power spectrum ~ $1/f^2$

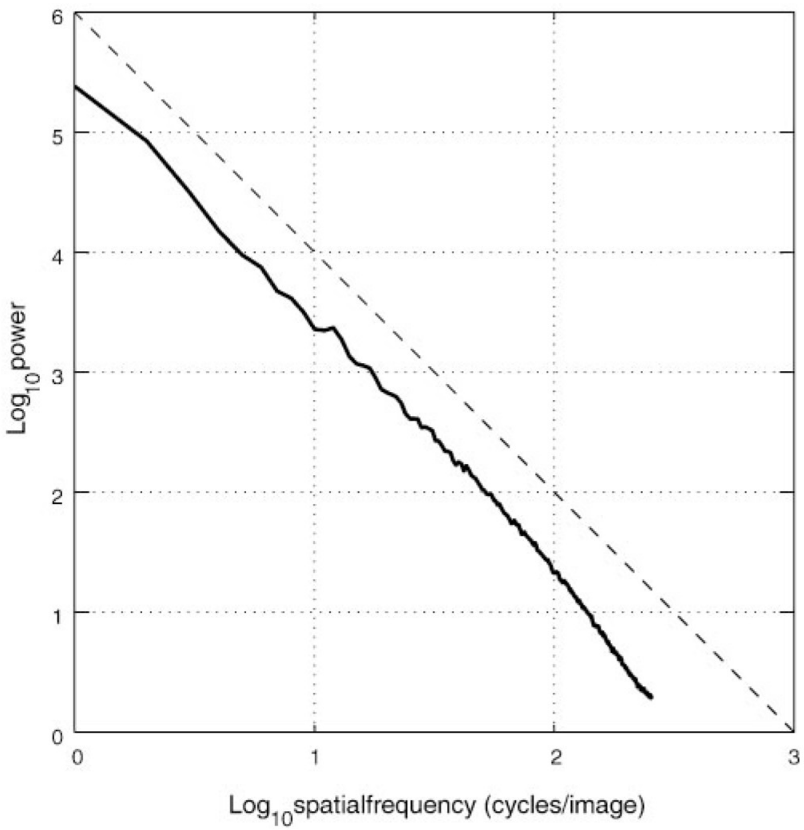



**Figure 4** Power spectrum of a natural image (solid line) averaged over all orientations, compared with $1/f^2$ (dashed line).

$$\log(f(w)) = \alpha \log(w) + c$$

Note: Scale invariance

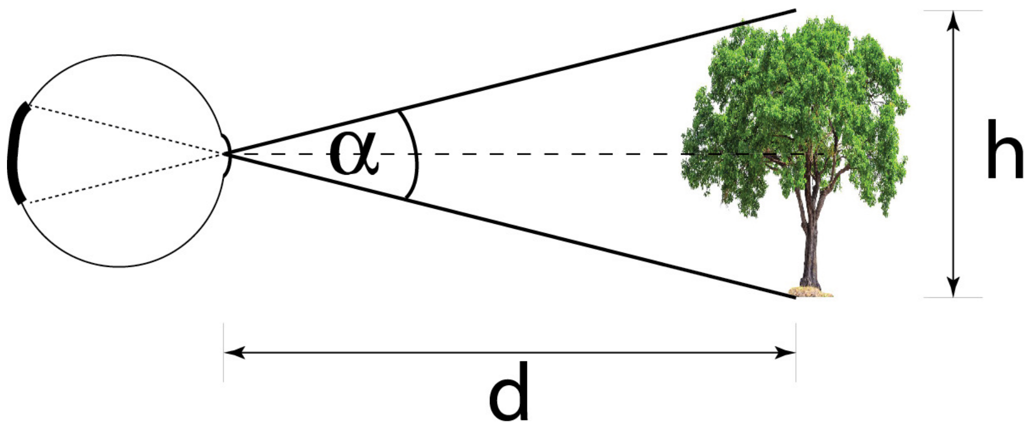
$$w' \to aw$$

$$\log(f(w')) = \beta \log(w) + d$$

There are multiple examples of power law distributions in physics, biology and social sciences

Simoncelli and Olshausen 2001

# Sizes are measured in degrees of visual angle




α / 2 = arctan (h/2d)

Size of the moon ~ 0.5 degrees
Thumb at arms length ~ 2 degrees

# Natural image statistics

## There are also strong correlations in time

The visual input is largely static, except for:

- External object movements
- Head movements
- Eye movements

The visual image is largely static over hundreds of milliseconds

**Silent Reading 225-250 ms fixation, 2 degrees saccade size (8-9 letters)**
**Scene Perception 260-330 ms fixation, 4 degrees saccade**

"Slowness" has been proposed as a constraint for learning about objects (Földiak 1991, Stringer et al 2006, Wiskott et al 2002, Li et al 2008)

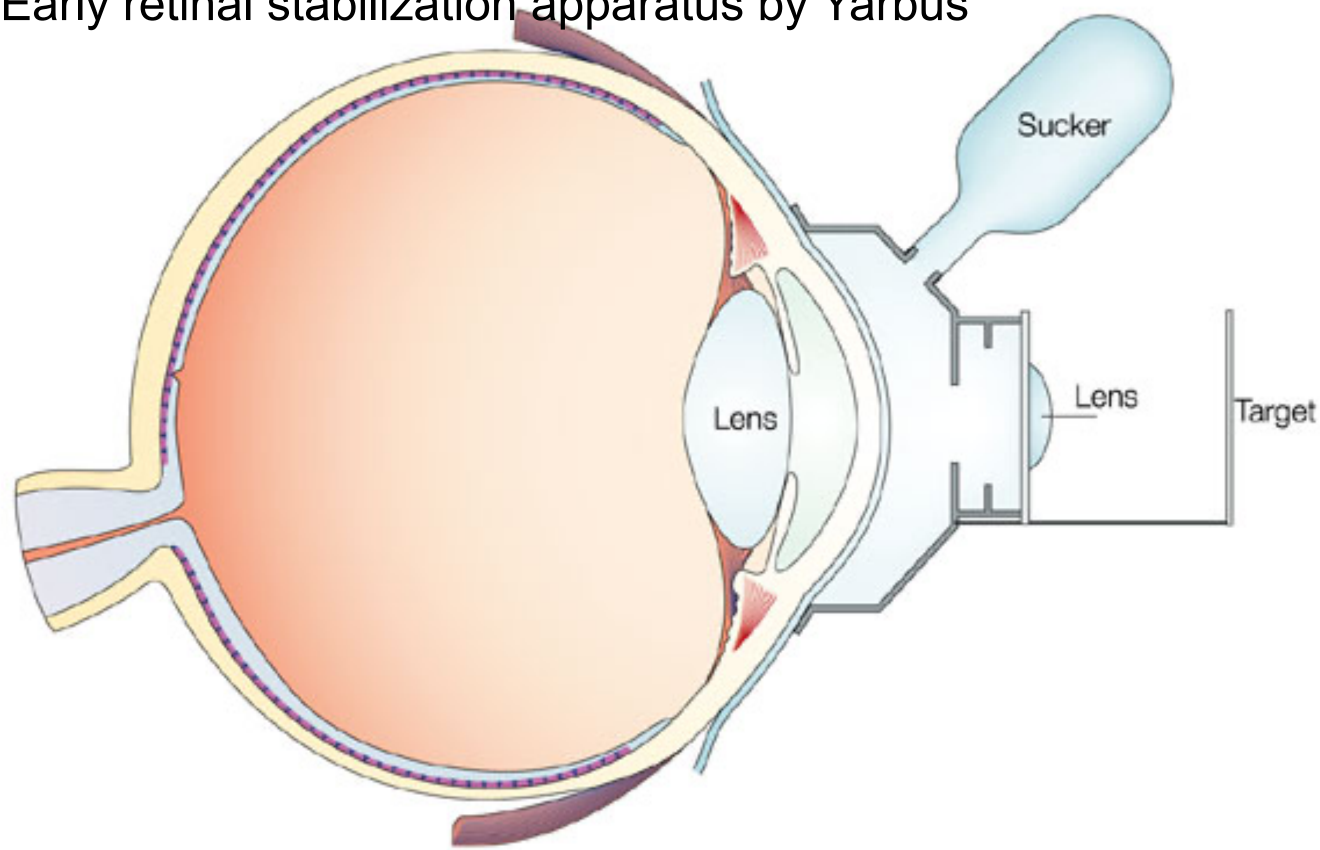
# Saccadic eye movements

0.033 secs

# Example pattern of fixations while examining an image

# Microsaccades are important for perception

Early retinal stabilization apparatus by Yarbus



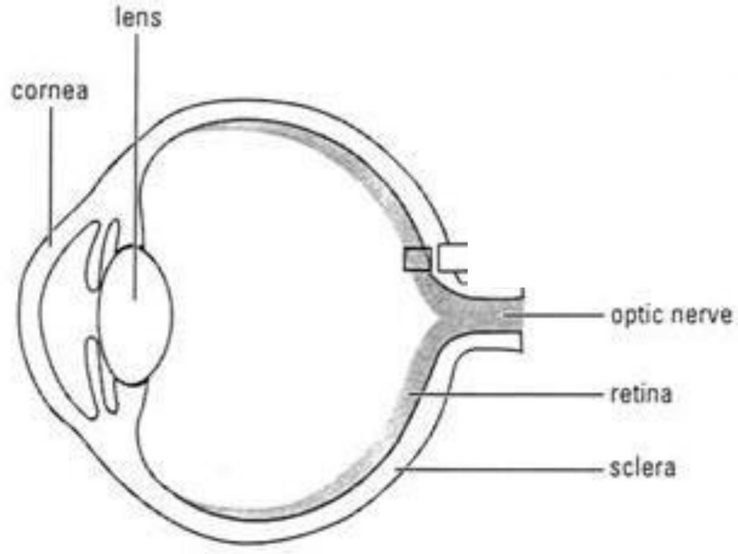

Martinez-Conde, Macknik, & Hubel, 2004

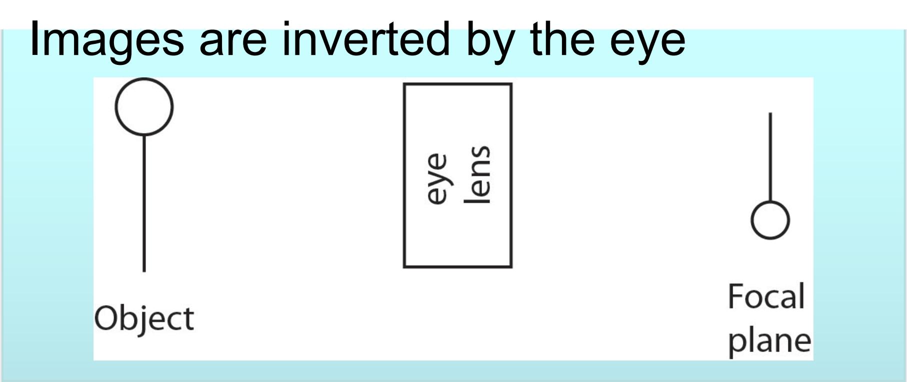
The eye constantly makes very small "fixational" movements, a.k.a. microsaccades.

When visual stimulation is held fixed relative to the retina, visual perception rapidly fades (Ratliff & Riggs, 1950).

These microsaccades may be critical for counteracting perceptual fading (McCamy, 2012)

# The image is focused onto the retina

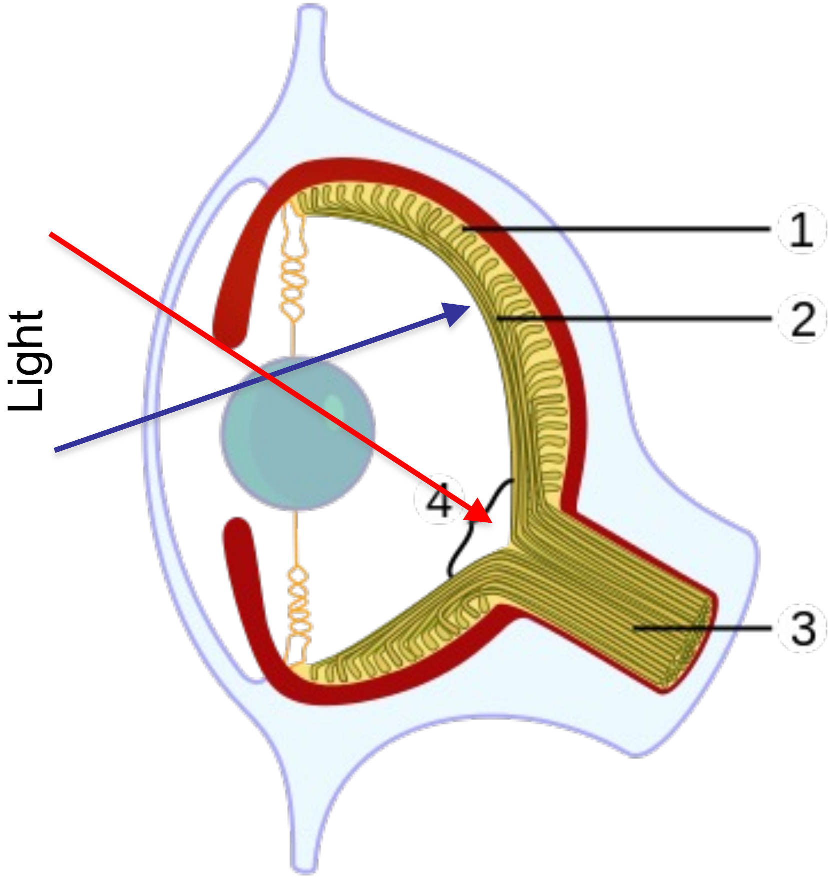


# Why don't we see everything upside down?

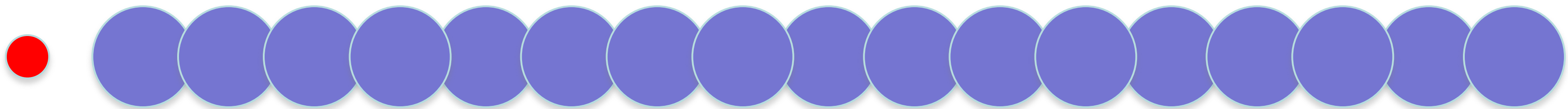
**Perception can adapt to reversing the image.**
Some preliminary experiments on vision without inversion of the retinal image. Stratton, 1896

# The blind spot

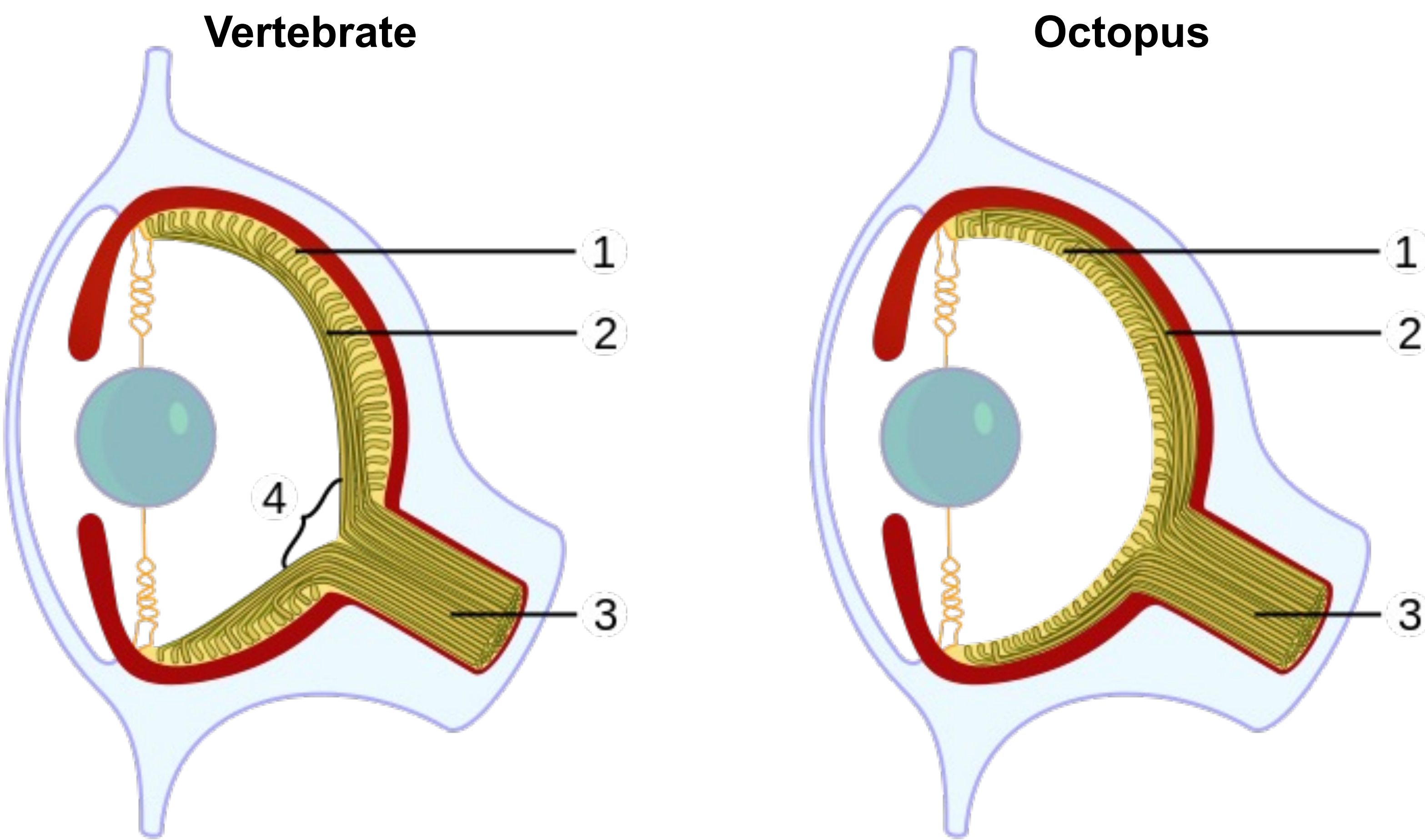


# “Seeing” the blind spot

Cover your left eye
Fixate on the red circle and do not move your eyes!



# Evolution of the retina




4. Blind spot is only present in vertebrates

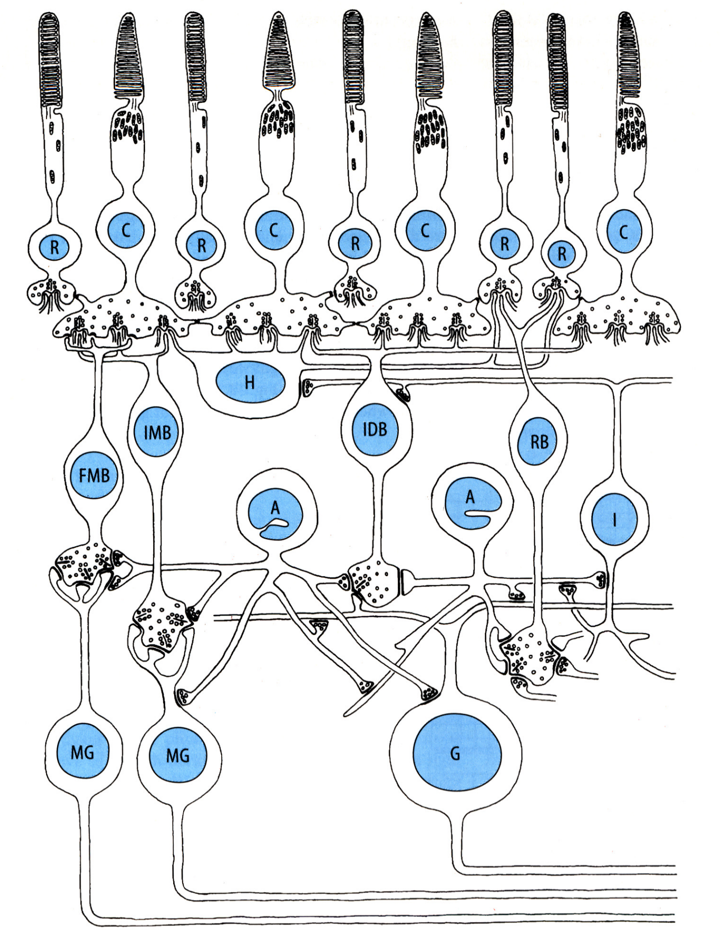
1. In octopi, light reaches directly the photoreceptors before reaching the ganglion cells

# Evolution of the eye:
# absurd in the highest degree

"To suppose that the eye, with all its inimitable contrivances for adjusting the focus to different distances, for admitting different mounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, **absurd in the highest degree**."

Charles Darwin, The origin of species, 1859

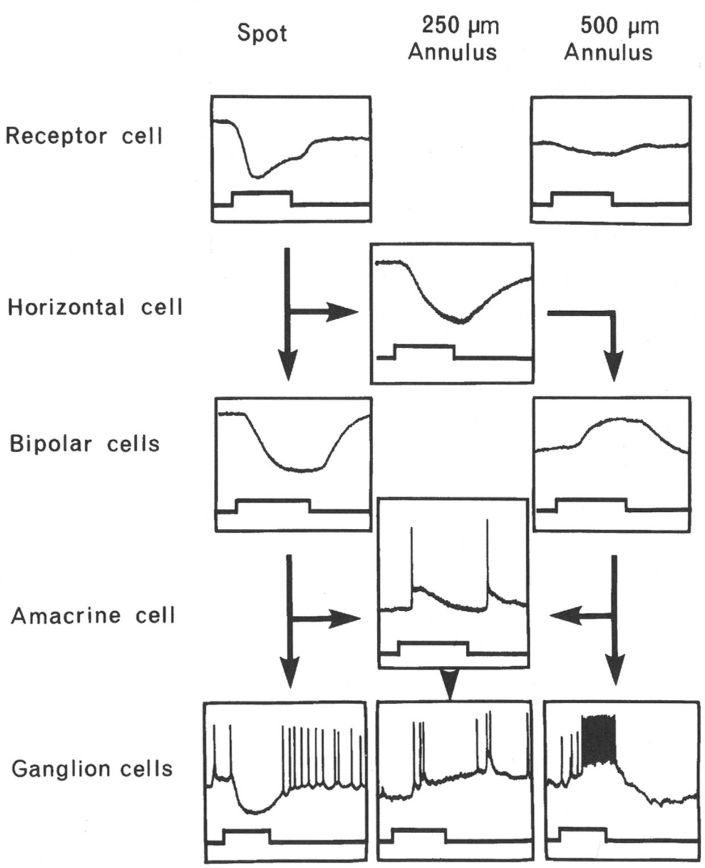
# The retina: An amazingly beautiful circuitry composed of many different cell types




- ~0.5 mm thick
- 5 x 5 cm retinal area
- Three cellular layers
- Rods (low-illumination conditions, ~$10^8$)
- Cones (color sensitivity, ~ $10^6$)
- Blind spot
- Fovea (rod free, ~0.5 mm, ~ 1.7 deg)
- Midget ganglion cells (small dendritic arbors)
- Parasol ganglion cells (large dendritic arbors)

Dowling (2007), Scholarpedia, 2:3487
Wandell (1995), Foundations of Vision. Sinauer Books

# The retina: Some cells fire action potentials whereas other cells show graded responses




- Photoreceptors transduce incoming light input into electrical signals

- Rod to bipolar convergence increases rod-pathway sensitivity

- Cones, rods, horizontal and bipolar cells are non-spiking neurons

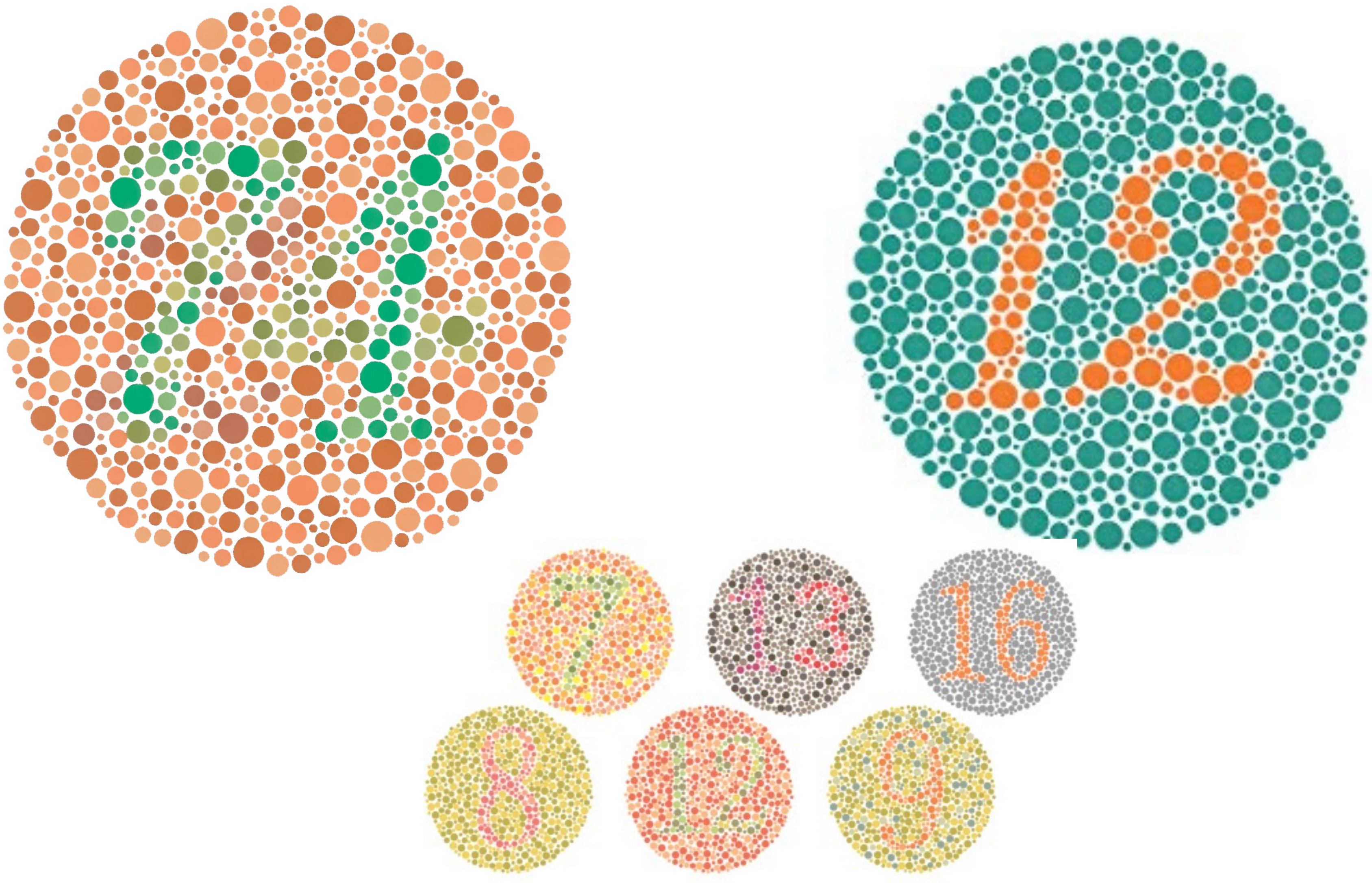
- Many different types of amacrine cells (some of which fire action potentials)

- Retinal ganglion cells fire action potentials and carry the output signals

John Dowling (2007), Scholarpedia, 2:3487.

# Rods see largely in grayscale

# Ishihara tests for color blindness



# Non-uniform sampling of the visual field
# There is much more detail at the *fovea*

# There is much more detail at the *fovea*

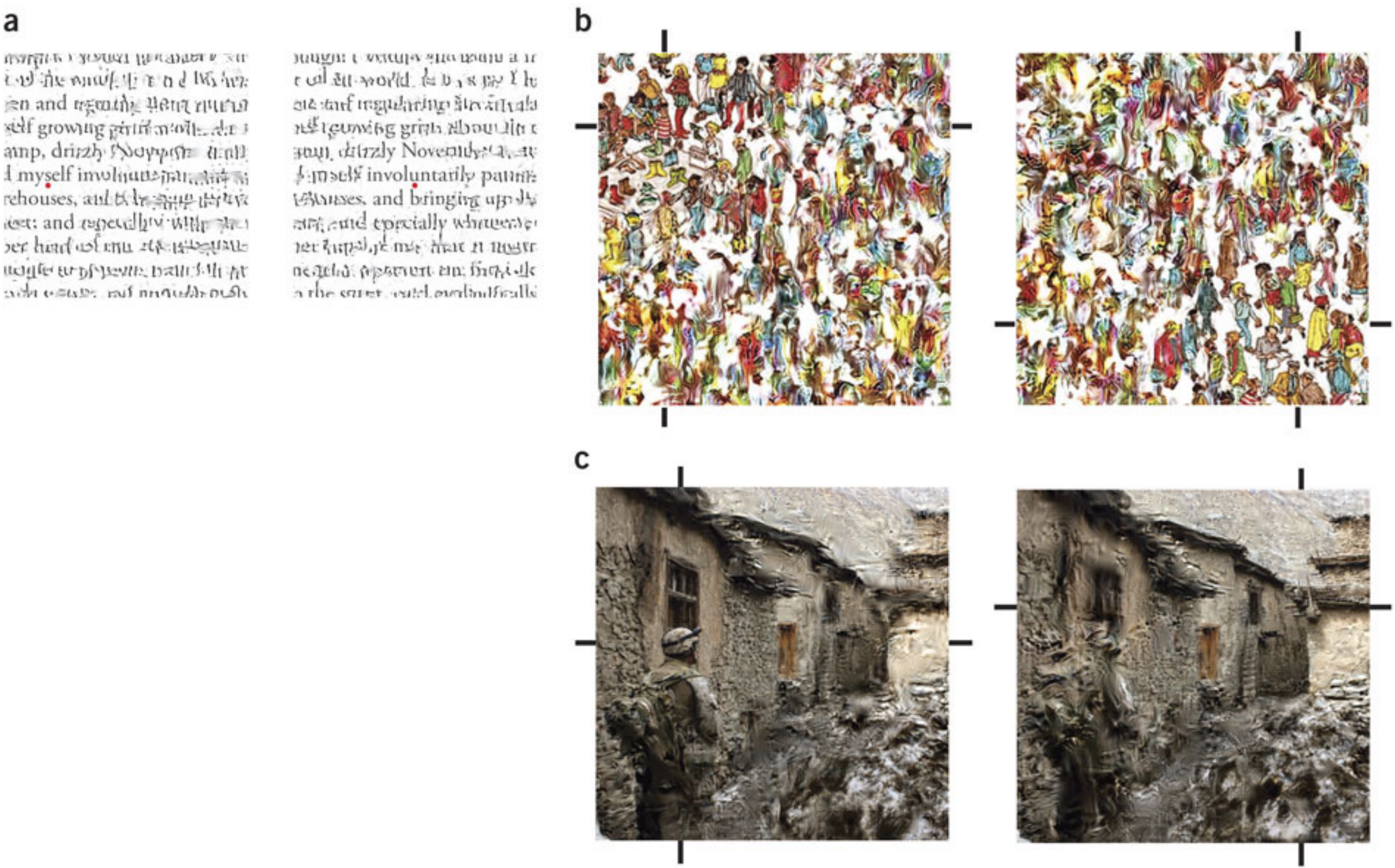
Try to read outside of your fixation spot

To be, or not to be, that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take Arms against a Sea of troubles,
And by opposing end them: to die, to sleep
No more; and by a sleep, to say we end
The heart-ache, and the thousand natural shocks
That Flesh is heir to? 'Tis a consummation
Devoutly to be wished. To die, to sleep,
perchance to **Dream**; aye, there's the rub,
For in that sleep of death, what dreams may come,
When we have shuffled off this mortal coil,
Must give us pause. There's the respect
That makes Calamity of so long life:
For who would bear the Whips and Scorns of time,
The Oppressor's wrong, the *proud* man's Contumely,
The pangs of *dispised* Love, the Law’s delay,
The insolence of Office, and the spurns

# There is much more detail at the *fovea*

Try to read outside of your fixation spot

# There is much more detail at the fovea



# Example psychophysics task

## Indistinguishable images due to reduced acuity in the periphery

Picture 1 is flashed
Picture 2 is flashed
Picture 3 is flashed

Was there an L or a T in the center?

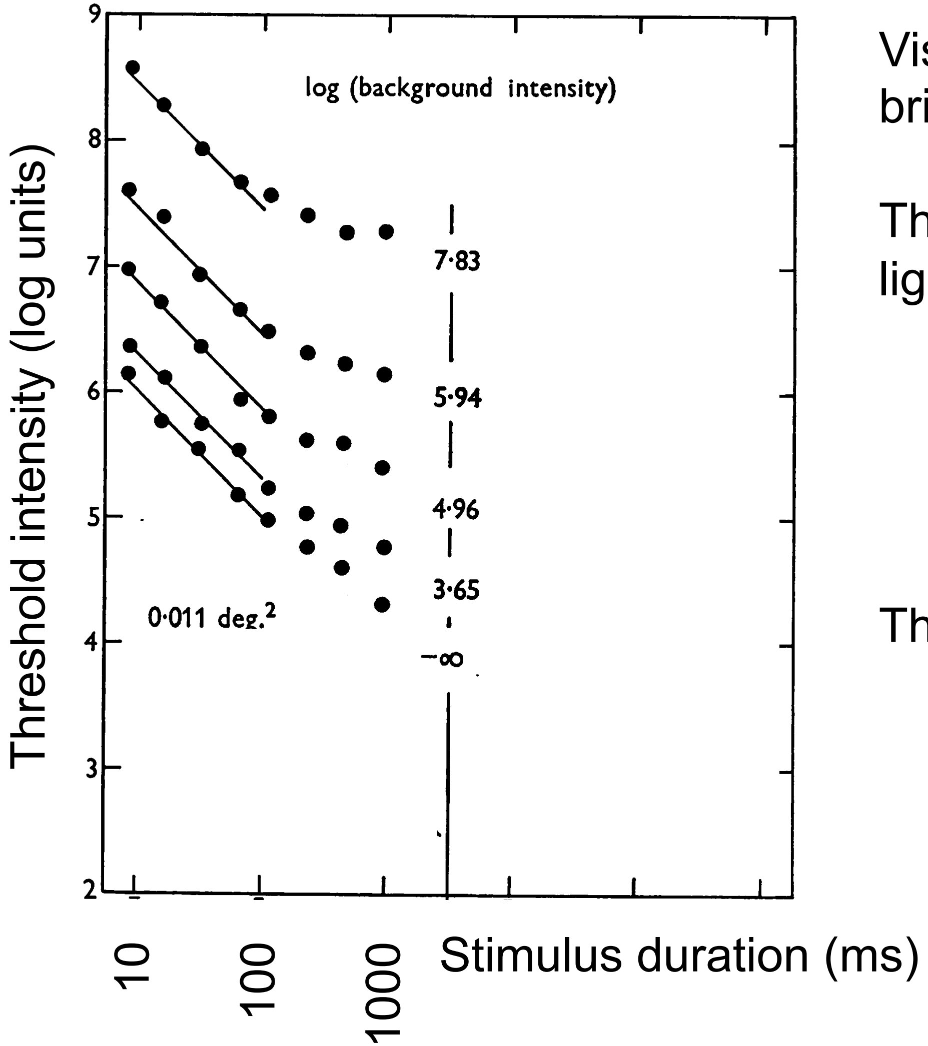
Was picture 3 more similar to picture 1 or picture 2?

Freeman and Simoncelli (2011). "Metamers of the ventral stream." Nat Neurosci **14**(9): 1195-1201


Sample 1 or 2?


1
2

# The retina has a huge dynamic range




Vision works well in moonless nights and in bright sunlight (~$10^9$ more light intensity)

There are several mechanisms that support light adaptation:

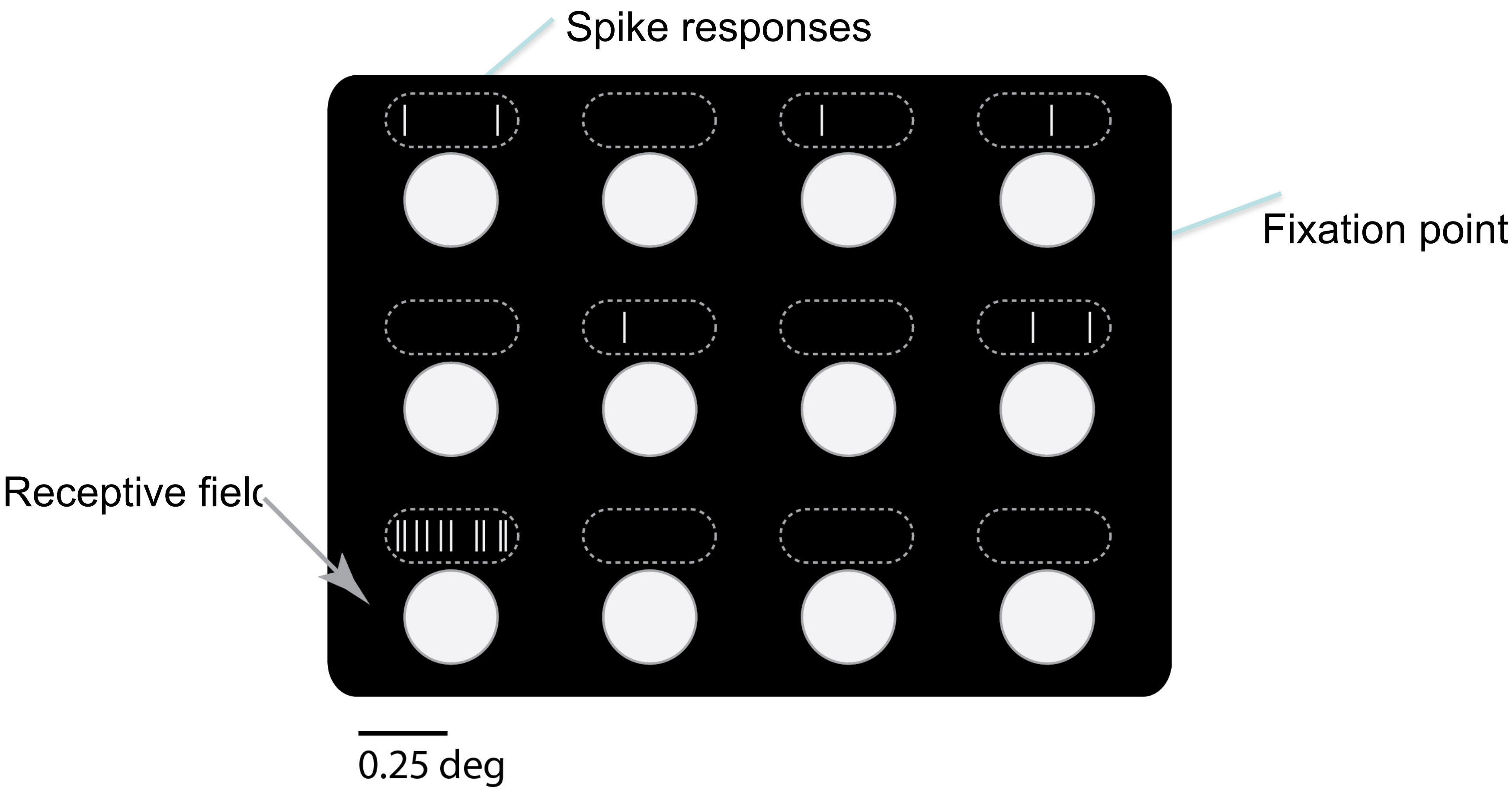
- Changing pupil size
- Reliance on rods vs. cones
- Photopigment bleaching
- Feedback from horizontal cells to photoreceptors

These mechanisms are relatively slow

Barlow, 1958

# The receptive field

## Neurons throughout the visual system are very picky about the stimulus location

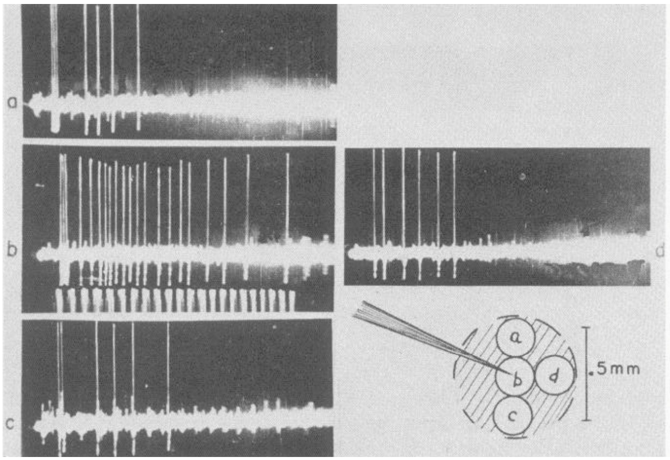



*This cartoon neuron responds only when a flash of light appears in the periphery, in the lower left quadrant*

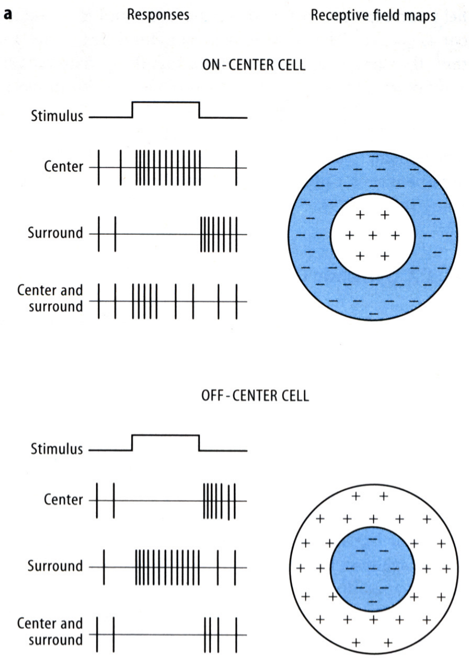
Blumberg and Kreiman, 2010

# Physiology of retinal ganglion cells

## The receptive field of most RGCs have a center-surround structure

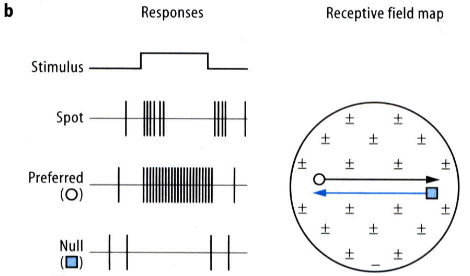



Kuffler, S. (1953)
*J. Neurophys.* **16**: 37-68




About 1.2 million projections from each retina towards the brain

Jonas et al., *Invest. Ophth. Vis. Sci.* 1992



# Diversity of retinal ganglion cells

A minority of RGCs have more complex response properties:

• Phasic cells respond briefly to stimulus onset, offset, or both
• Some phasic cells respond selectively to edge orientation
• Suppressed-by-contrast cells fire except when an edge is present in receptive field
• Bistratified RGCs lack surrounds and are color-sensitive
• Color-opponent cells have centers and surrounds with opposing color preferences
• Intrinsically photosensitive RGCs contain photoreceptors and project to regions controlling pupil size, circadian rhythm, etc.
• Direction-sensitive cells respond to direction of motion of light or dark spots

These cells likely account for approximately 10% of RGCs
Unclear to what extent, these cells contribute to visual object recognition

Stone and Fukuda, *Journal of Neurophysiology* 1974
Cleland and Levick, *Journal of Neurophysiology* 1974
Berson et al., *Science* 2002

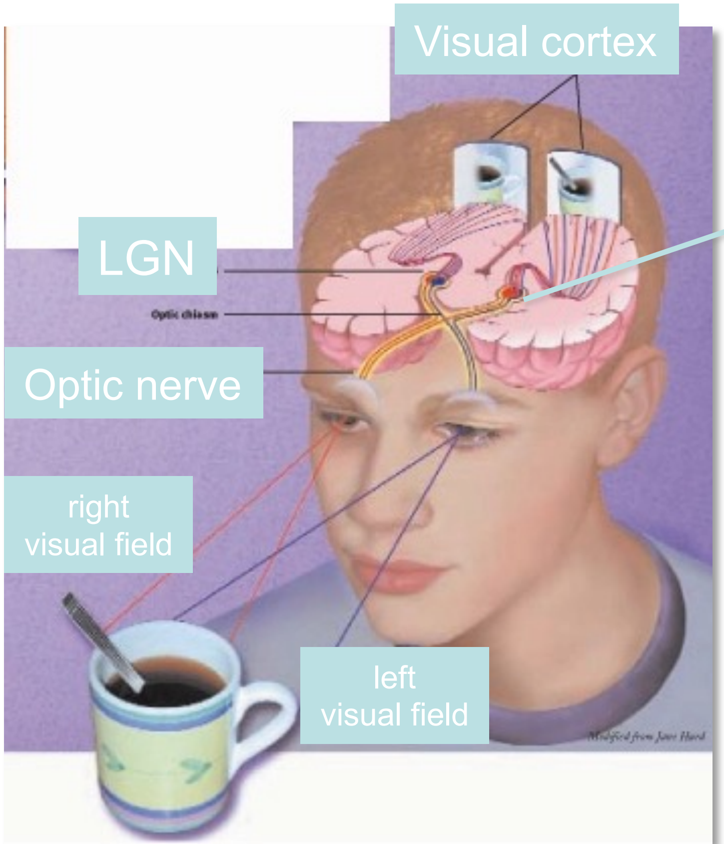
# Eye versus digital cameras

| | One eye | Camera (phone) |
|---|---|---|
| Angle of view | ~ 120 degrees | ~ 70 degrees |
| Sampling | Eccentricity dependent | Uniform |
| Resolution | ~ 10 "MP" in the fovea | ~ 20 MP (million pixels) |
| Sensitivity | High sensitivity in low light conditions | Struggle in low light and requires very long integration time |
| Energy efficiency | Small fraction of ~3 meals a day, say 14 hours of continuous use. (2500 calories / day ~ 0.12 Watts) | Continuous video recording: ~ 3 hours on one battery life (~5 Watts, hence > 12 watts a day ) |
| Malleability | Hard to change (without additional devices like glasses, microscopes, telescopes) | Can expand in many ways |
| Post-processing | The human brain! | Minimal computer vision* |

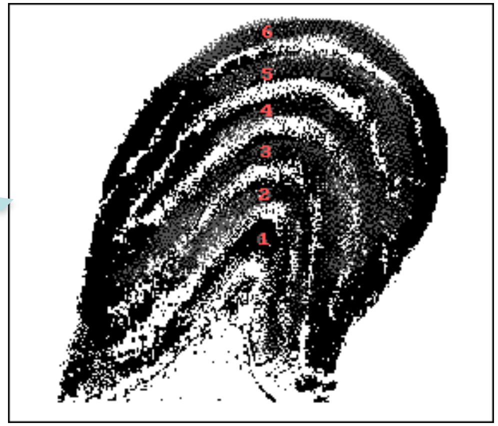
# Eyes in non-human animals

| | Humans | Other animals | Examples |
|---|---|---|---|
| Wavelength | "Visible" spectrum | Also UV, IR | mice, dogs, birds: UV<br>snakes: IR |
| Cones | 3 | 1-30 | bat, racoon: 1<br>cats, dogs: 2<br>some dragon flies: 30 |
| Polarization | No | Some animals | cuttlefish |
| Number of eyes | 2 | Up to 10 | spiders: 8-12<br>horseshoe crabs: 10 |
| Angle | Binocular: ~120$^o$<br>Up to 180$^o$ | Up to 300$^o$ | rabbits: no binocular zone in the center<br>cows, horses: 300$^o$ |
| Eyes | Correlated | Sometimes independent | Chameleon: can move eyes independently |
| Resolution | ~100M receptors | Wide variation | Star fish: 200 receptors<br>Birds: higher acuity than humans |

# To cortex, through the thalamus:
## The Lateral Geniculate Nucleus (LGN)




Modified from Society for Neuroscience Brain Facts




- LGN main visual part of the thalamus
- 6 layers, contralateral visual hemifield
- The visual field is represented multiple times in the LGN

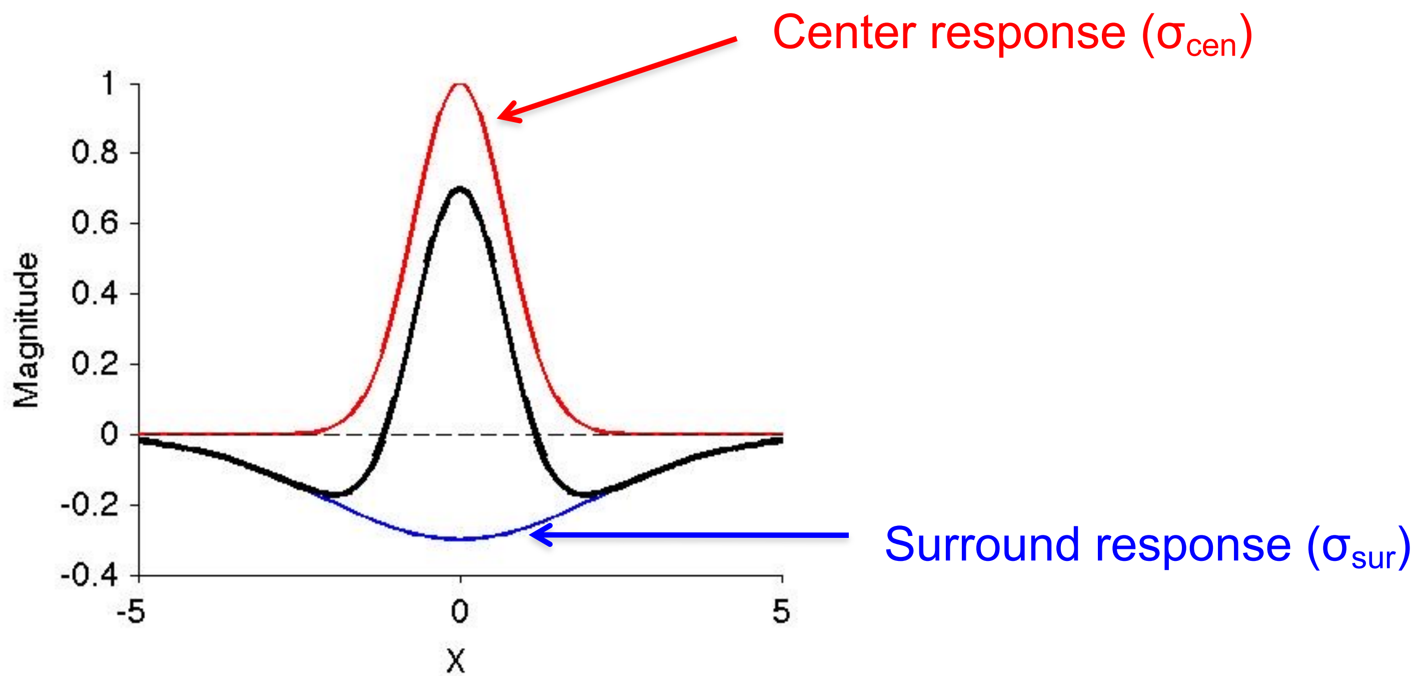
- Layers 2, 3 and 5 receive *ipsilateral* eye's input
- Layers 1, 4 and 6 receive *contralateral* eye's input
- Layers 1-2: *magnocellular* layers, input from parasol retinal ganglion cells (~ motion)
- Layers 3-6: *parvocelluar* layers, input from midget ganglion cells (~ color)
- On and Off center cells are present in all layers
- LGN does not project back to the retina

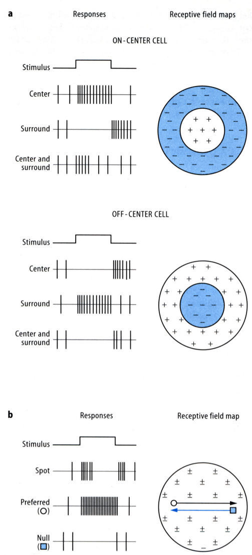
NOTE: Most of the input to the LGN comes from visual cortex and not from the retina! (e.g. Douglas and Martin 2004)

Wandell (1995)

# Modeling receptive fields: difference of Gaussians

The center-surround structure can be described by a difference of gaussians (Mexican-hat)




$$D(x,y) = \pm\left(\frac{1}{2\pi\sigma_{cen}^2}\exp\left[-\frac{x^2+y^2}{2\sigma_{cen}^2}\right] - \frac{B}{2\pi\sigma_{sur}^2}\exp\left[-\frac{x^2+y^2}{2\sigma_{sur}^2}\right]\right)$$

Dayan and Abbott. (2001) Theoretical Neuroscience. The MIT Press

# Neurons respond with transient bursts of activity

Dynamic receptive fields in the retina/LGN

$$D(x,y,t) = \pm\left(\frac{D_{cen}(t)}{2\pi\sigma^2_{cen}}\exp\left[-\frac{x^2+y^2}{2\sigma^2_{cen}}\right] - \frac{BD_{sur}(t)}{2\pi\sigma^2_{sur}}\exp\left[-\frac{x^2+y^2}{2\sigma^2_{sur}}\right]\right)$$

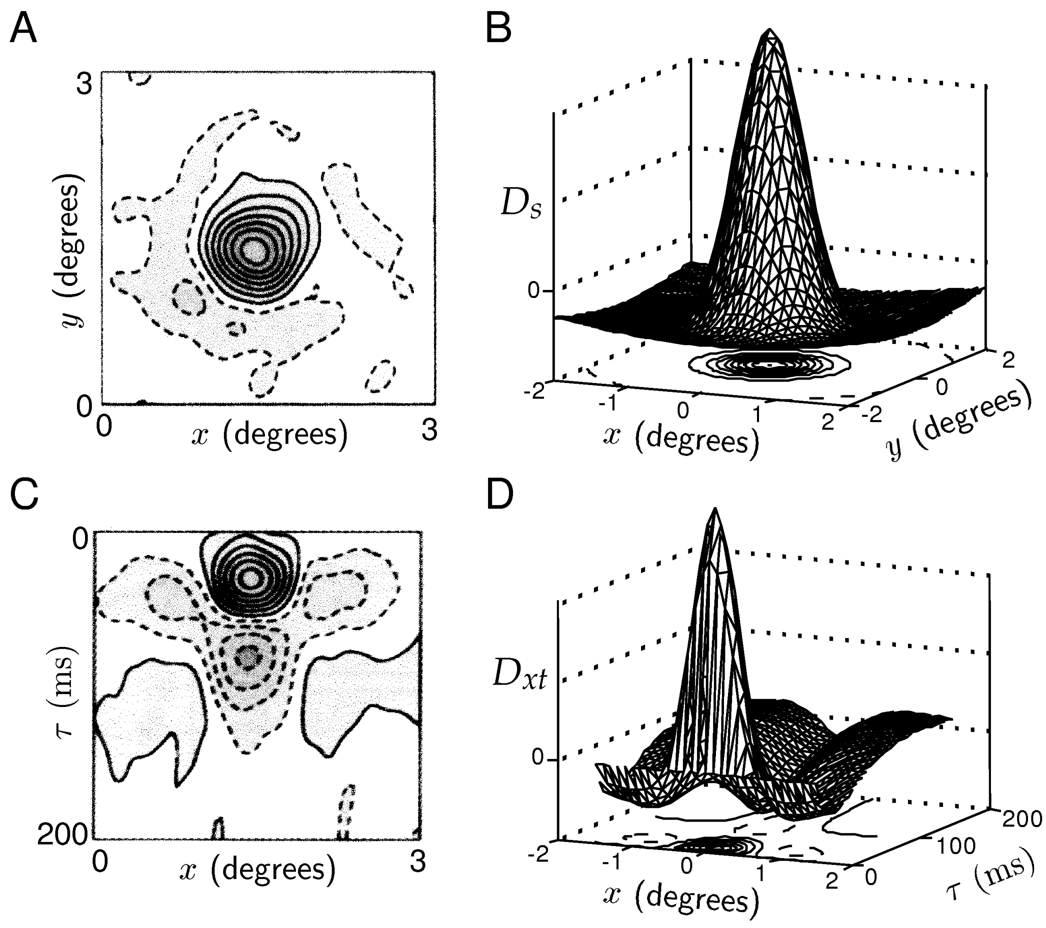
$$D_{cen}(t) = \alpha^2_{cen} t\exp[-\alpha_{cen} t] - \beta^2_{cen} t\exp[-\beta_{cen} t]$$

$$D_{sur}(t) = \alpha^2_{sur} t\exp[-\alpha_{sur} t] - \beta^2_{sur} t\exp[-\beta_{sur} t]$$

Dayan and Abbott. (2001) Theoretical Neuroscience. The MIT Press

# Difference of Gaussians in space and time

## The center-surround structure can also be seen in receptive field dynamics




Dayan and Abbott. (2001) Theoretical Neuroscience. The MIT Press

# Information from the retina goes to many places
## Subcortical visual pathways

Retinal projections

**Lateral geniculate nucleus (LGN)**

Superior Colliculi — Main visual pathway in birds, reptiles, fish

Implicated in saccade generation in mammals

Suprachiasmatic Nucleus — Hypothalamus: involved in circadian rhythms

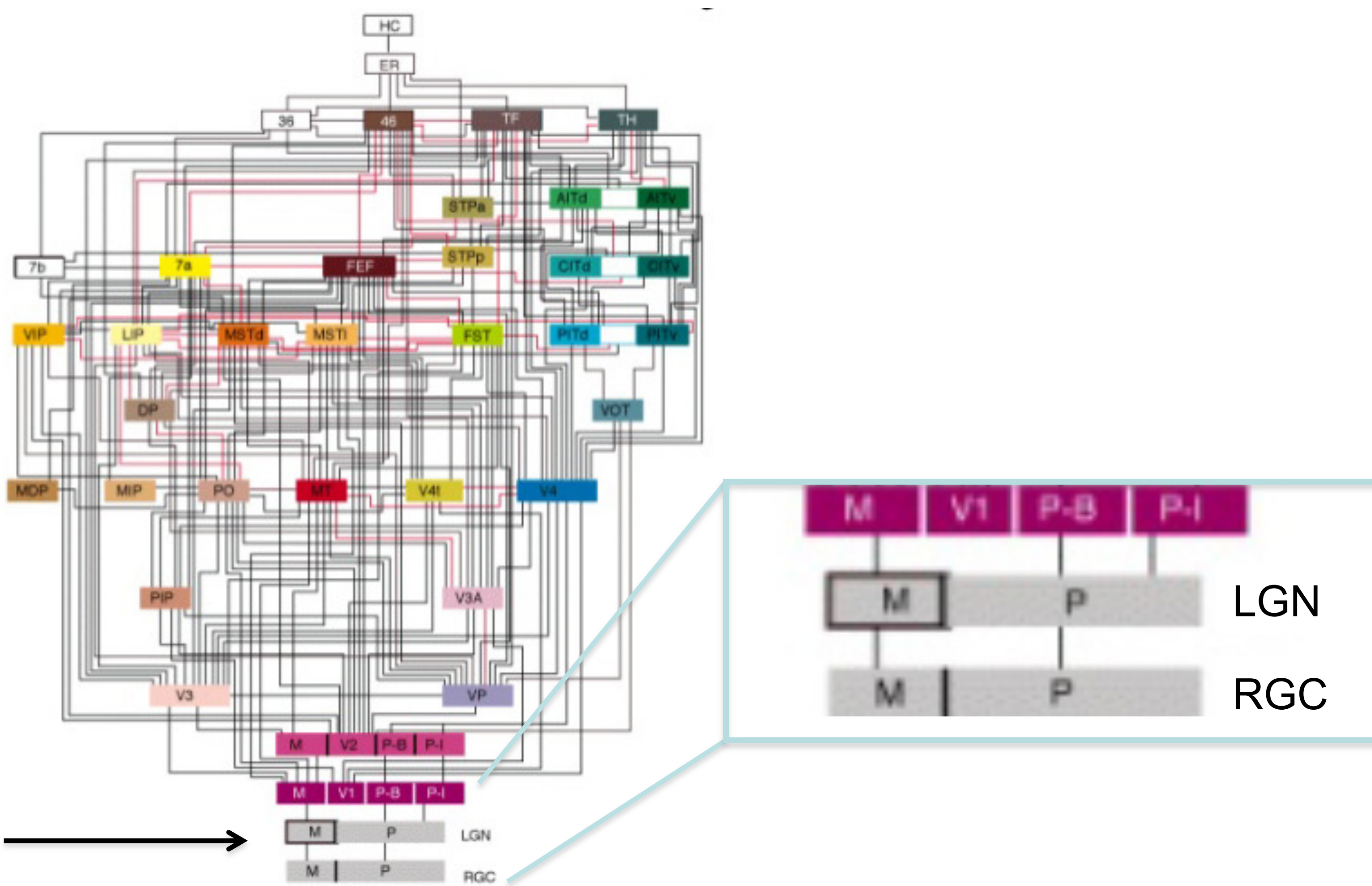
Pretectum

Pregeniculate

Accesory optic system

Primates can recognize objects after lesions to the Superior Colliculus or Suprachiasmatic Nucleus, but not after lesions to V1 (Gross 1994).

# Visual system circuitry




You are here

Felleman and Van Essen. *Cerebral Cortex* 1991

# Cited works (Class notes @ http://tinyurl.com/vision-class)


- Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. The Journal of physiology, 141(2), 337-350.
- Barlow, H. (1972). Single units and sensation: a neuron doctrine for perception. Perception 1, 371-394.
- Berson, D. M., Dunn, F. A., & Takao, M. (2002). Phototransduction by retinal ganglion cells that set the circadian clock. Science, 295(5557), 1070-1073.
- Blumberg, J., & Kreiman, G. (2010). How cortical neurons help us see: visual recognition in the human brain. The Journal of clinical investigation, 120(9), 3054-3063.
- Cleland, B. G., & Levick, W. R. (1974). Brisk and sluggish concentrically organized ganglion cells in the cat's retina. The Journal of Physiology, 240(2), 421-456.
- Dayan, P., and Abbott, L. (2001). Theoretical Neuroscience (Cambridge: MIT Press).
- Douglas, R.J., and Martin, K.A. (2004). Neuronal circuits of the neocortex. Annu Rev Neurosci 27, 419-451.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1(1), 1-47.
- Foldiak, P. (1991). Learning Invariance from Transformation Sequences. Neural Computation 3, 194-200.
- Freeman and Simoncelli (2011). Metamers of the ventral stream. Nat Neurosci 14(9): 1195-1201
- Gross, C.G. (1994). How inferior temporal cortex became a visual area. Cerebral Cortex 5, 455-469.
- Jonas, J. B., Schmidt, A. M., Müller-Bergh, J. A., Schlötzer-Schrehardt, U. M., & Naumann, G. O. (1992). Human optic nerve fiber count and optic disc size. *Investigative ophthalmology & visual science*, *33*(6), 2012-2018.
- Kuffler, S. (1953). Discharge patterns and functional organization of mammalian retina. Journal of Neurophysiology 16, 37-68.
- Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. Science, 321(5895), 1502-1507.
- Livingstone, M. (2002). Vision and Art: The Biology of Seeing. Harry N. Abrams.
- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. Nature Reviews Neuroscience, 5(3), 229-240.
- McCamy, M. B., Otero-Millan, J., Macknik, S. L., Yang, Y., Troncoso, X. G., Baer, S. M., & Martinez-Conde, S. (2012). Microsaccadic efficacy and contribution to foveal and peripheral vision. The Journal of Neuroscience, 32(27), 9194-9204.
- Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607-609.
- Ratliff, F., & Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. Journal of experimental psychology, 40(6), 687.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychol Bull 124, 372-422.
- Simoncelli, E., and Olshausen, B. (2001). Natural Image Statistics and Neural Representation. Annual Review of Neuroscience 24, 193-216.
- Stone, J., & Fukuda, Y. (1974). Properties of cat retinal ganglion cells: a comparison of W-cells with X-and Y-cells. Journal of Neurophysiology, 37(4), 722-748.
- Stratton, G. (1896). Some preliminary experiments on vision without inversion of the retinal image. Psychological Review 3, 611-617.
- Stringer, S.M., Perry, G., Rolls, E.T., and Proske, J.H. (2006). Learning invariant object recognition in the visual system with continuous transformations. Biol Cybern 94, 128-142.
- Wandell, B.A. (1995). Foundations of vision (Sunderland: Sinauer Associates Inc.).
- Wiskott, L., and Sejnowski, T.J. (2002). Slow feature analysis: unsupervised learning of invariances. Neural Comput 14, 715-770.
- Yau, K. (1994). Phototransduction mechanism in retinal rods and cones. Investigative Opthalmology and Visual Science 35, 9-32.