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

Web site: http://tinyurl.com/visionclass
→ Class notes, Class slides, Readings Assignments
Location: Biolabs 2062
Time: Mondays 03:00 – 05:00

Lectures:
Faculty: Gabriel Kreiman and invited guests
TA: Emma Giles

Contact information:
Gabriel Kreiman
gabriel.kreiman@tch.harvard.edu
617-919-2530

Emma Giles
emmagiles@g.harvard.edu

Office Hours: After Class. Mondays 5pm, or by appointment

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

Class 1 [09/10/2018]. Introduction to pattern recognition [Kreiman]

Class 2 [09/17/2018]. Why is vision difficult? Natural image statistics. The retina. [Kreiman]

Class 3 [09/24/2018]. Lesions and neurological studies [Kreiman].

Class 4 [10/01/2018]. Psychophysics of visual object recognition [Sarit Szpiro]

October 8: University Holiday

Class 5 [10/15/2018]. Primary visual cortex [Hartmann]

Class 6 [10/22/2018]. Adventures into *terra incognita* [Frederico Azevedo]

Class 7 [10/29/2018]. High-level visual cognition [Diego Mendoza-Haliday]

Class 8 [11/05/2018]. Correlation and causality. Electrical stimulation in visual cortex [Kreiman]

Class 9 [11/12/2018]. Visual consciousness [Kreiman]

Class 10 [11/19/2018]. Computational models of neurons and neural networks. [Kreiman]

Class 11 [11/26/2018]. Computer vision. Artificial Intelligence in Visual Cognition [Bill Lotter]

Class 12 [12/03/2018]. The operating system for vision. [Xavier Boix]

FINAL EXAM, PAPER DUE 12/13/2018. No extensions.

Starting from the very beginning

- Let there be light, and there was light.

- Objects reflect light

- Light photons impinge on the retina

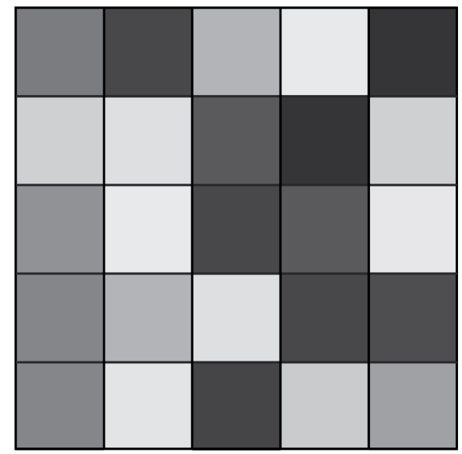
- The retina conveys visual information to the brain

An oversimplified first-order description:

The retina functions as a very sophisticated and spectacular digital camera

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

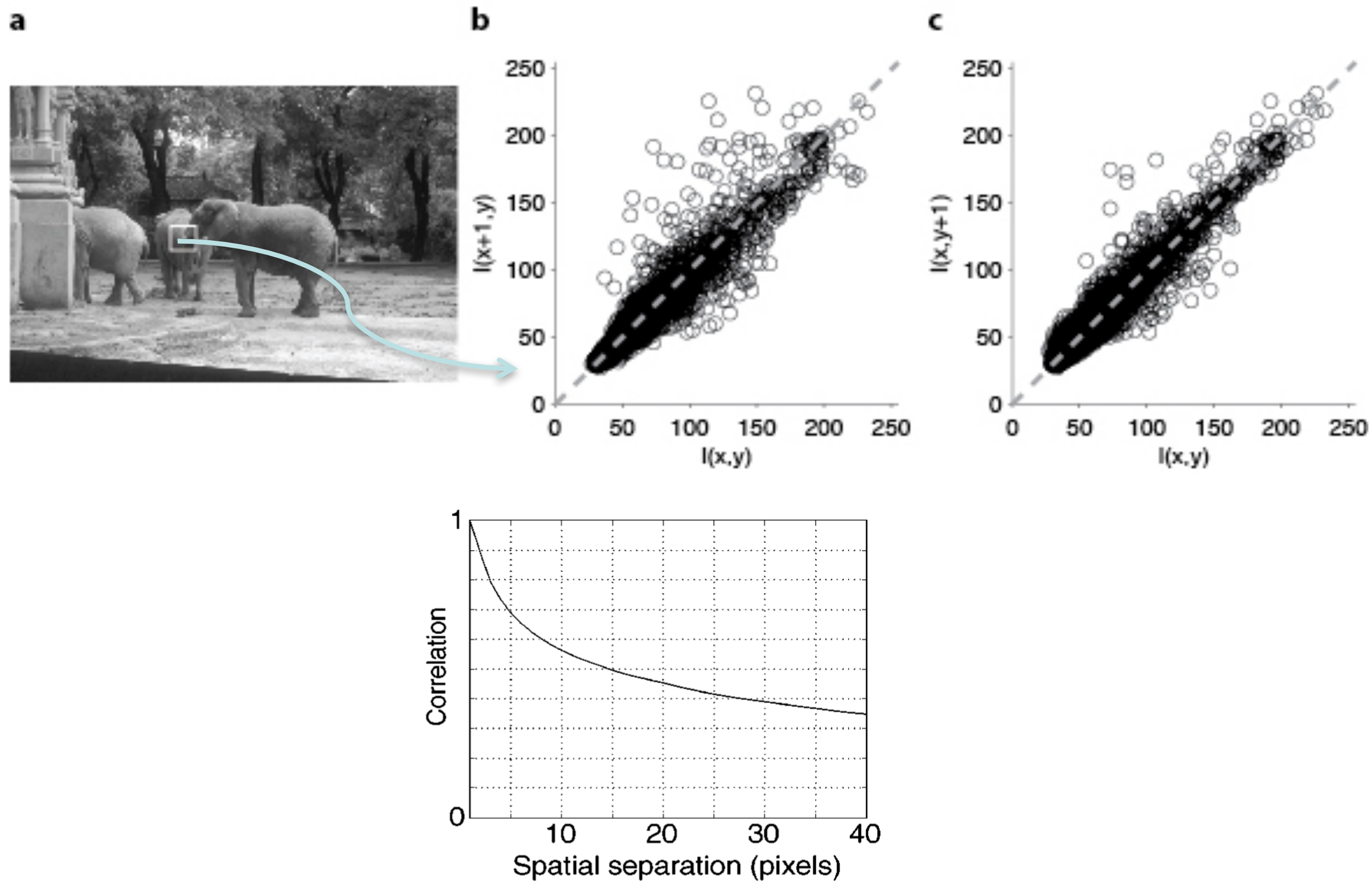
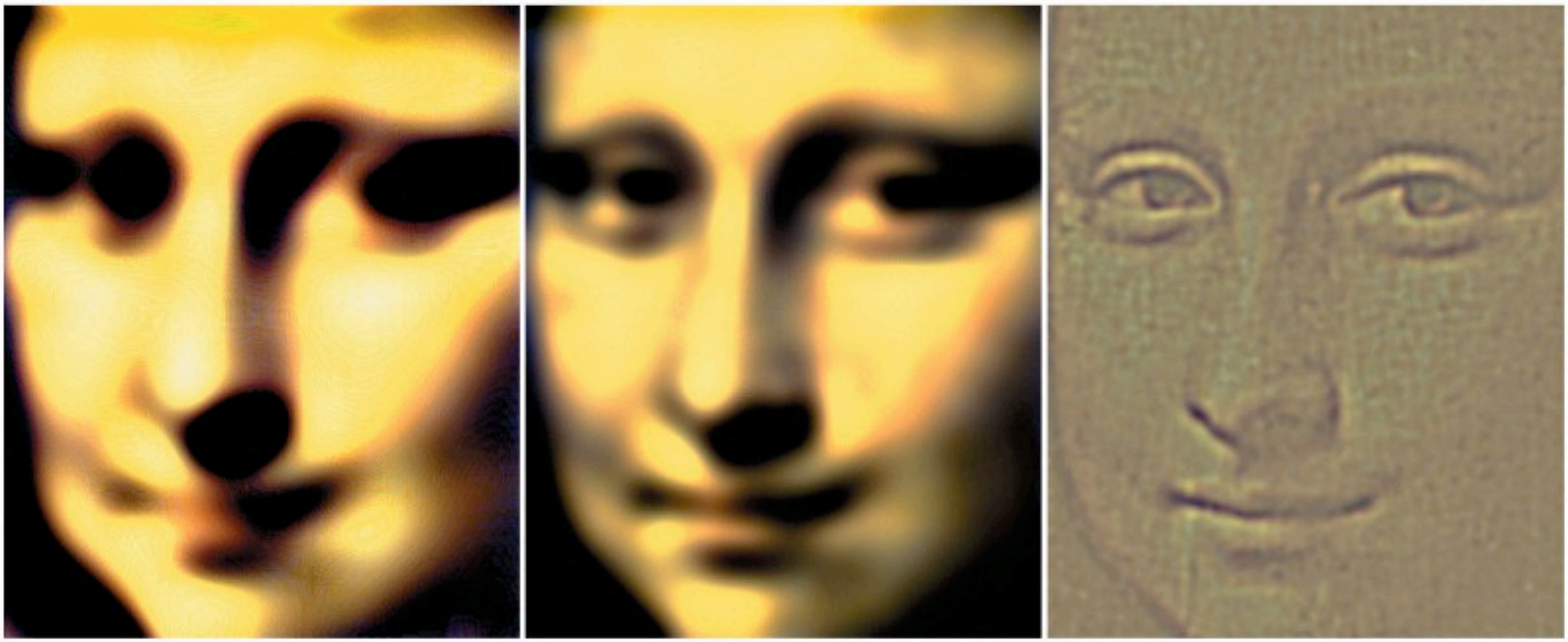


Figure 3 (*a*) Joint distributions of image pixel intensities separated by three different distances. (*b*) Autocorrelation function.

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 ~ $1/f^2$

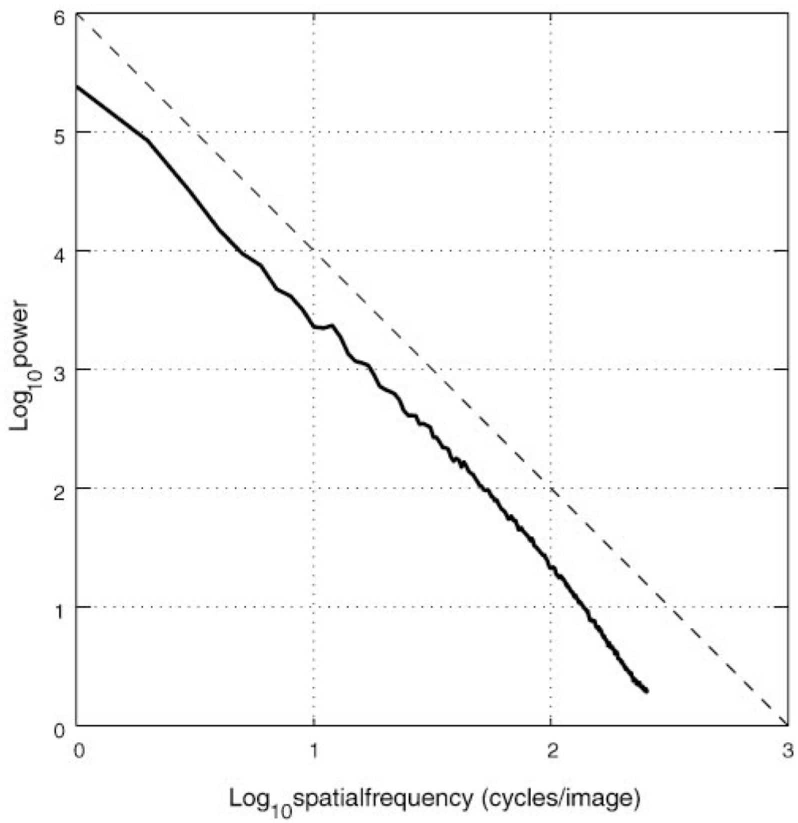


Figure 4 Power spectrum of a natural image (solid line) averaged over all orientations, compared with $1/f^2$ (dashed line).

Simoncelli and Olshausen 2001

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

Note: Scale invariance

$$w' \rightarrow aw$$

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

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

Natural image statistics
There are also strong correlations in time

The visual input is largely static, except for:

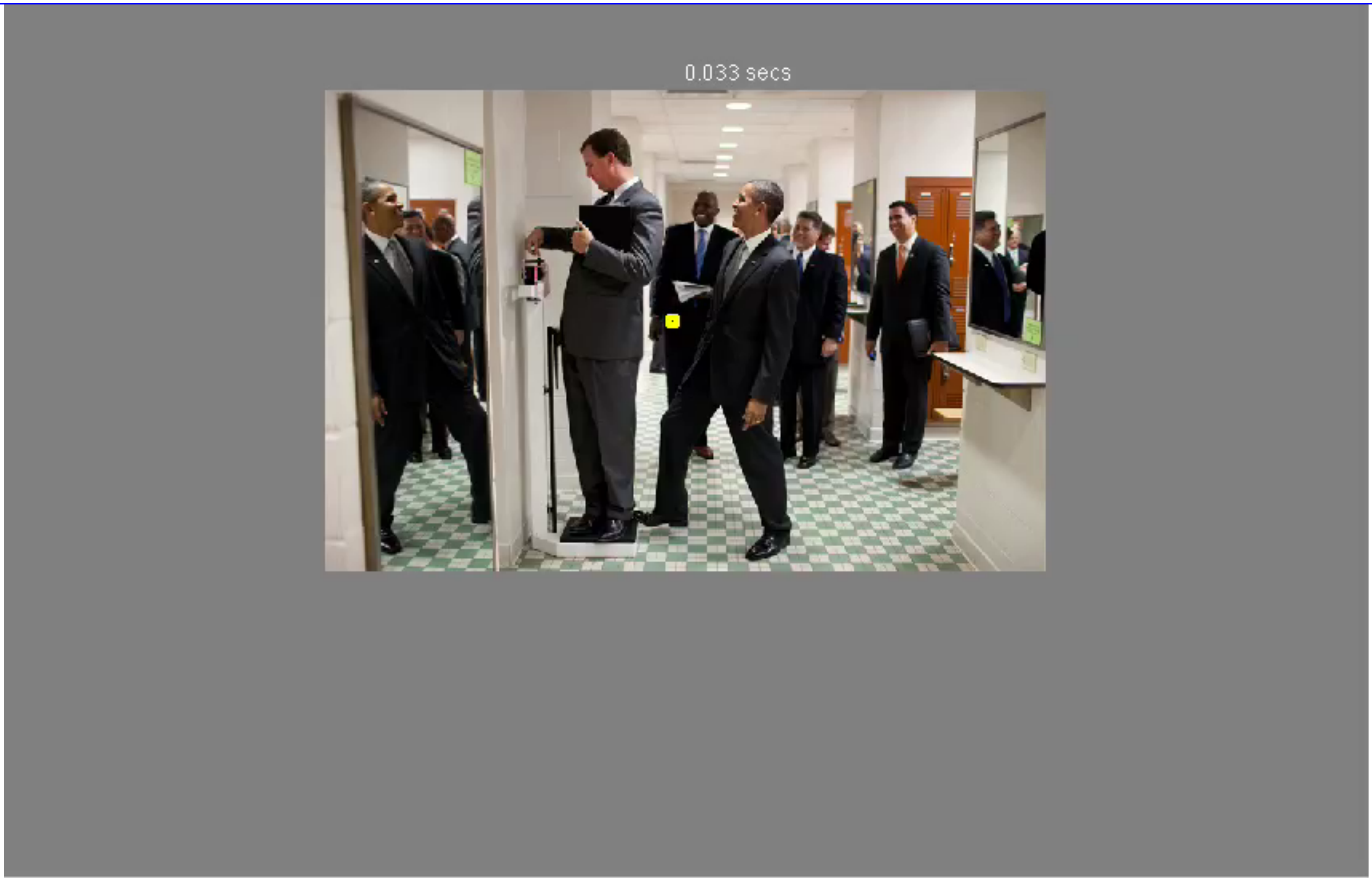
- External object movements
- Head movements
- Eye movements

The visual image is largely static over hundreds of milliseconds

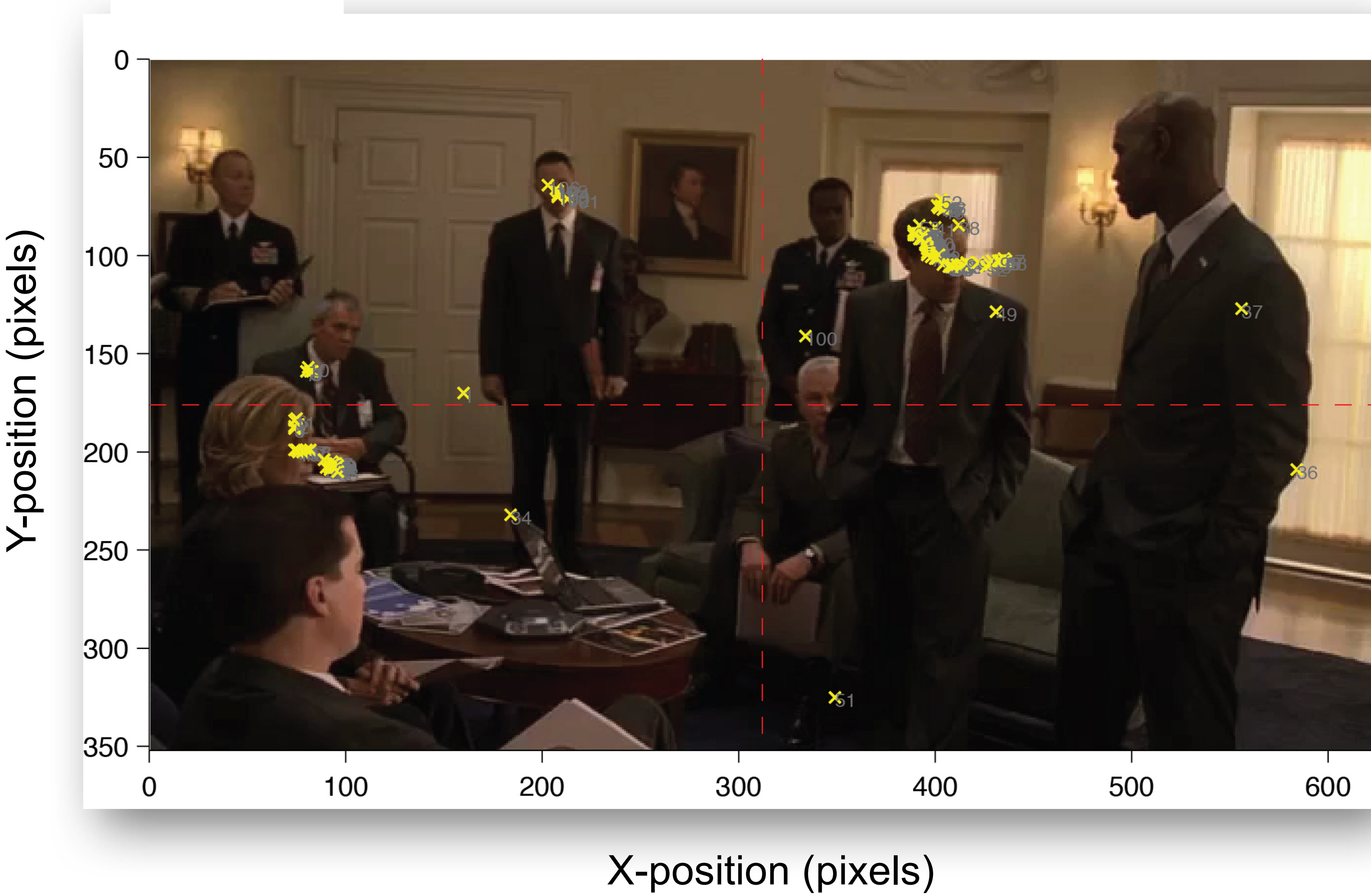
Silent Reading 225-250 ms fixation, 2 degrees saccade size (8-9 letters)
Scene Perception 260-330 ms fixation, 4 degrees saccade

"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

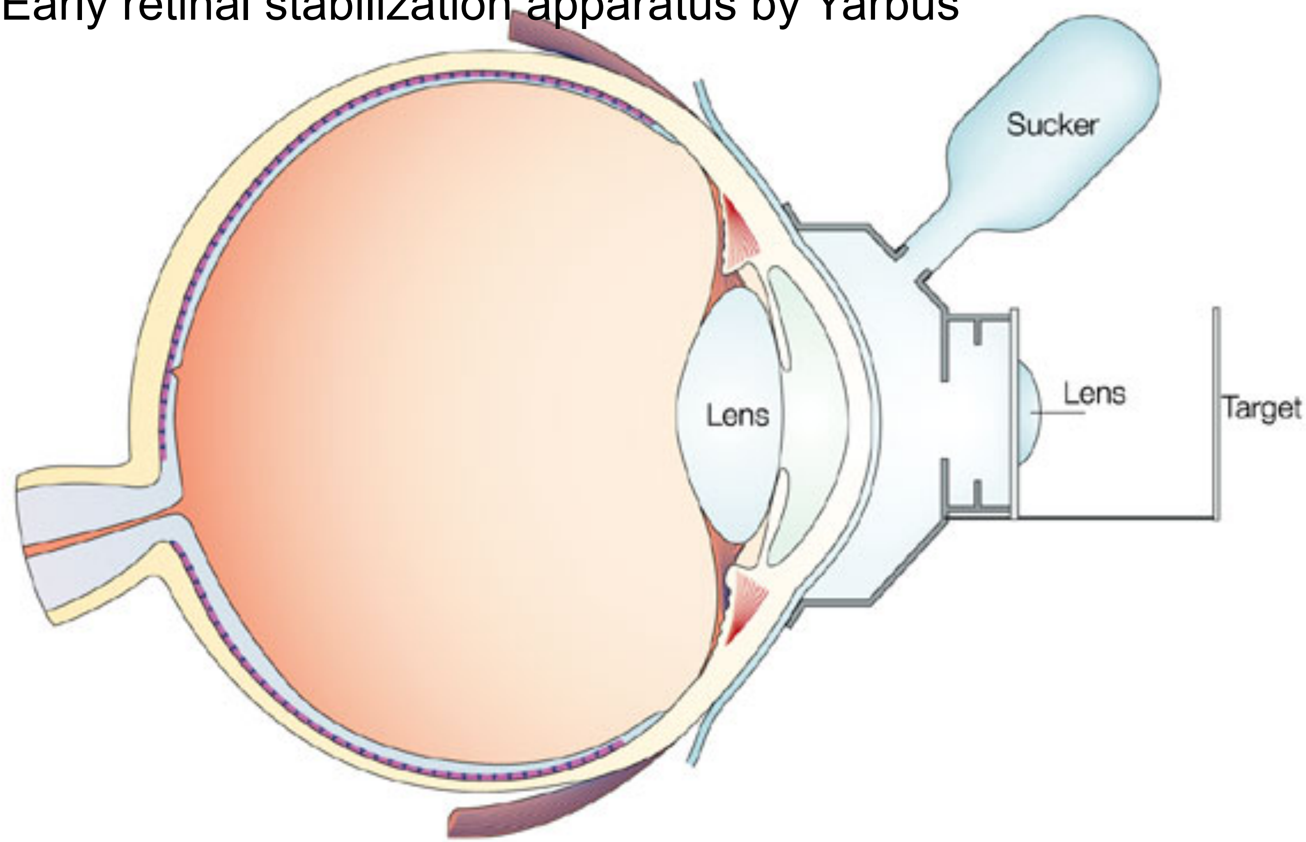


Example pattern of fixations during a movie



Microsaccades are important for perception

Early retinal stabilization apparatus by Yarbus



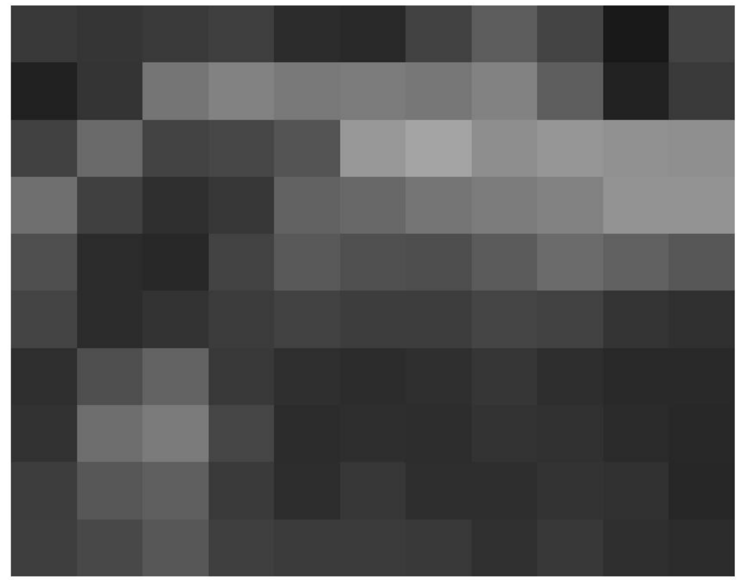
Martinez-Conde, Macknik, & Hubel, 2004

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).

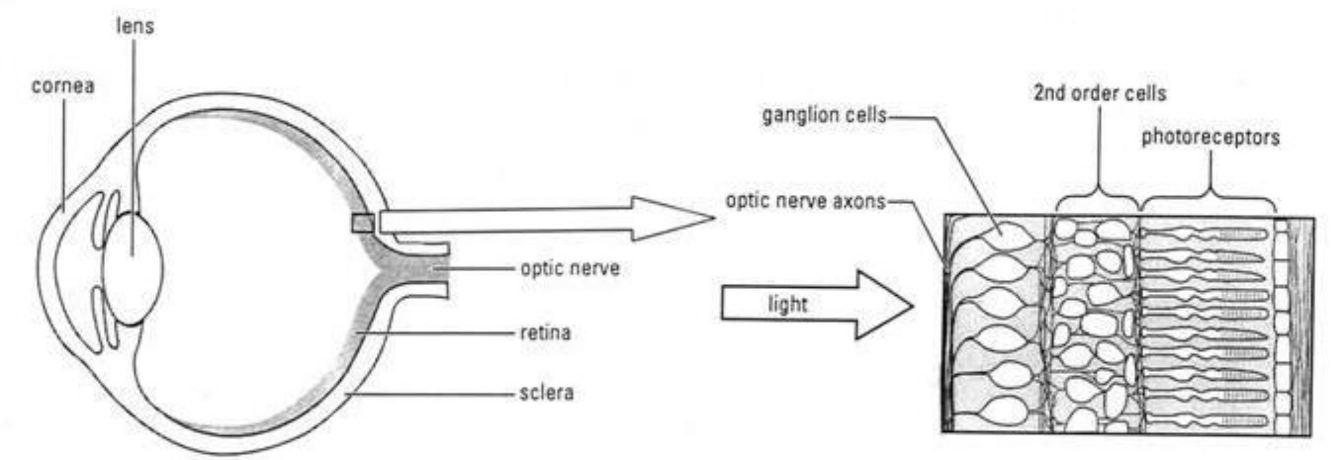
These microsaccades may be critical for counteracting perceptual fading (McCamy, 2012)

An image as a collection of pixels

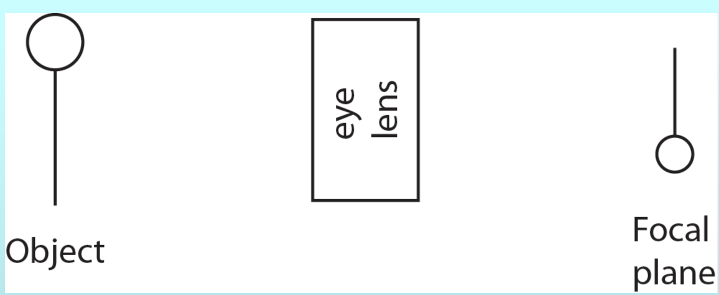


57	53	58	63	44	41	66	93	68	25	67
33	52	117	130	121	124	119	130	94	34	58
65	106	67	71	84	152	164	142	150	145	143
111	64	47	55	98	104	117	124	130	147	147
79	44	40	67	89	80	78	91	107	97	87
68	44	51	60	66	61	61	69	66	52	48
47	79	99	57	47	44	47	54	46	41	41
50	110	123	70	44	46	45	51	49	43	40
61	87	95	58	45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

The image is focused onto the retina



Images are inverted by the eye

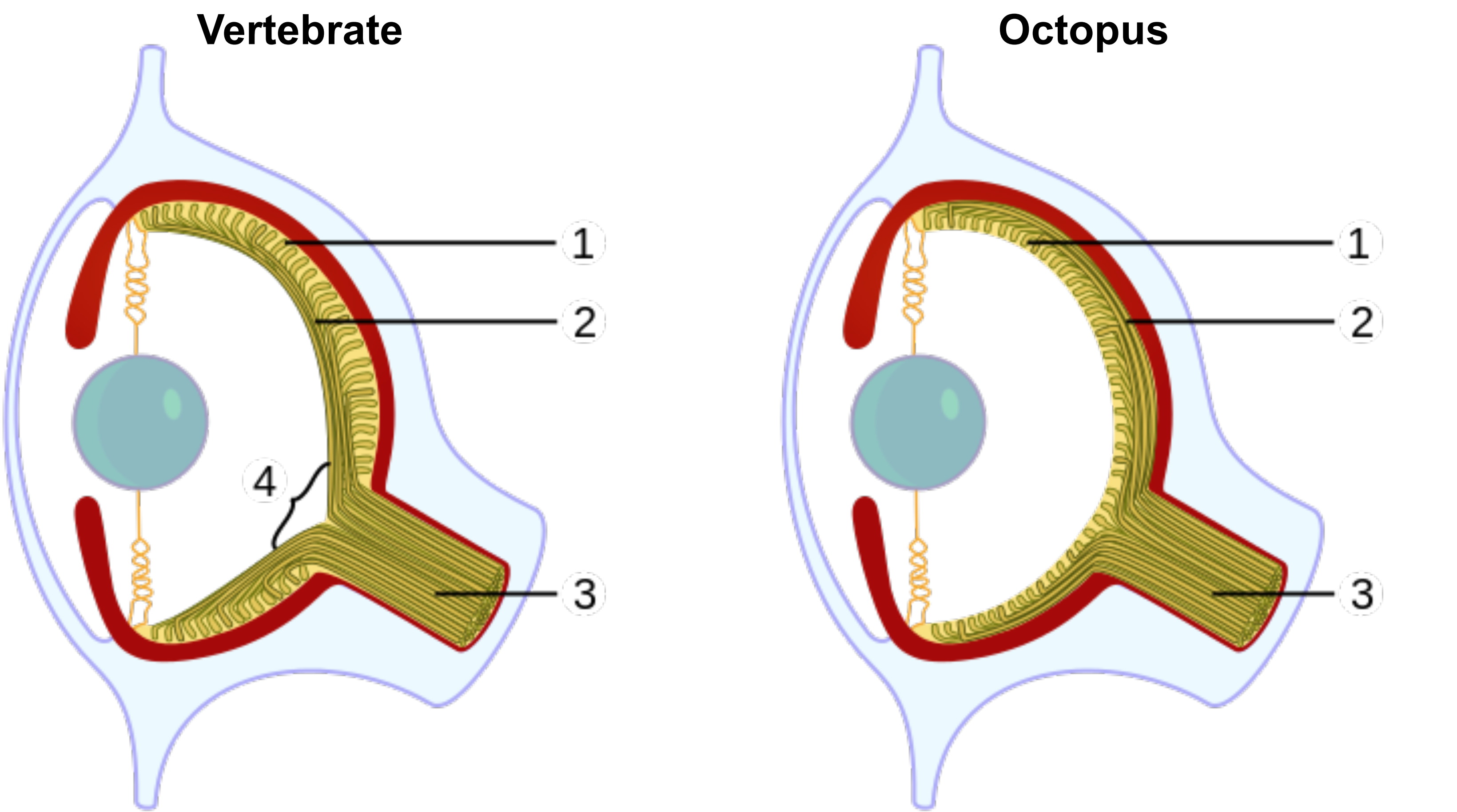


Why don't we see everything upside down?

Perception can adapt to reversing the image:
SOME PRELIMINARY EXPERIMENTS ON VISION WITHOUT INVERSION OF THE RETINAL IMAGE.

GEORGE M. STRATTON, 1896

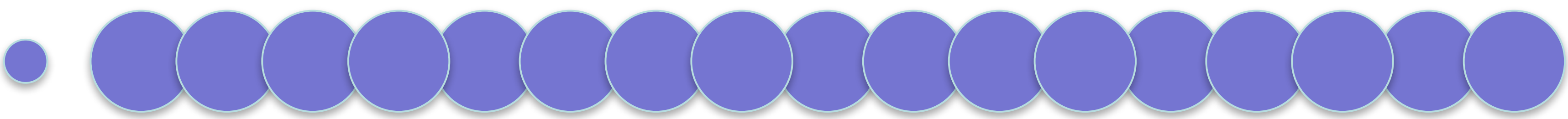
Evolution of the retina



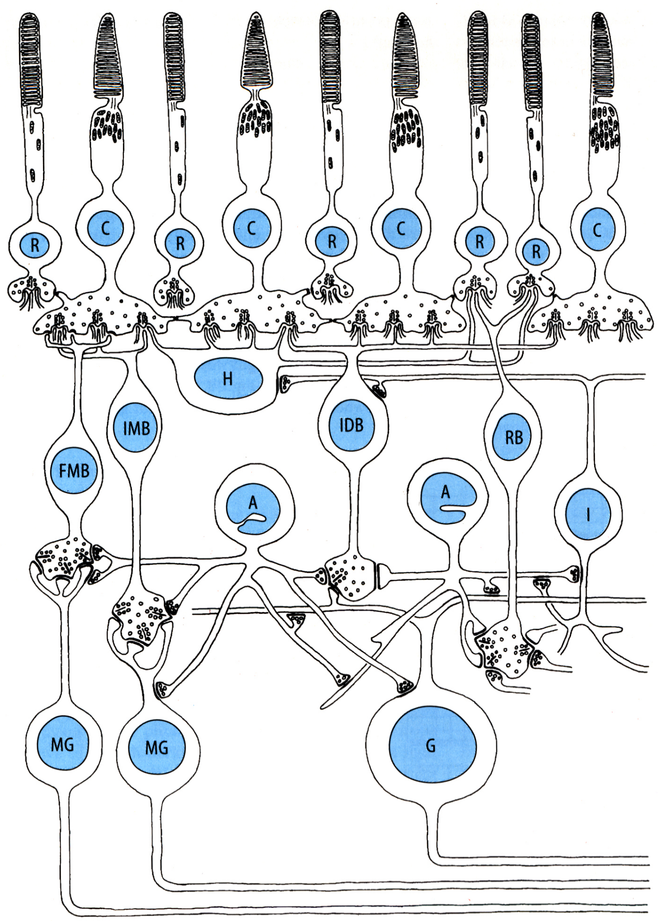
4. Blind spot is only present in vertebrates

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

“Seeing” the blind spot



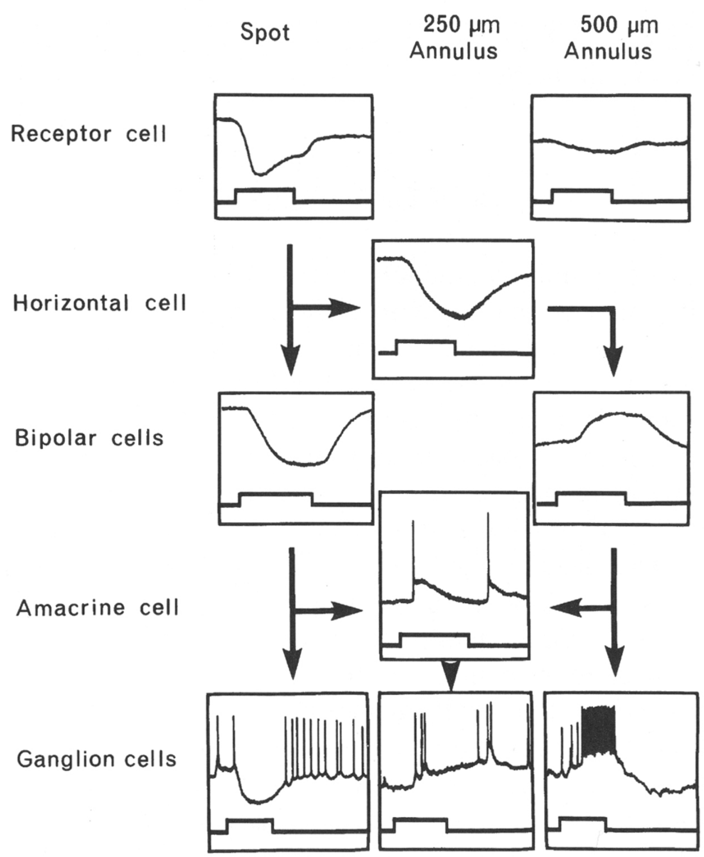
The retina: An amazingly beautiful circuitry composed of many different cell types



- ~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

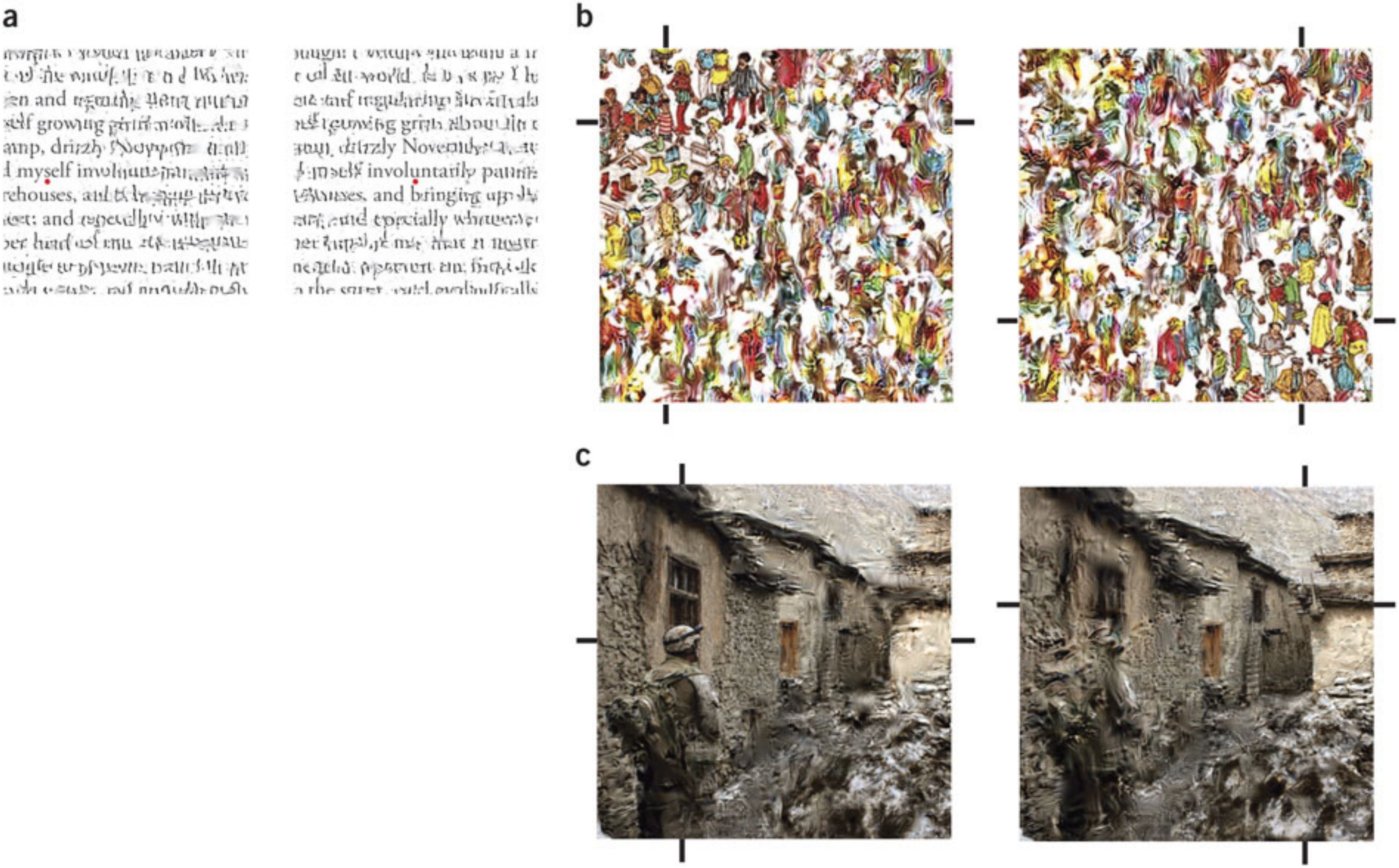


- Photoreceptors transduce incoming light input into electrical signals
- Rod to bipolar convergence increases rod-pathway sensitivity
- Cones, rods, horizontal and bipolar cells are non-spiking neurons
- Many different types of amacrine cells (some of which fire action potentials)
- Retinal ganglion cells fire action potentials and carry the output signals

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

Rods see largely in grayscale

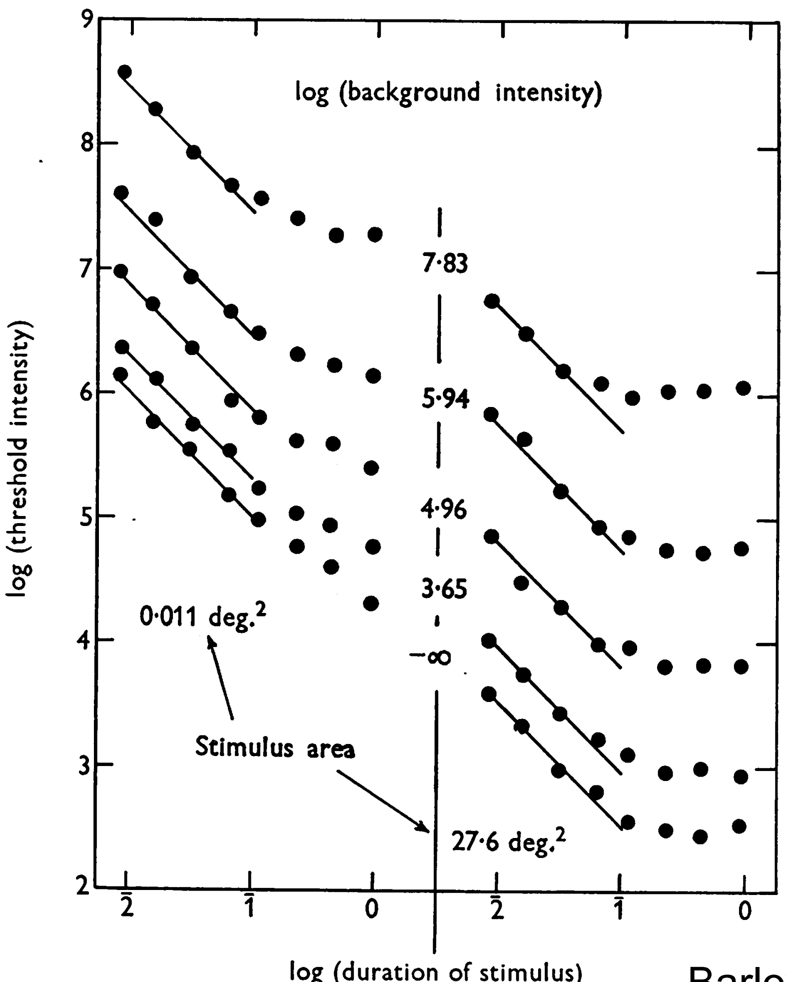
There is much more detail at the *fovea*



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


Sample 1 or 2?

The retina has a huge dynamic range



Barlow, 1958

Vision works well in moonless nights and in bright sunlight, a huge range of light intensity

There are several mechanisms that support this adaptation:

Changing pupil size

Reliance on rods vs. cones

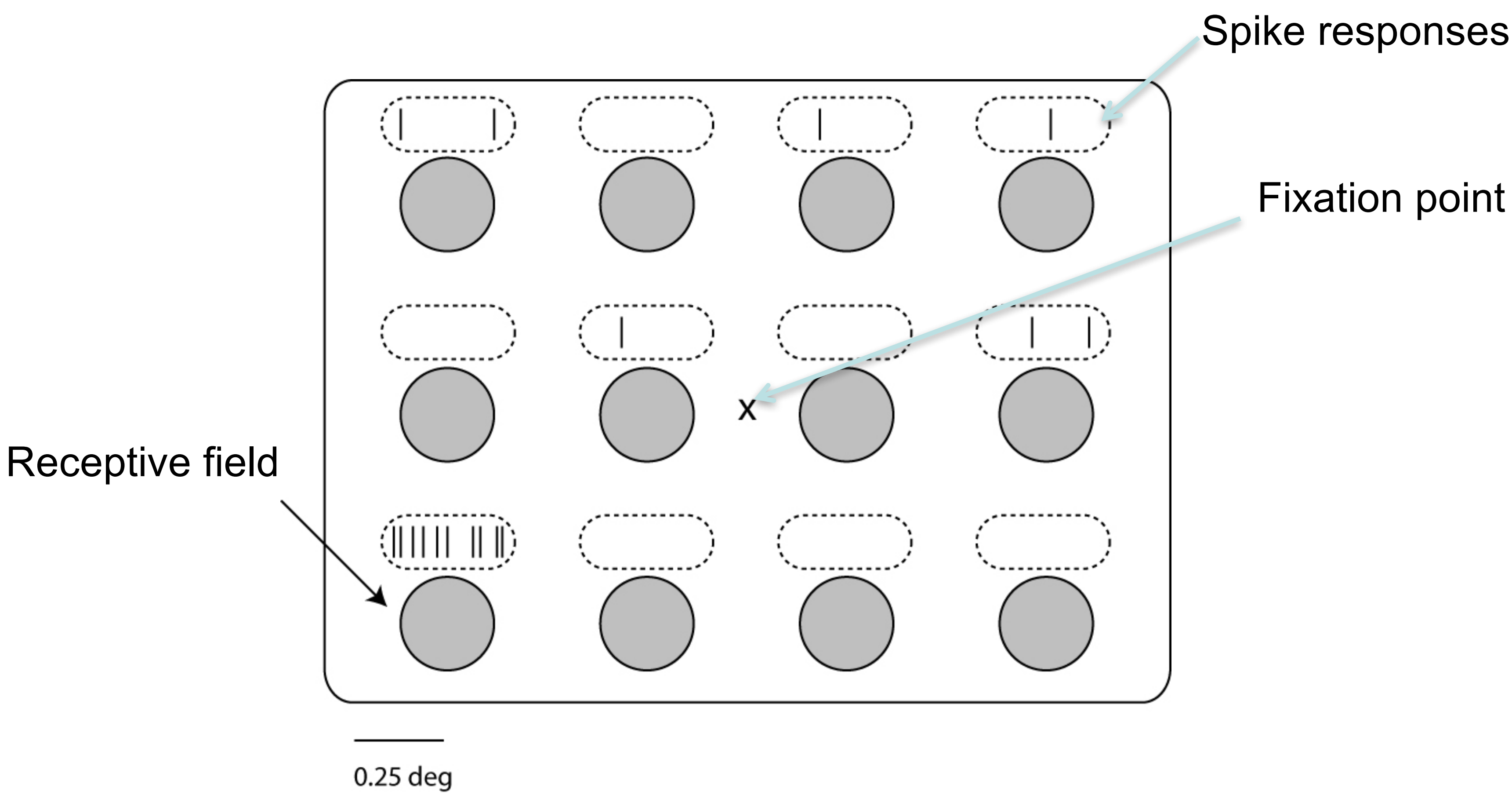
Photopigment bleaching

Feedback from horizontal cells to photoreceptors

These mechanisms are relatively slow

The receptive field

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

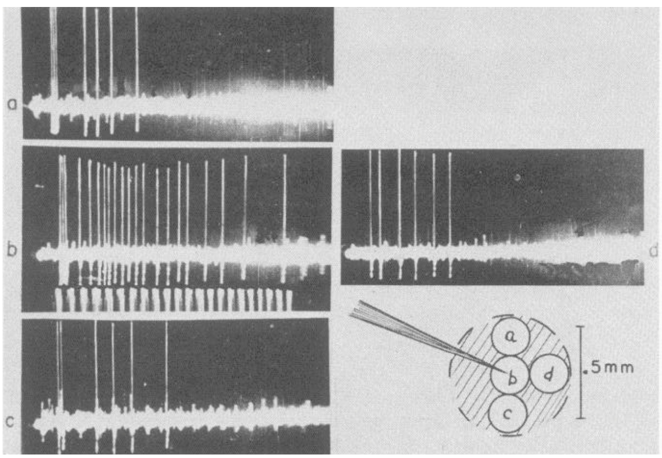


This cartoon neuron responds only when a flash of light appears in the periphery, in the lower left quadrant

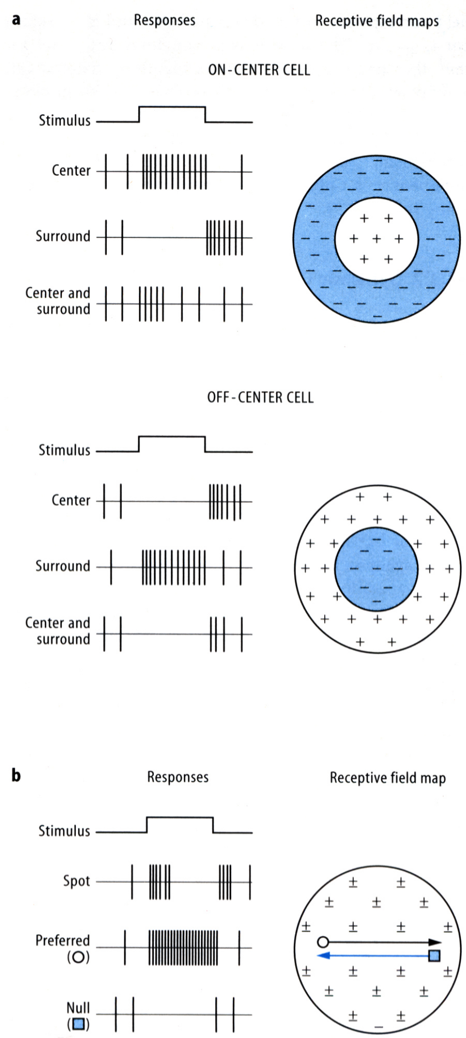
Blumberg and Kreiman, 2010

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

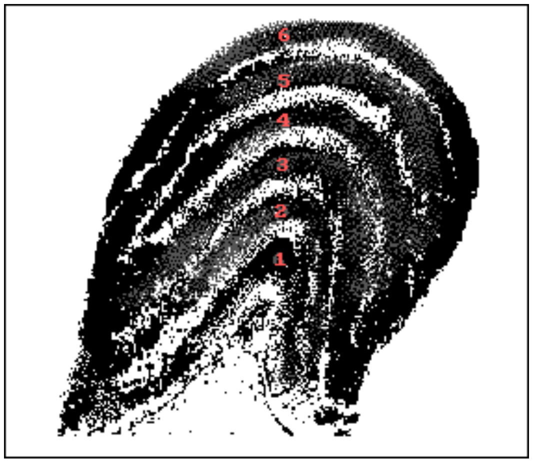
Diversity of retinal ganglion cells

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

To cortex, through the thalamus



The lateral geniculate nucleus (LGN) is the main visual part of the thalamus:

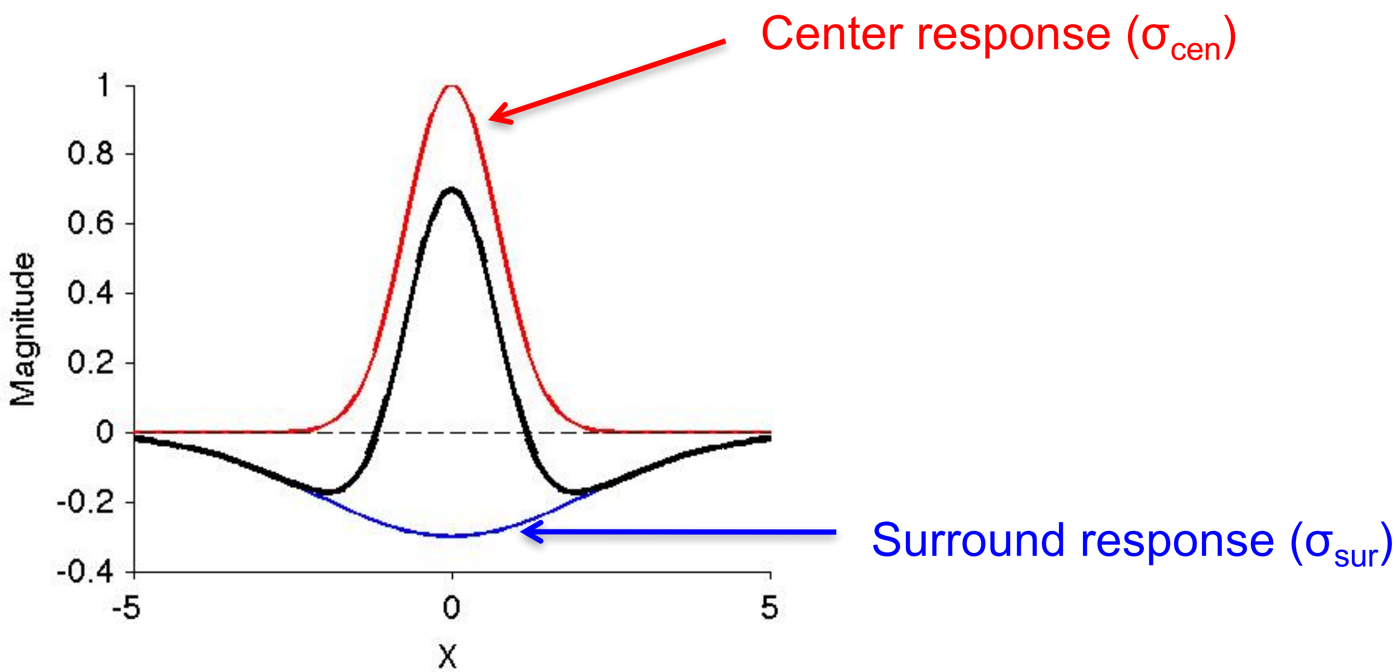
- •6 layers, contralateral visual hemifield
- •Layers 2, 3 and 5 receive ipsilateral eye's input
- •Layers 1, 4 and 6 receive contralateral eye's input
- •Layers 1-2: *magnocellular* layers that receive input from parasol ganglion cells
- •Layers 3-6: *parvocelluar* layers that receive input from midget ganglion cells
- •Between the layers: *koniocellular* layers that receive input from bistratified retinal ganglion cells
- •Right and left visual hemifields are separate in the LGN
- •Right and left eyes are separate in the LGN
- •The visual field is represented multiple times in the LGN
- •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), Foundations of Vision. Sinauer Books

Difference of Gaussians

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



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

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

Neurons respond with transient bursts of activity

Dynamic receptive fields in the retina/LGN

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

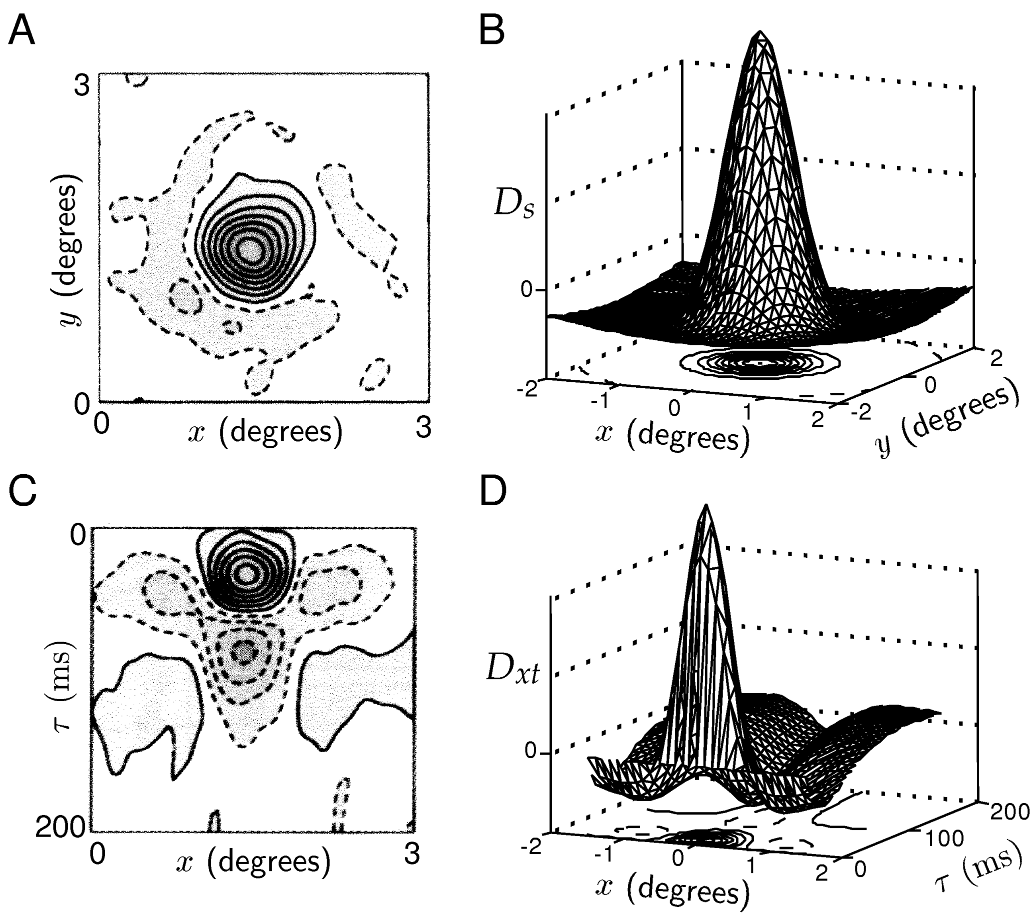
$$D_{cen}(t) = \alpha_{cen}^2 t\exp[-\alpha_{cen}t] - \beta_{cen}^2 t\exp[-\beta_{cen}t]$$

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

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

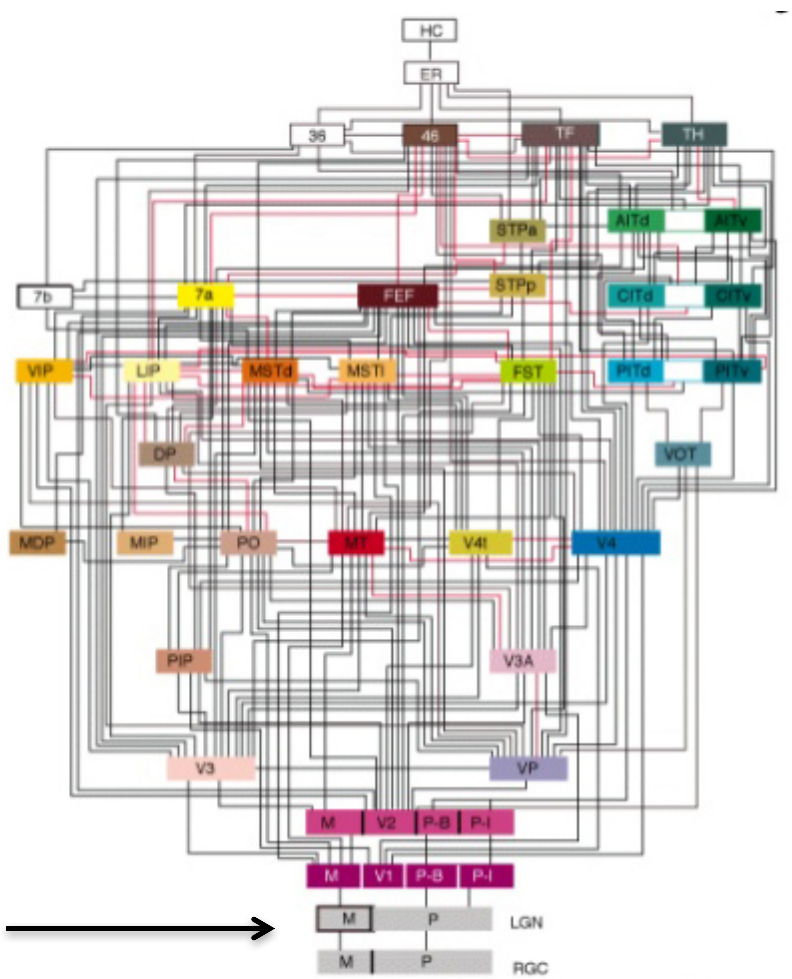
Subcortical visual pathways

Retinal projections

- Lateral geniculate nucleus (LGN) – Thalamus
- 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 but not after lesions to V1 (Gross 1994 for historical overview).

Visual system circuitry



You are here

Felleman and Van Essen. *Cerebral Cortex* 1991

Further reading

Further reading

- Class notes: http://tinyurl.com/vision-class
- Wandell B. Foundations of Vision. Sinauer Books 1995.
- Dayan and Abbott. Theoretical Neuroscience. MIT Press 2001.

Some of the original articles cited in class (see lecture notes for full list)

- Simoncelli and Olshausen. Annual Review of Neuroscience 2001
- Dowling J. Scholarpedia 2007.
- Felleman and Van Essen. Cerebral Cortex 1991.
- Blumberg and Kreiman. Journal of Clinical Investigation 2010.
- Kuffler. Journal of Neurophysiology 1953.
- Foldiak. Neural Computation 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.
- Smith, E.C., and Lewicki, M.S. (2006). Efficient auditory coding. Nature 439, 978-982.
- 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.