Early Vision:
from retina to visual cortex

Jonathan Pillow
NEU 502a

Jan 30, 2025

the retina

(“smart” film in your camera)

What does the retina do?

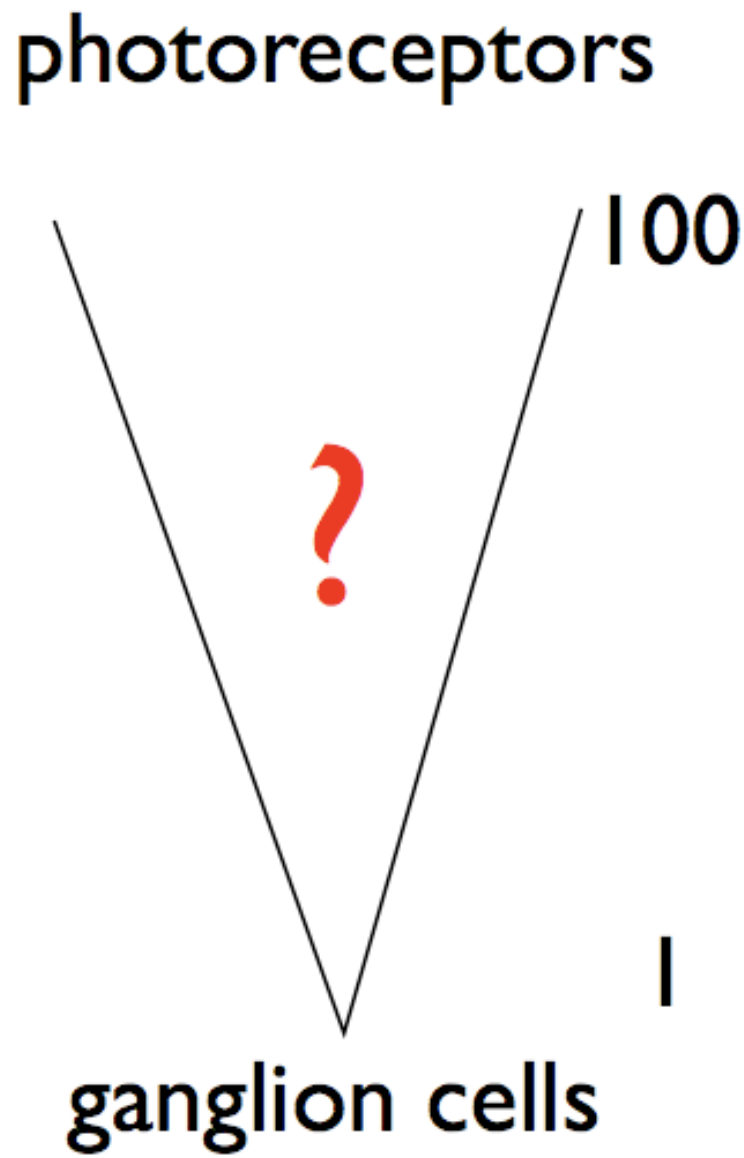
1. Transduction

• Conversion of energy from one form to another (i.e., "light" into "electrical energy")

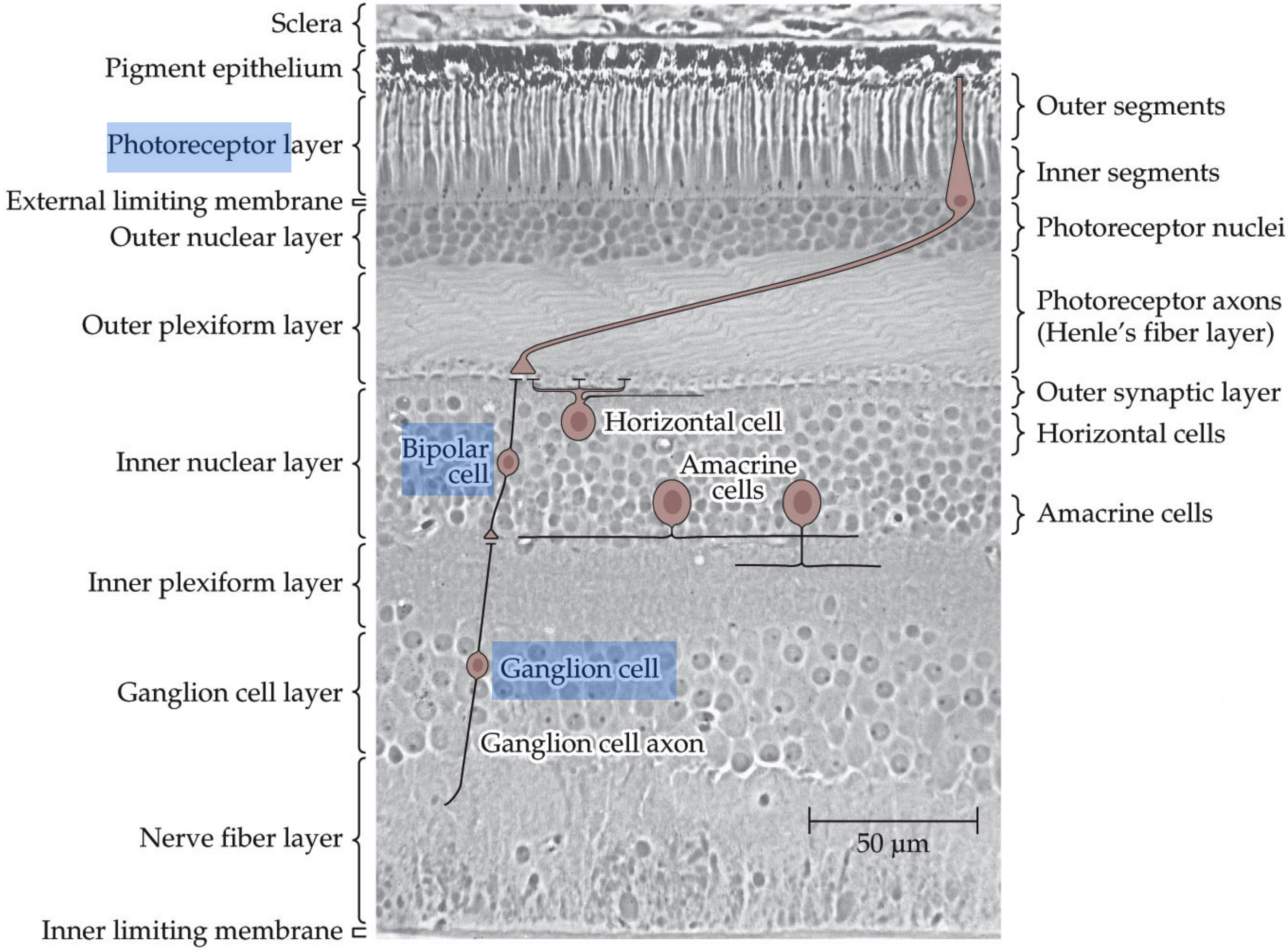
2. Processing

• **Amplification** of very weak signals
  (1-2 photons can be detected!)

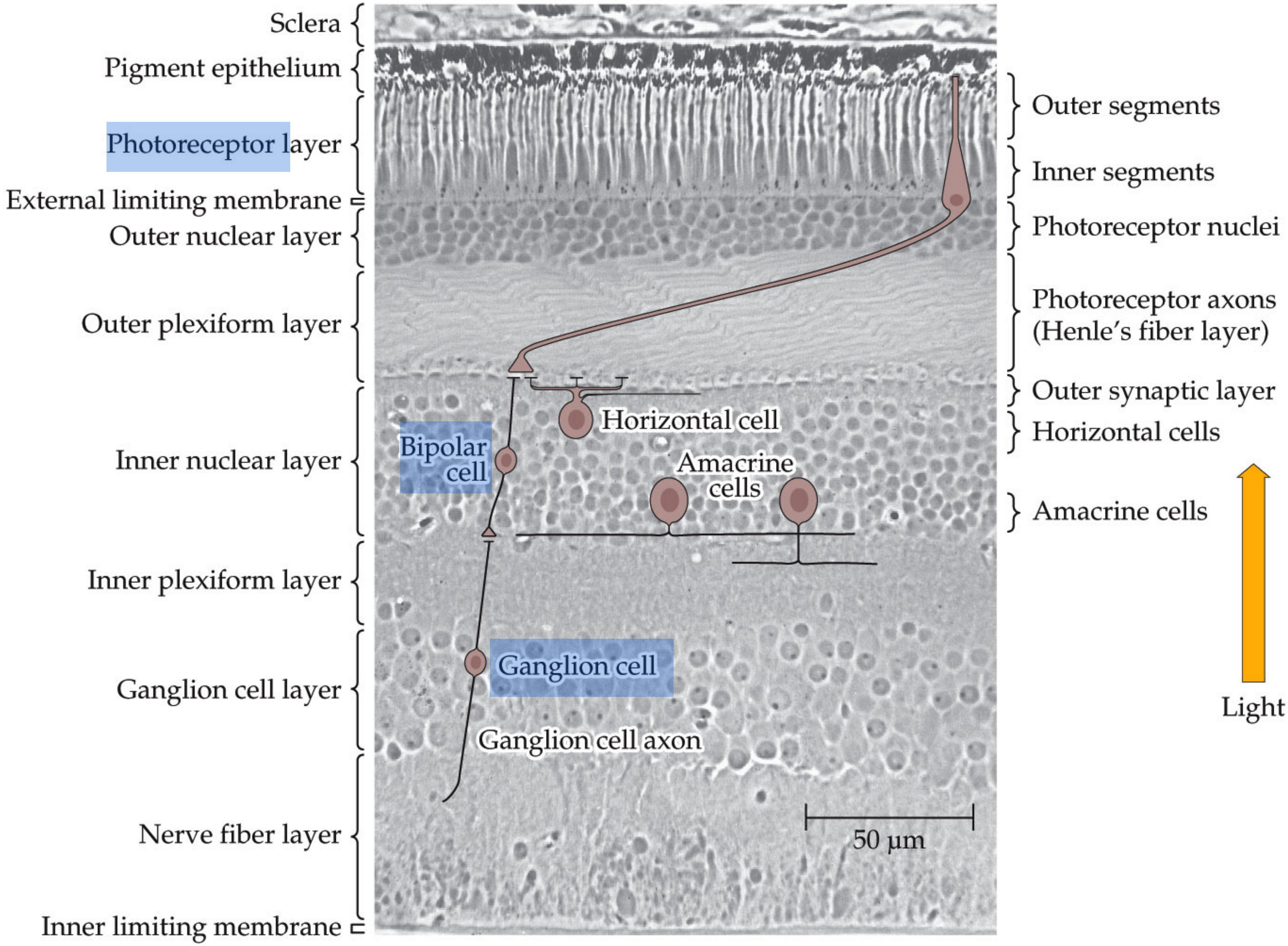
• **Compression** of image into more compact form so that information can be efficiently sent to the brain
  optic nerve = "bottleneck"
  analogy: jpeg compression of images

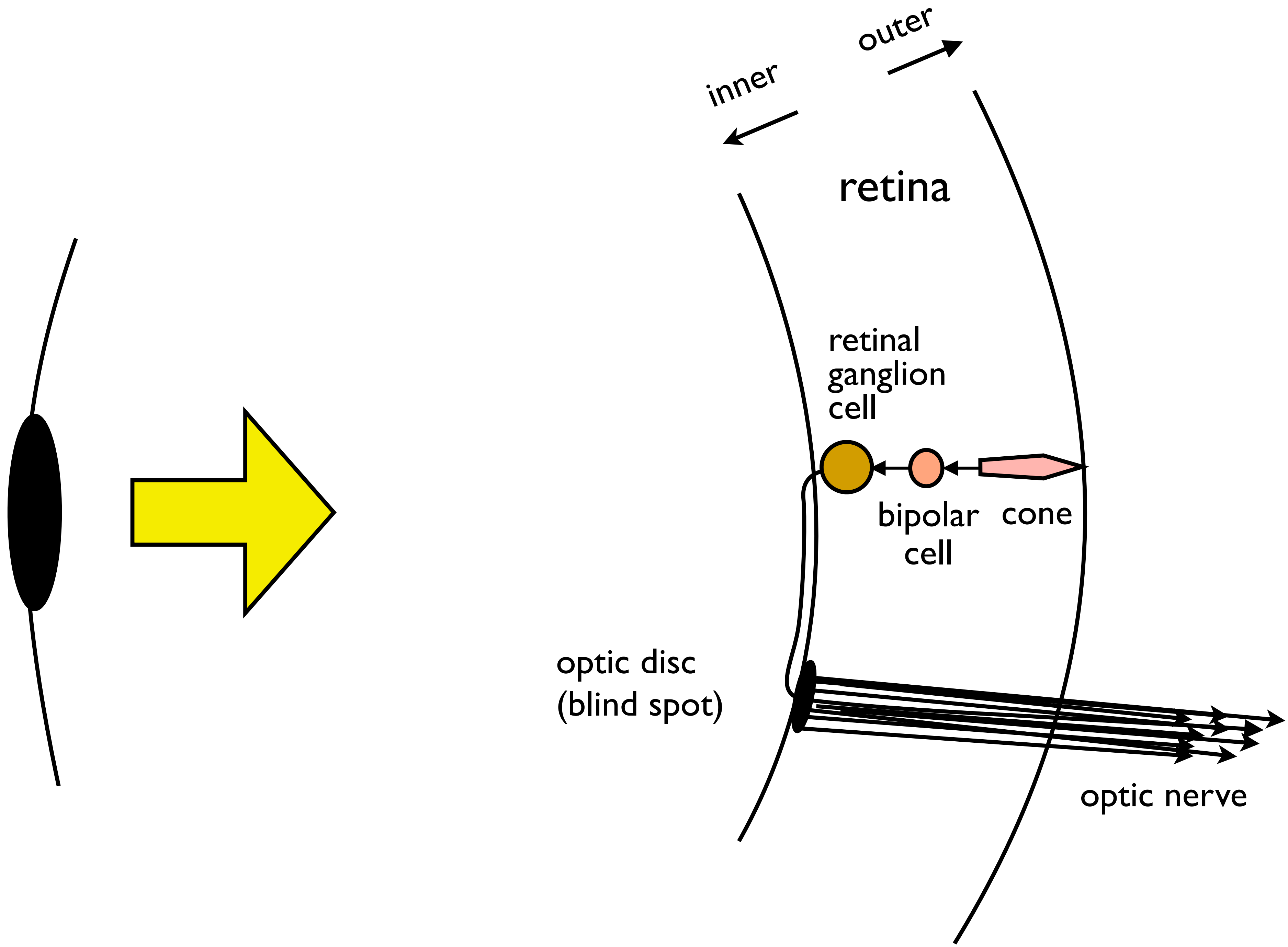

photoreceptors
100
?
1
ganglion cells

Basic anatomy: photomicrograph of the retina



Basic anatomy: photomicrograph of the retina




outer
inner
retina
retinal
ganglion
cell
bipolar
cell
cone
optic disc
(blind spot)
optic nerve

What’s crazy about this is that the light has to pass through all the other junk in our eye before getting to photoreceptors!

Cephalopods (squid, octopus): did it right.

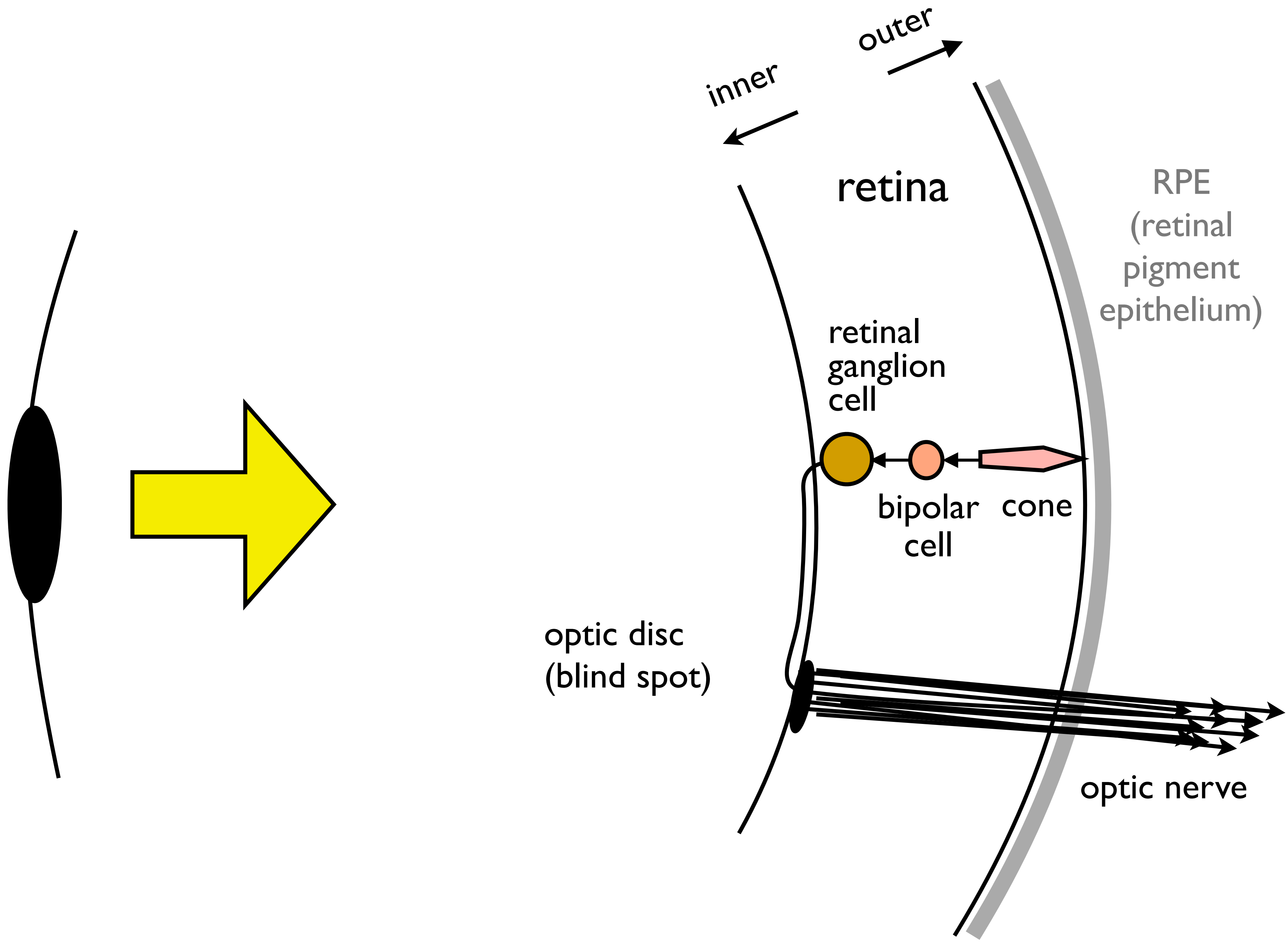
- photoreceptors in innermost layer, no blind spot!

<u>Debate</u>:

1. accident of evolution?

   OR

2. better to have photoreceptors near blood supply?

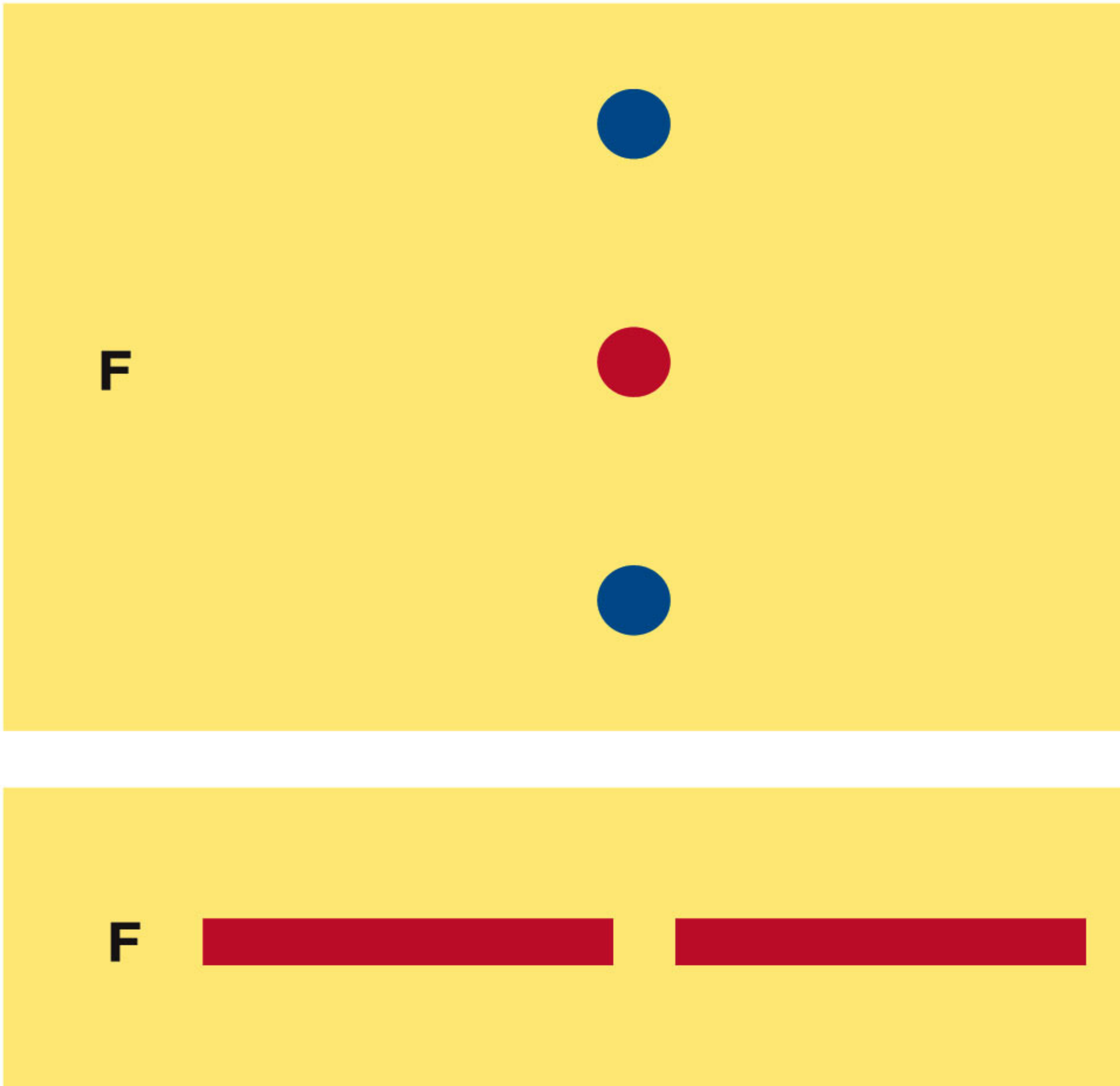


outer
inner
retina
RPE
(retinal
pigment
epithelium)
retinal
ganglion
cell
bipolar
cell
cone
optic disc
(blind spot)
optic nerve

blind spot demo

(a)

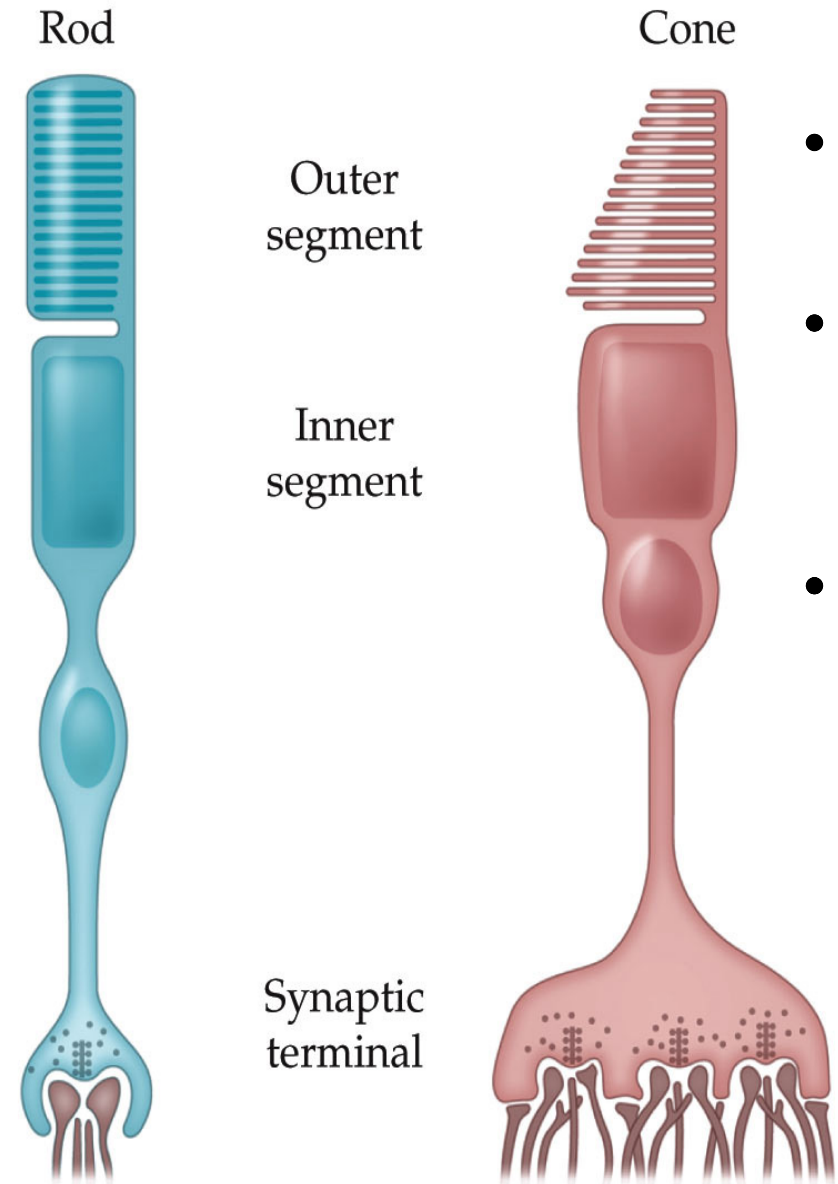
(b)



phototransduction: converting light to electrical signals

rods

- respond in low light ("scotopic")
- only one kind: don't process color
- 90M in humans



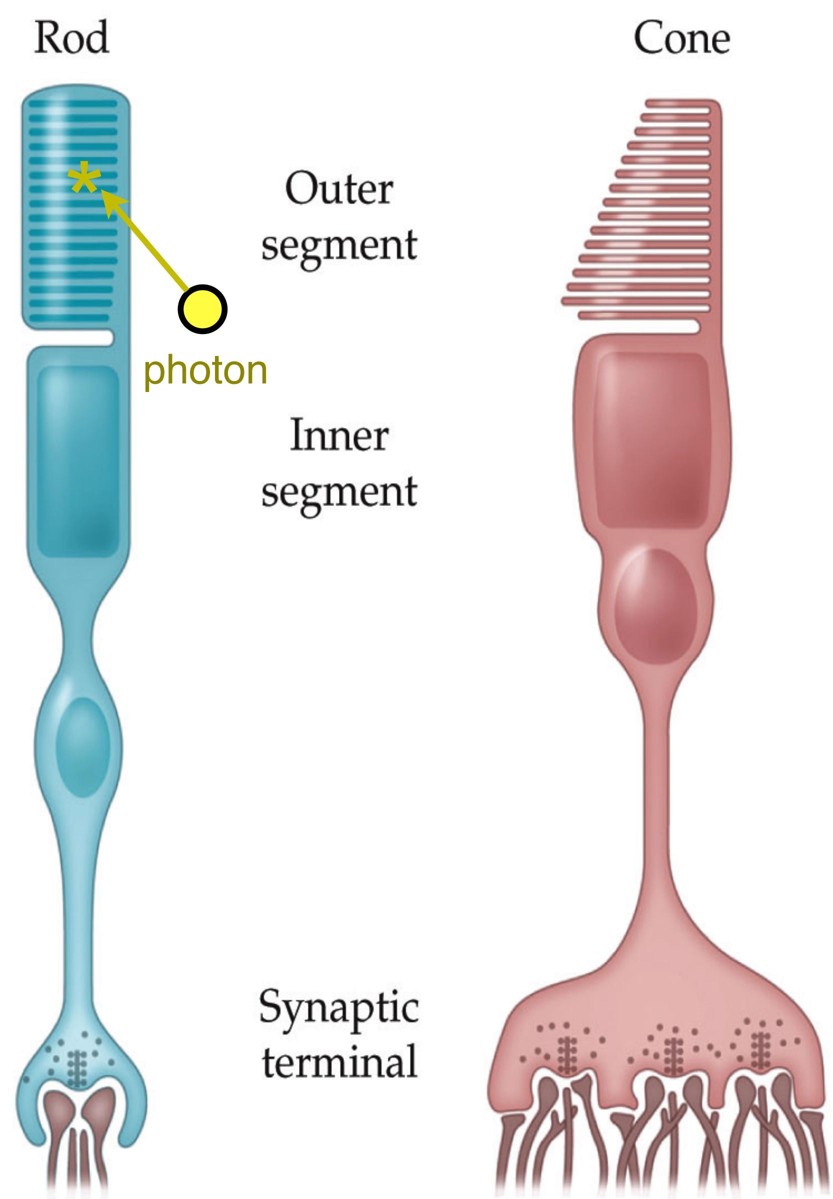
cones

- respond in daylight ("photopic")
- 3 different kinds: responsible for color processing
- 4-5M in humans

phototransduction: converting light to electrical signals

<u>outer segments</u>

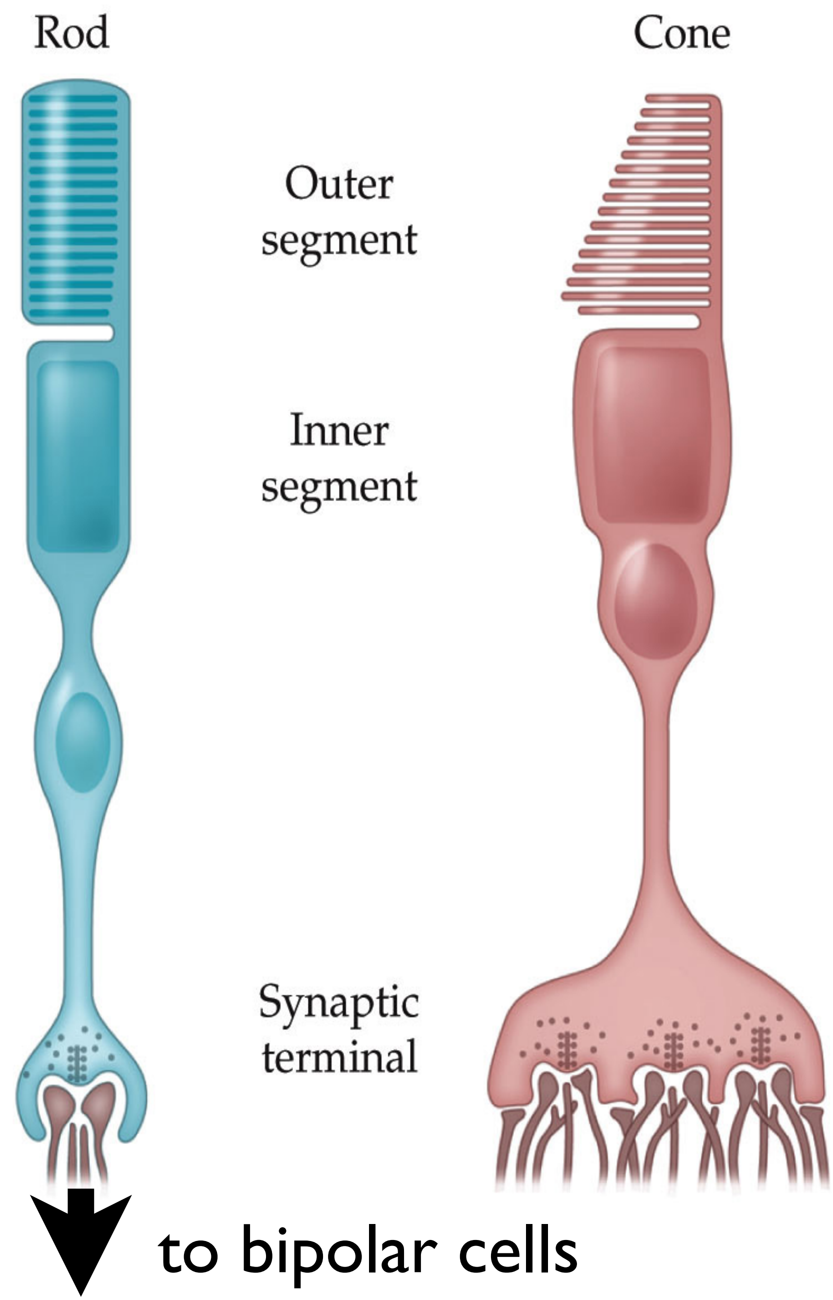
- packed with discs
- discs have **opsins** (proteins that change shape when they absorb a photon - amazing!)



- different opsins sensitive to different wavelengths of light
- **rhodopsin**: opsin in rods
- **photopigment**: general term for molecules that are photosensitive (like opsins)

dark current

- In the dark, membrane channels in rods and cones are open by default (unusual!)
- current flows in continuously
- membrane is *depolarized* (less negative)



- neurotransmitter is released at a high rate

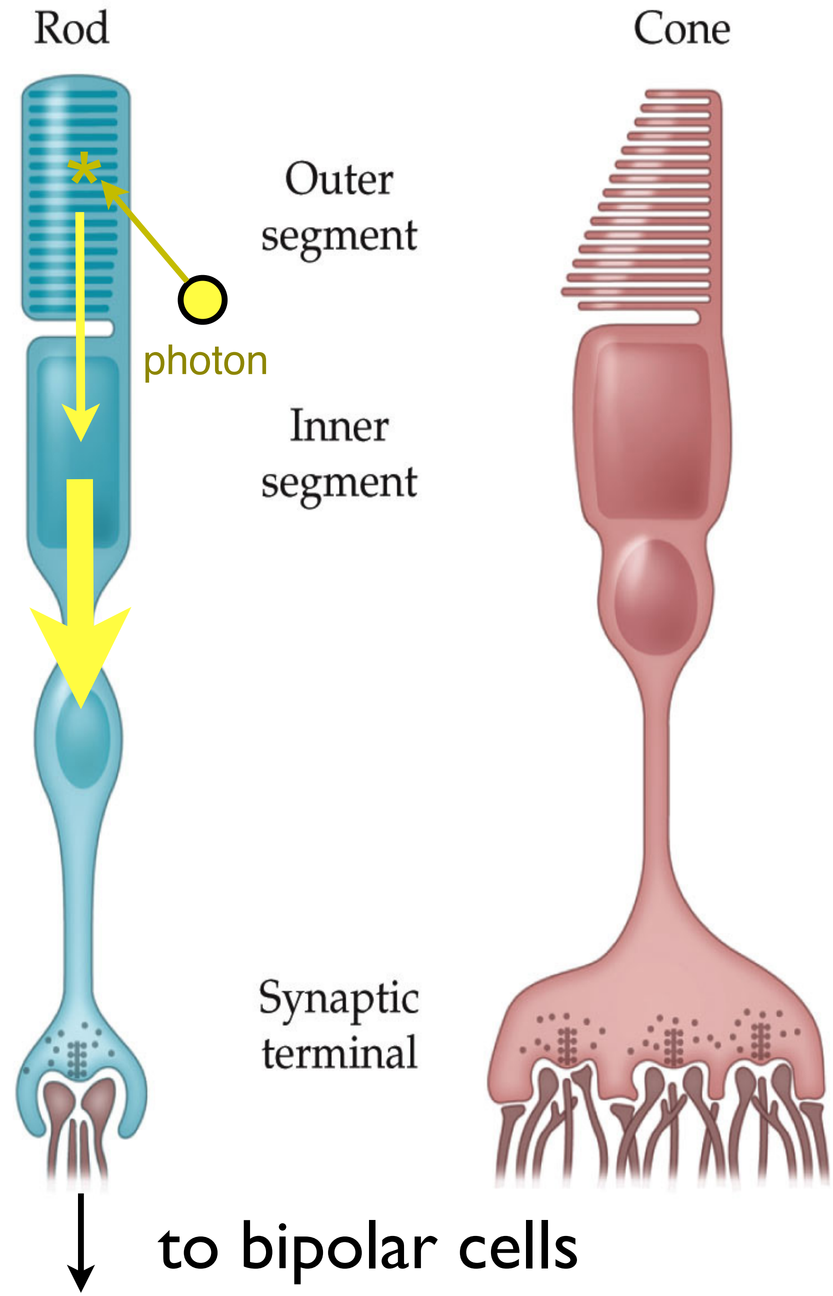
transduction & signal amplification

• photon is absorbed by an opsin

• channels close (dark current turns off)

• membrane becomes *more* polarized (more negative)

• neurotransmitter is released at a lower rate



transduction & signal amplification

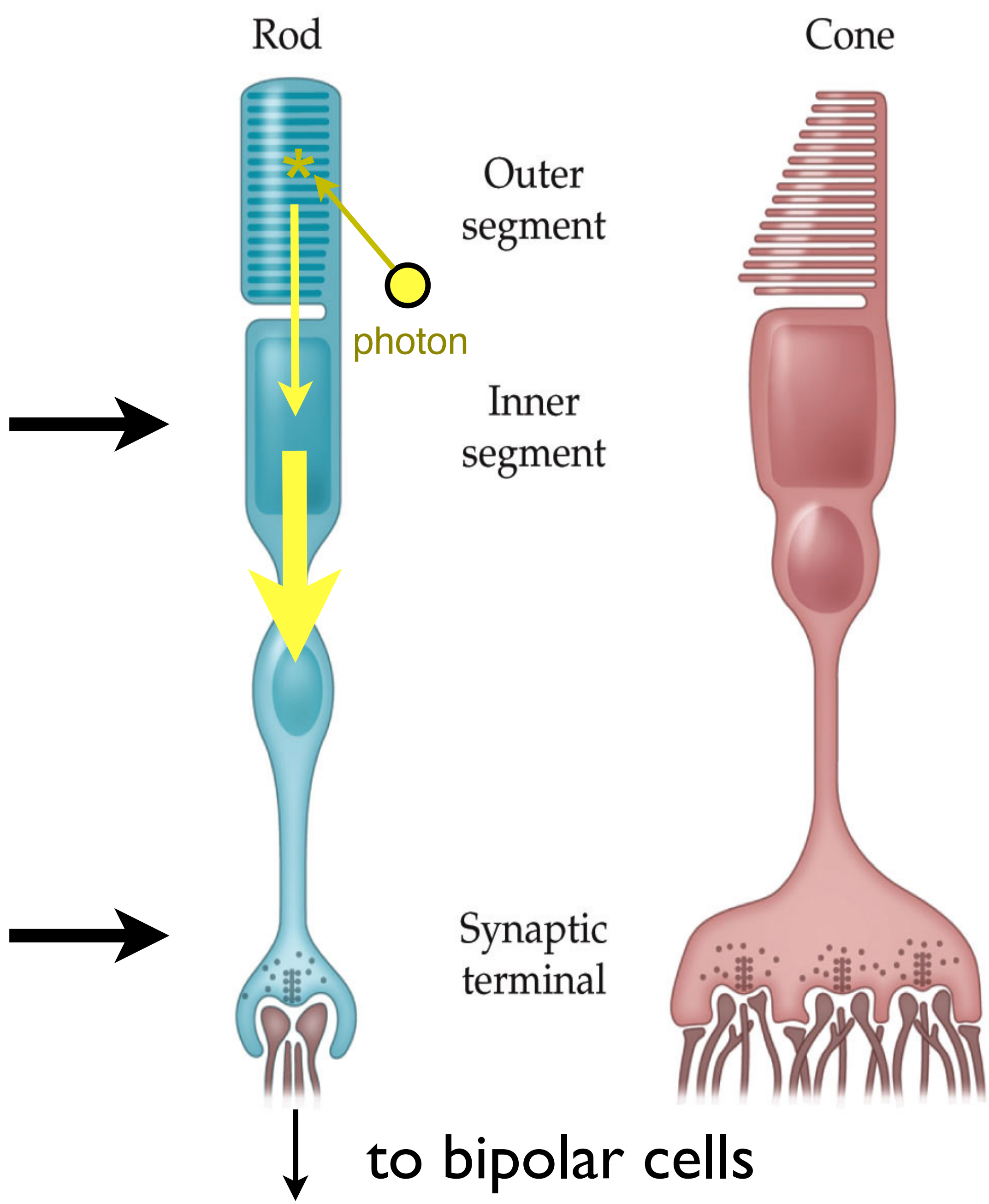
<u>inner segments</u>

machinery for amplifying signals from outer segment

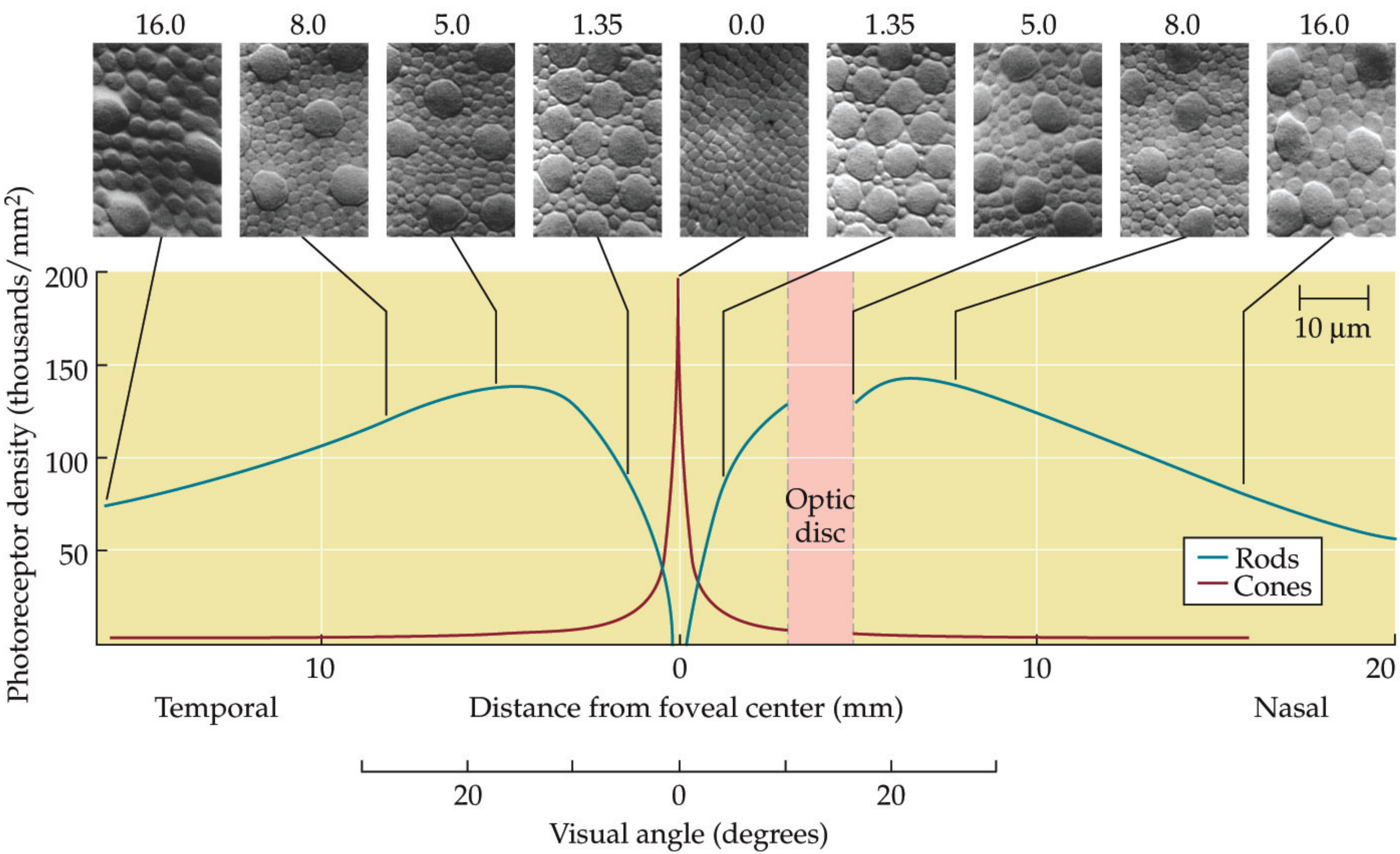
<u>neurotransmitter release</u>

graded potential

(*not* spikes!)



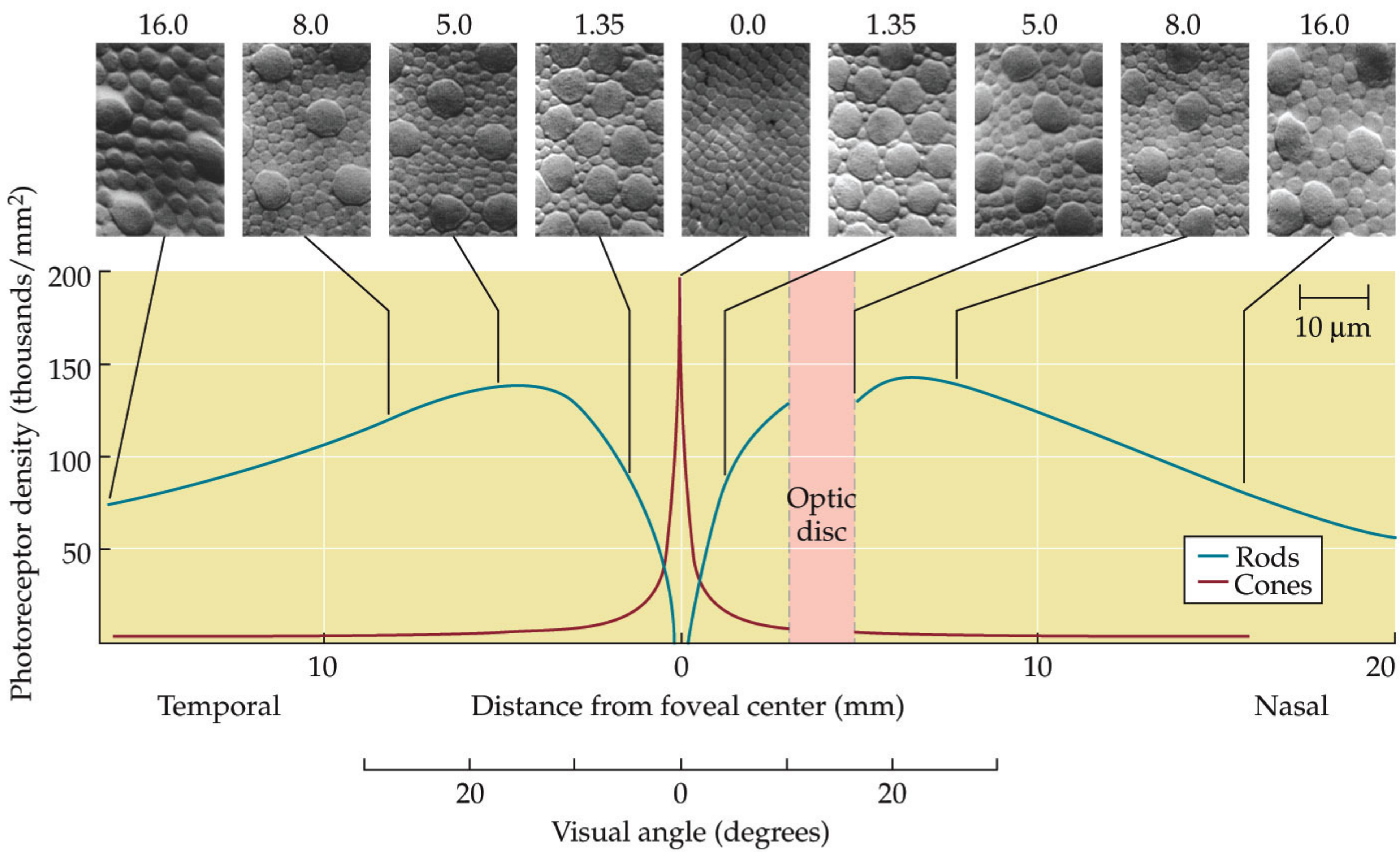
Photoreceptors: not evenly distributed across the retina



- fovea: mostly cones
- periphery: mostly rods

Q: what are the implications of this?

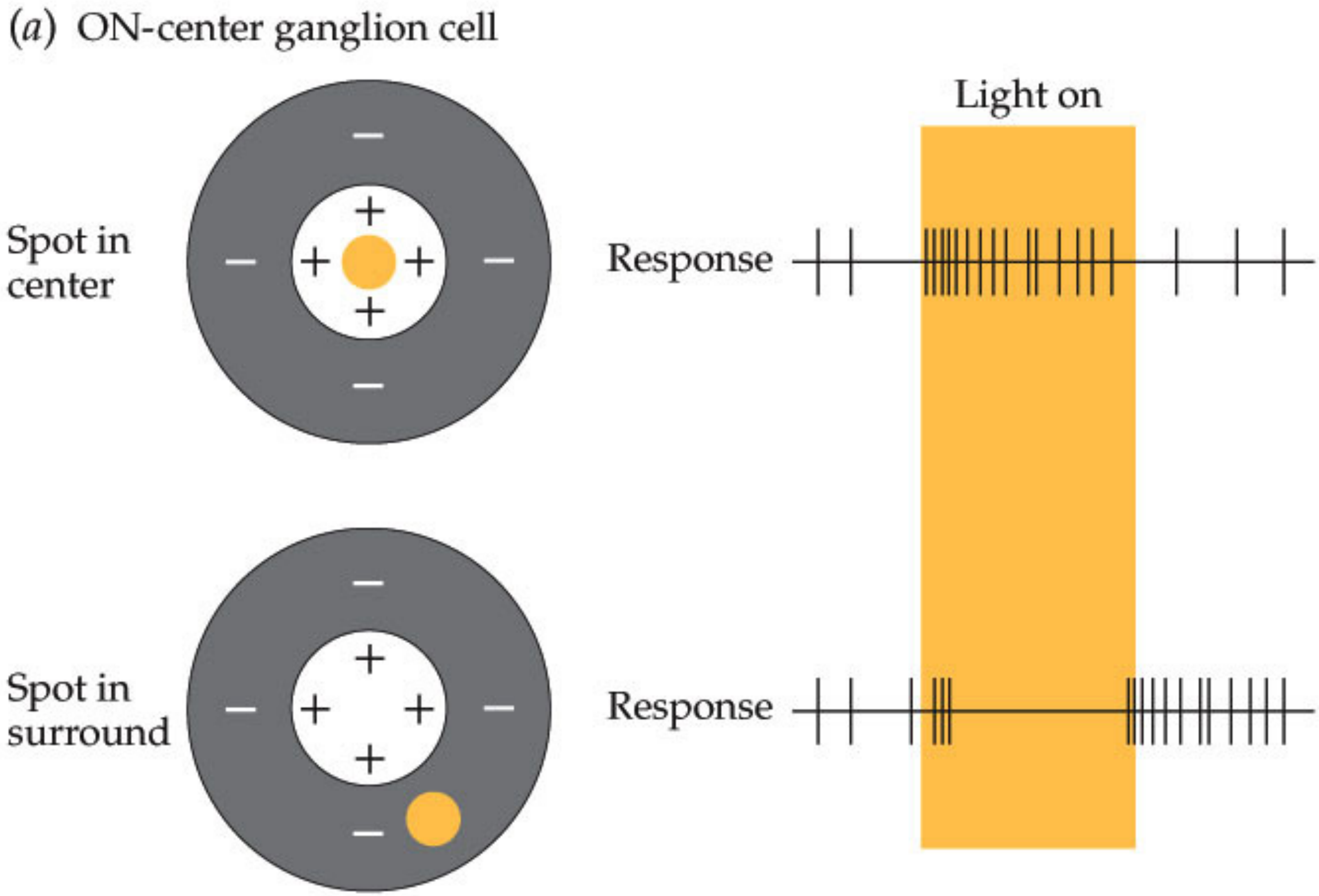
Photoreceptors: not evenly distributed across the retina



- not much color vision in the periphery
- highest sensitivity to dim lights: 5° eccentricity

Retinal information processing: receptive fields

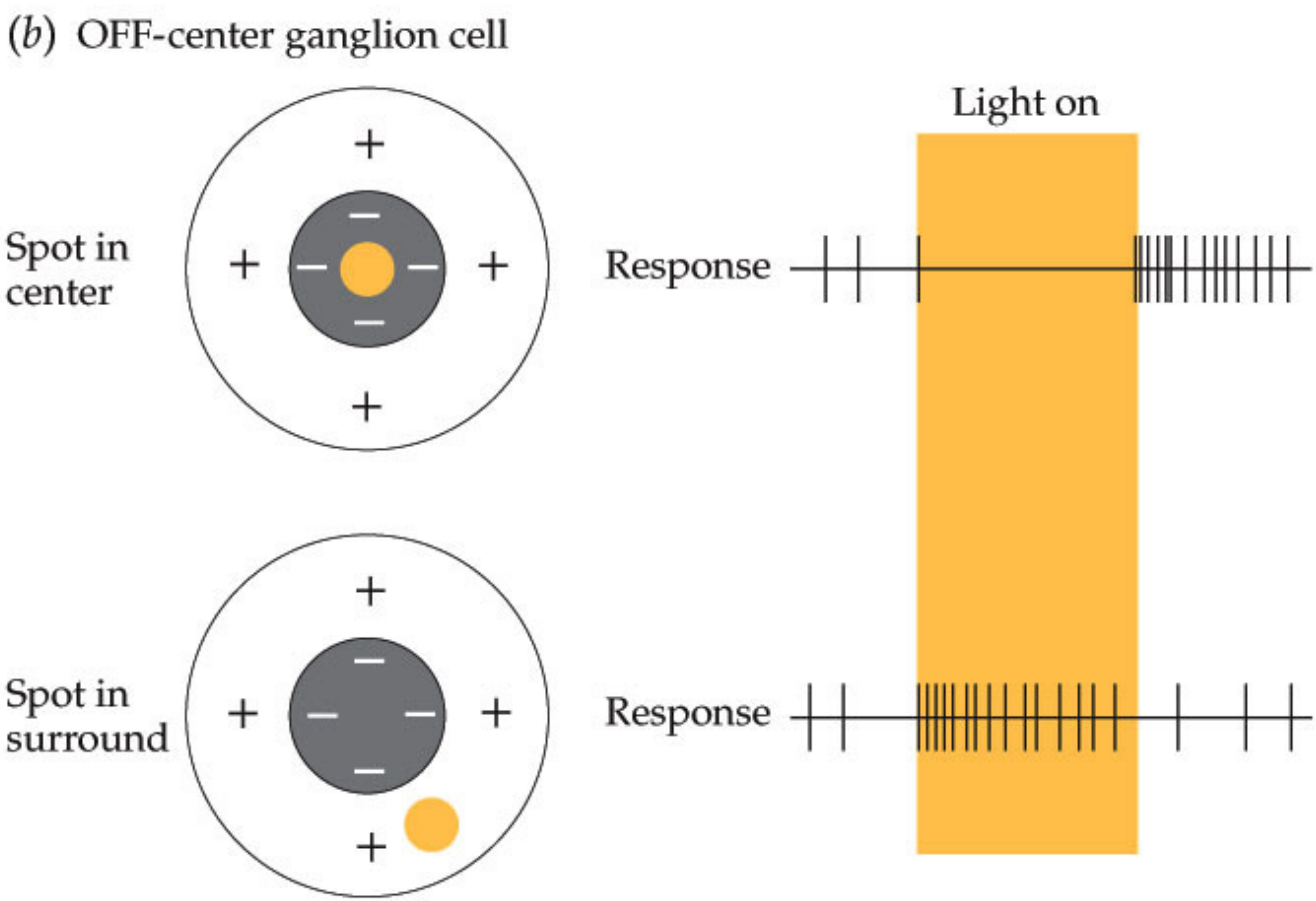
"ON" Cell



[Kuffler 1952]

Retinal information processing: receptive fields

"OFF" Cell



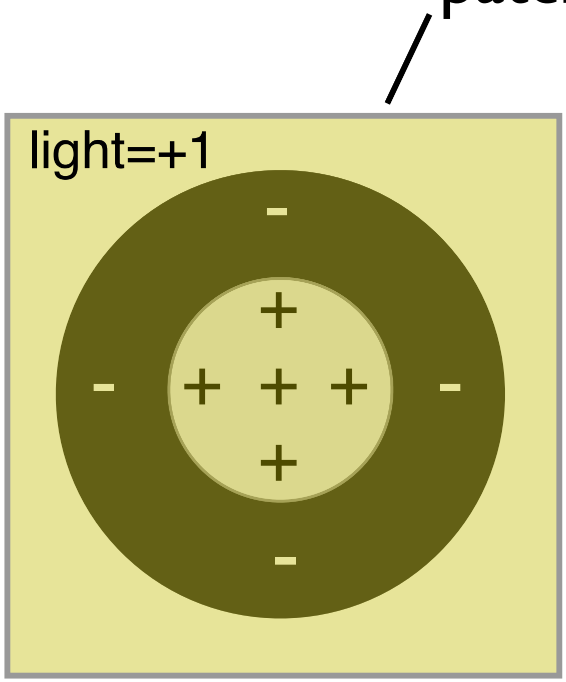
[Kuffler 1952]

<u>Receptive field</u>: "what makes a neuron fire"

- weighting function that the neuron uses to add up its inputs

Response to a dim light

patch of light



light level

$$1\times(+5) \;+\; 1\times(-4) \;=\; +1 \text{ spikes}$$

"center" weight

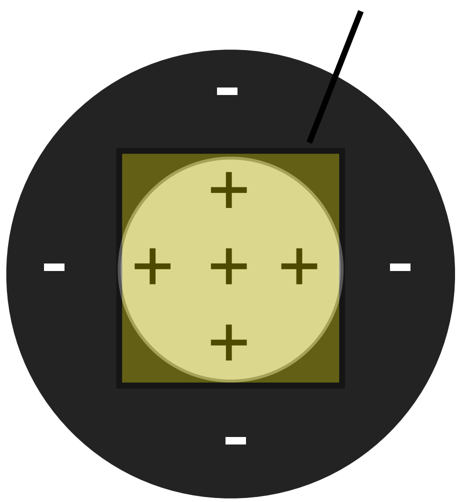
"surround" weight

Receptive field: "what makes a neuron fire"

- weighting function that the neuron uses to add up its inputs

Response to a spot of light

patch of bright light



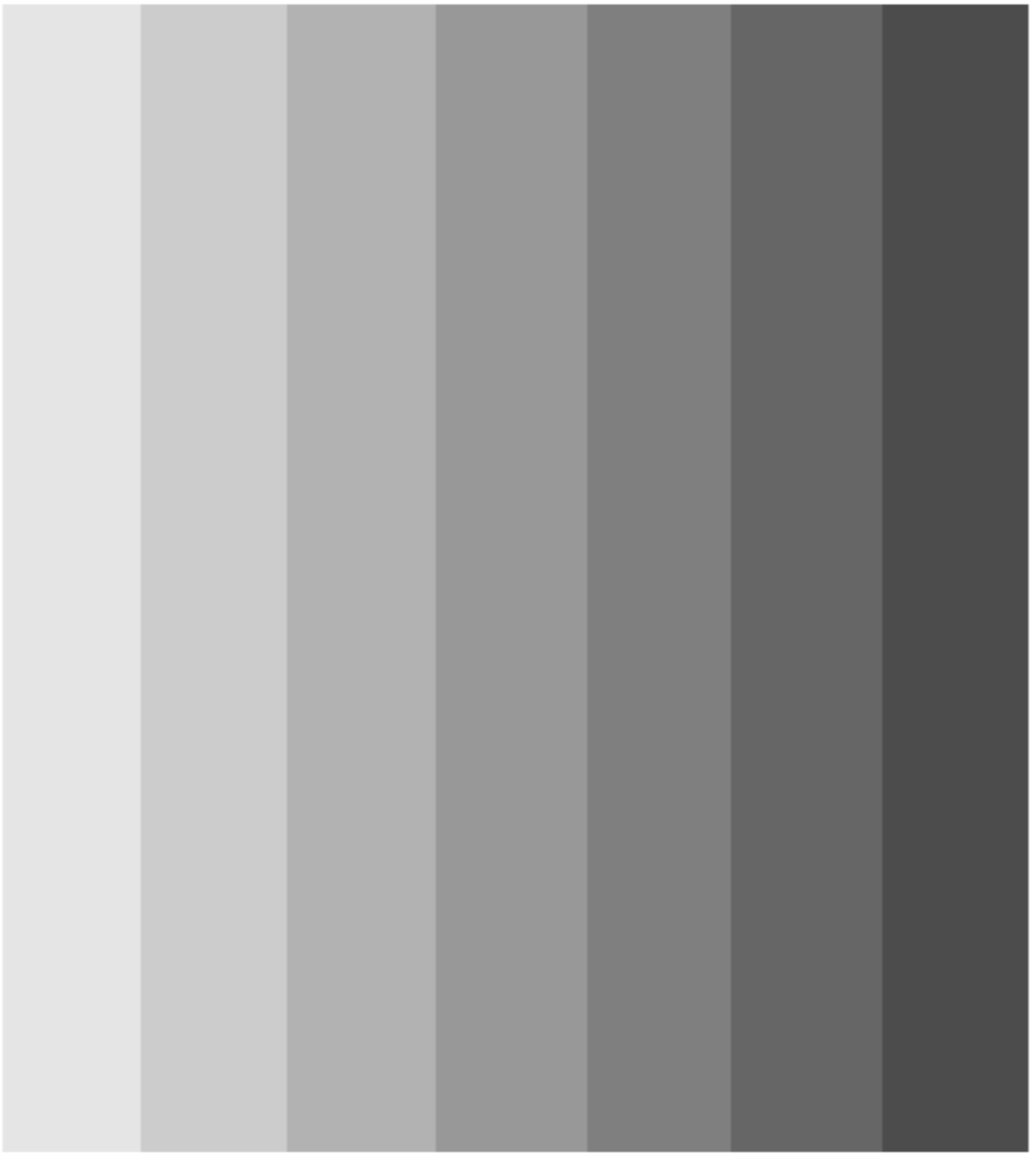
light level

$1\times(+5) \quad + \quad 0\times(-4) \quad = \quad +5$ spikes

"center" weight

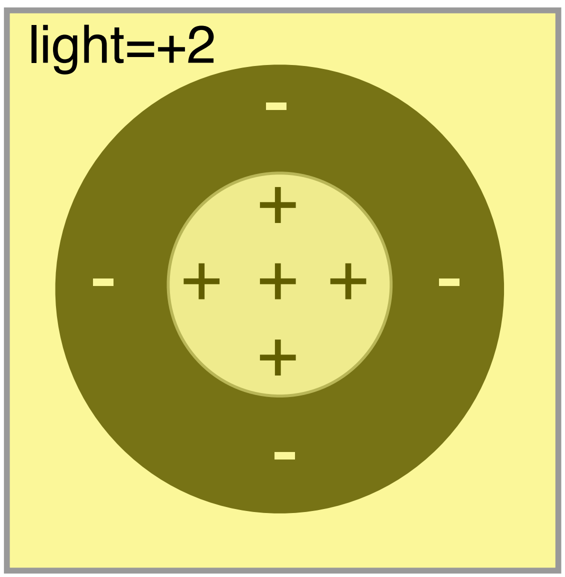
"surround" weight

Mach Bands



Each stripe has constant luminance ("light level")

Response to a bright light



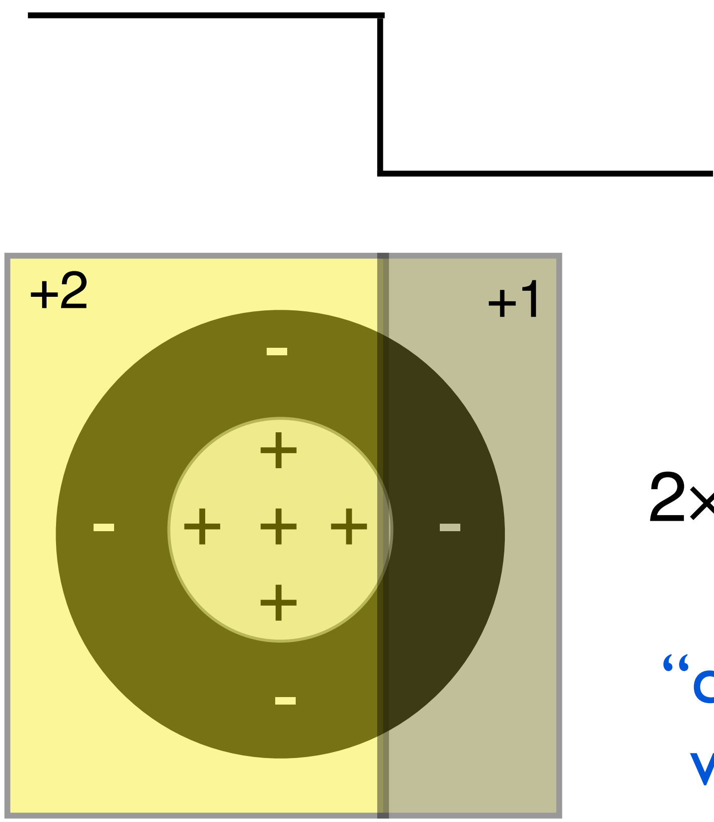
higher light level

$$2\times(+5) \quad + \quad 2\times(-4) \quad = \quad +2 \text{ spikes}$$

“center” weight

“surround” weight

Response to an edge

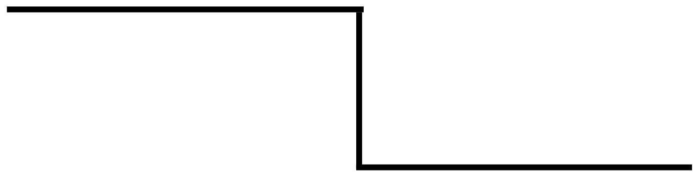


$$2\times(+5) \;+\; 2\times(-3) + 1\times(-1) \;=\; +3 \text{ spikes}$$

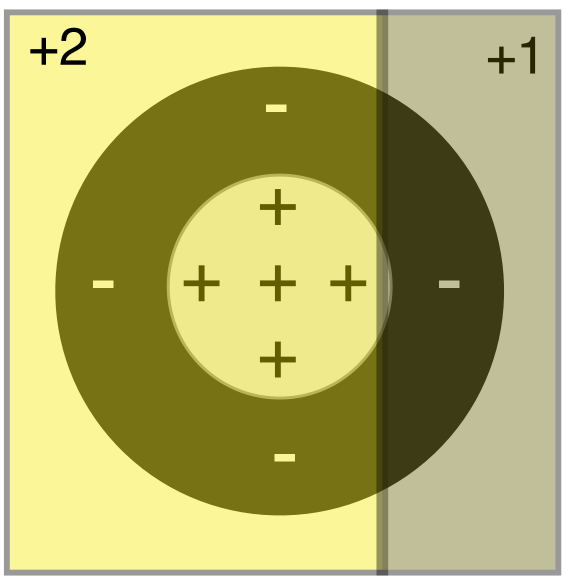
"center" weight

"surround" weight

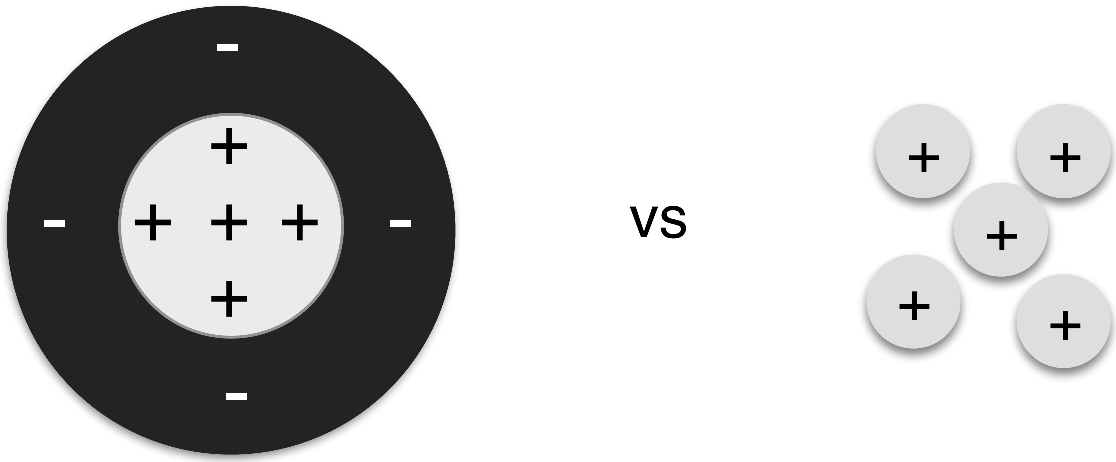
Mach Band response



+2	+2	+2	+3	0	+1	+1	+1
+2	+2	+2	+3	0	+1	+1	+1
+2	+2	+2	+3	0	+1	+1	+1
+2	+2	+2	+3	0	+1	+1	+1
+2	+2	+2	+3	0	+1	+1	+1
+2	+2	+2	+3	0	+1	+1	+1



$$\underbrace{2\times(+5)}_{\text{“center” weight}} + \underbrace{2\times(-3) + 1\times(-1)}_{\text{“surround” weight}} = +3 \text{ spikes}$$



It makes intuitive sense to think that encoding the contrast (local changes in light) in RGC responses might be more efficient than sending the raw light levels (eg raw photoreceptor responses).

Can we make the notion of "efficiency" precise?

Barlow 1961
Atick & Redlich 1990

Efficient Coding Hypothesis:

- goal of nervous system: maximize information about environment
  (one of the core "big ideas" in theoretical neuroscience)

redundancy: $$R = 1 - \frac{I}{C}$$

mutual information

channel capacity

Barlow 1961
Atick & Redlich 1990

Efficient Coding Hypothesis:

- goal of nervous system: maximize information about environment
  (one of the core "big ideas" in theoretical neuroscience)

redundancy: $$R = 1 - \frac{I}{C}$$

mutual information

channel capacity

mutual information:

$$I(x,y) = \underbrace{H(y)}_{\text{response entropy}} - \underbrace{H(y|x)}_{\text{"noise" entropy}}$$

- avg # yes/no questions you can answer about x given y ("bits")
- entropy: $H(x) = -\sum p(x) \log p(x)$

channel capacity:

$$C = \sup_{P_x} I(x,y)$$

- upper bound on mutual information
- determined by physical properties of encoder

Barlow 1961
Atick & Redlich 1990

Barlow's original version:

redundancy: $R = 1 - \dfrac{I}{C}$ mutual information

mutual information:

$$I(x,y) = \underbrace{H(y)}_{\text{response entropy}} - \underbrace{\cancel{H(y|x)}}_{\text{"noise" entropy}}$$

if responses are noiseless

Barlow 1961
Atick & Redlich 1990

Barlow's original version:

response entropy

redundancy: $$R = 1 - \frac{H(Y)}{C}$$

mutual information:

$$I(x,y) = \underbrace{H(y)}_{\text{response entropy}} - \underbrace{\cancel{H(y|x)}}_{\text{"noise" entropy}}$$

noiseless system

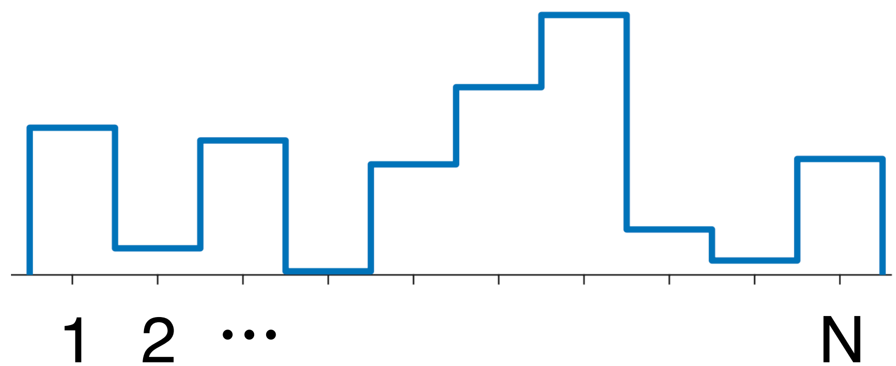
$\Longrightarrow$ brain should maximize response entropy

- use full dynamic range
- decorrelate ("reduce redundancy")

• mega impact: huge number of theory and experimental papers focused on decorrelation / information-maximizing codes in the brain

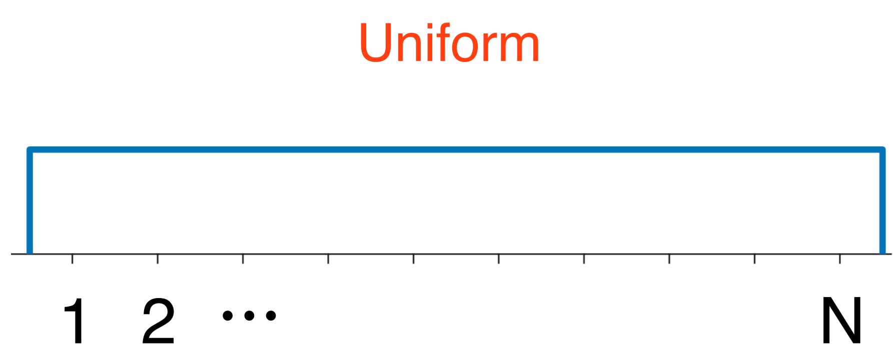
Maximum entropy distributions

- Q: what is the maximum entropy (discrete) distribution on *N* bins?



Maximum entropy distributions

- Q: what is the maximum entropy (discrete) distribution on *N* bins?

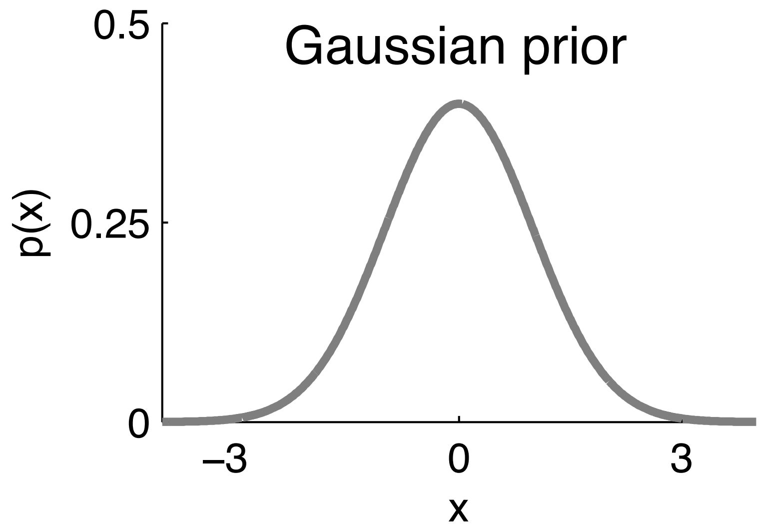


Application Example: single neuron encoding stimuli from a distribution P(x)

stimulus prior $$x \sim P(x)$$

noiseless, discrete encoding $$y = f(x), \qquad y \in \{y_1, y_2, \ldots, y_n\}$$

Q: what solution for infomax?



Application Example: single neuron encoding stimuli from a distribution P(x)

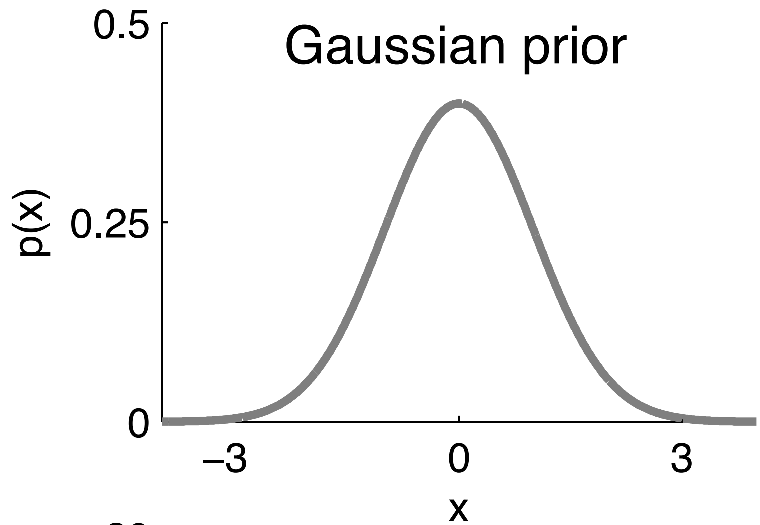
stimulus prior $$x \sim P(x)$$

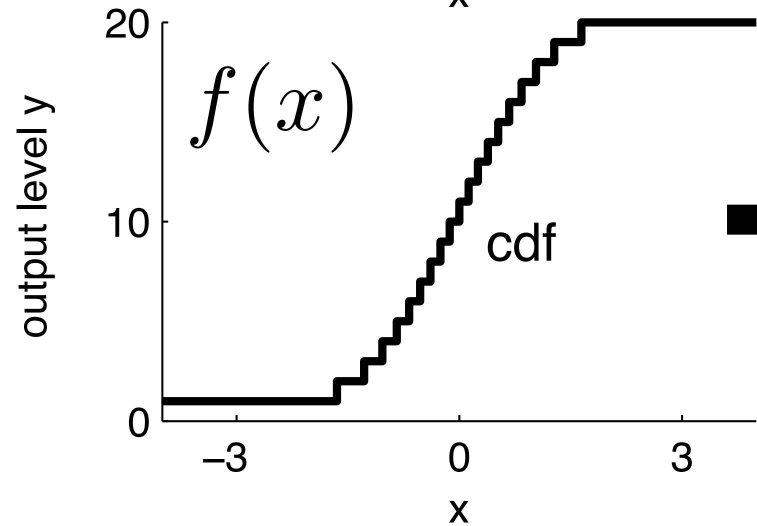
noiseless, discrete encoding $$y = f(x), \qquad y \in \{y_1, y_2, \ldots, y_n\}$$

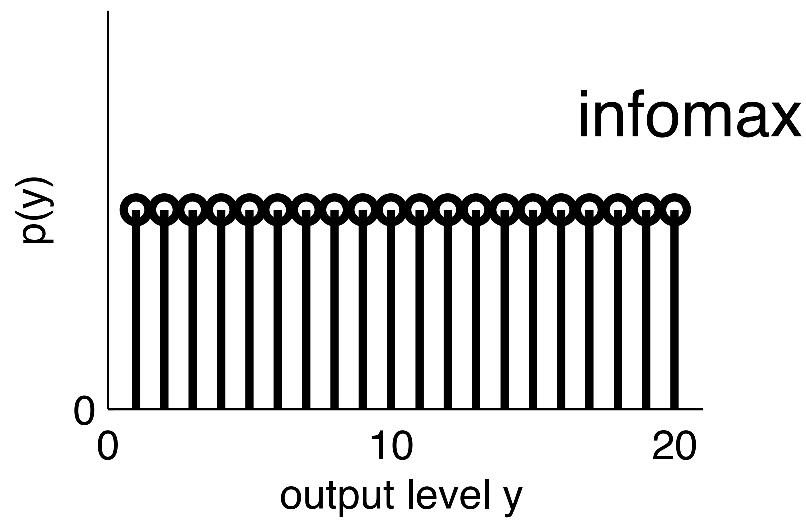
Q: what solution for infomax?

A: histogram-equalization

$$I(X,Y) = H(Y) - \cancel{H(Y|X)}$$

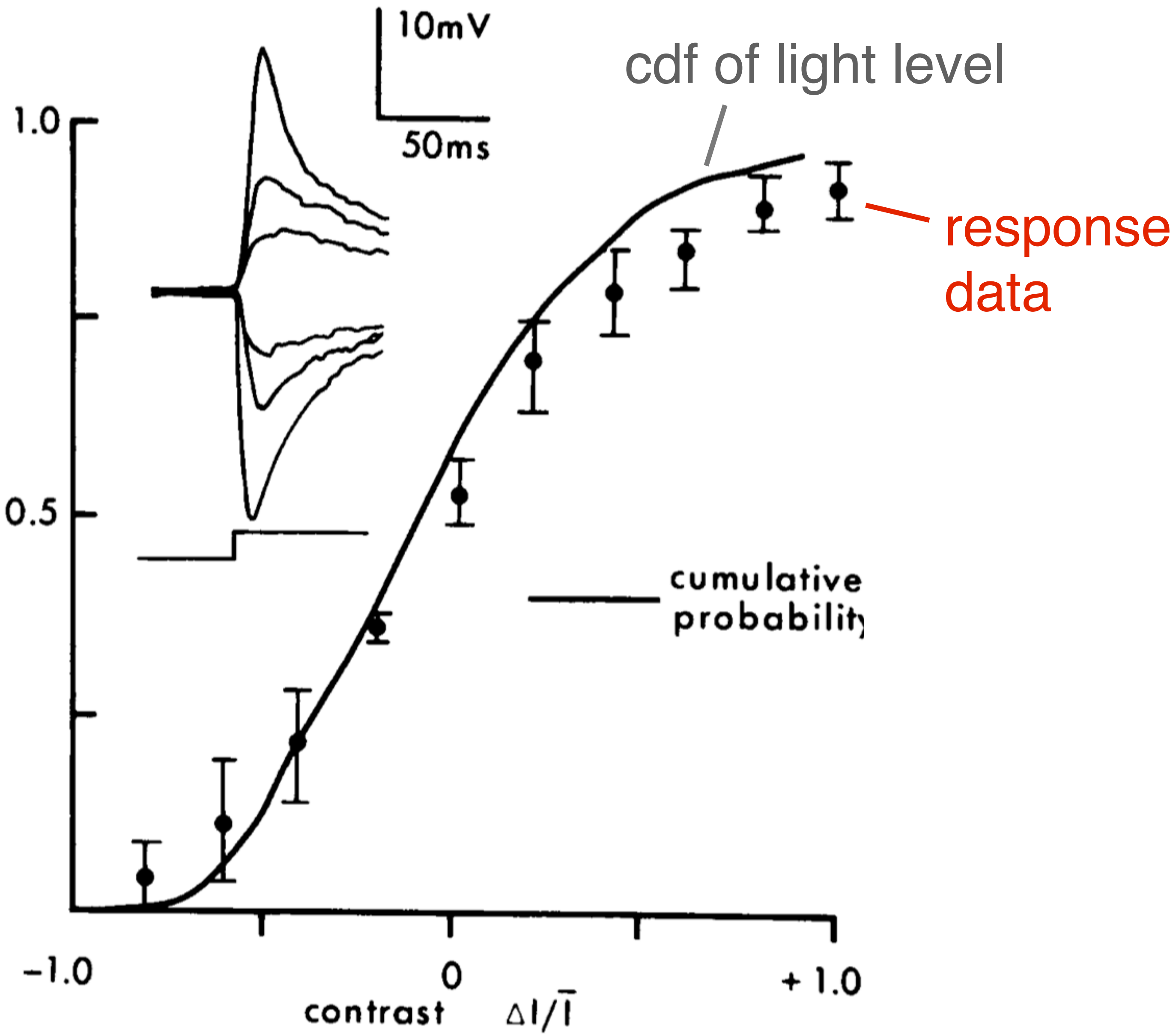






Laughlin 1981: blowfly light response

• first major validation of Barlow's theory



What about multiple neurons?

We want: joint distribution $p(r_1, r_2)$

that achieves maximum entropy $H(r_1, r_2)$

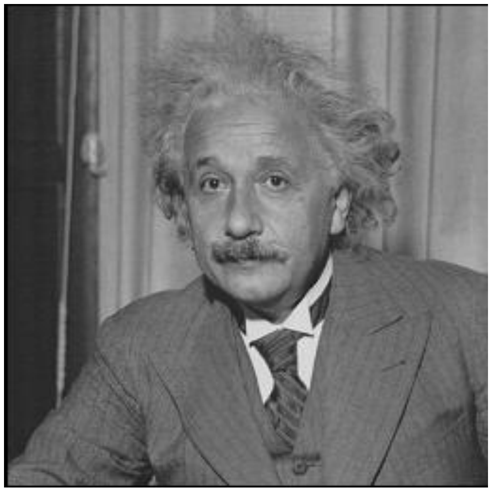
given fixed marginals $p(r_1)$ and $p(r_2)$

$$H(r_1, r_2) = H(r_1) + H(r_2) - I(r_1, r_2)$$

- This is clearly maximized when neurons are independent, i.e., $I(r_1, r_2) = 0$
- **solution**: neurons should be (marginally) independent!

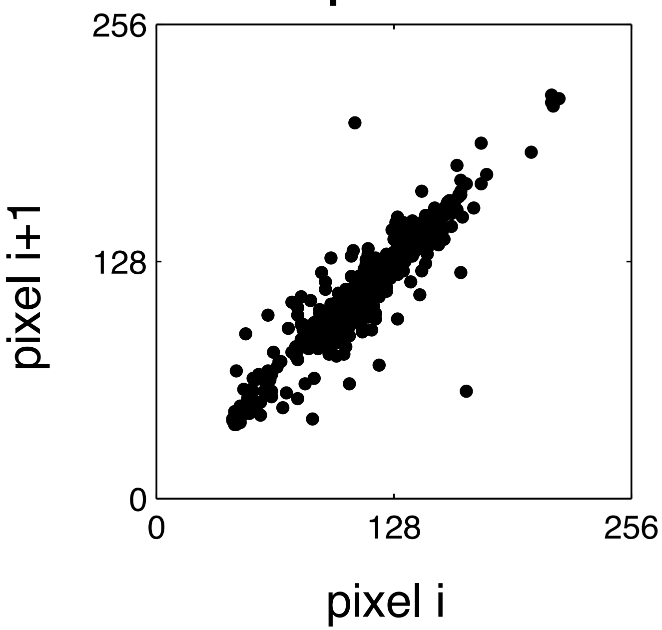
basic intuition

natural image



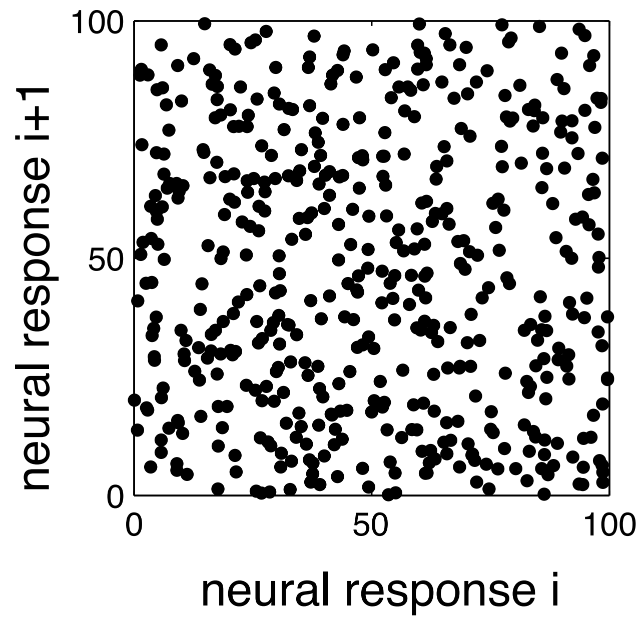
nearby pixels exhibit
strong dependencies

pixels



desired
encoding

neural representation



efficient coding: take-home

1. For single neurons: given a constraint on the response (e.g., maximum or mean rate), information transfer maximized if marginal response distribution is a maximum-entropy subject to that constraint

$$p(r) \propto e^{\lambda f(r)}$$

2. For multiple neurons: information is maximized when response distributions are independent (aka "redundancy reduction")

$$p(r_1, r_2 \ldots, r_n) = p(r_1)p(r_2) \cdots p(r_n)$$

break?

We’ve discussed:

• how the retina processes that image to extract contrast (with “center-surround” receptive fields)

• efficient coding: normative theory of retinal coding

Next:

• how does visual cortex process information?

early visual pathway

optic nerve

optic chiasm

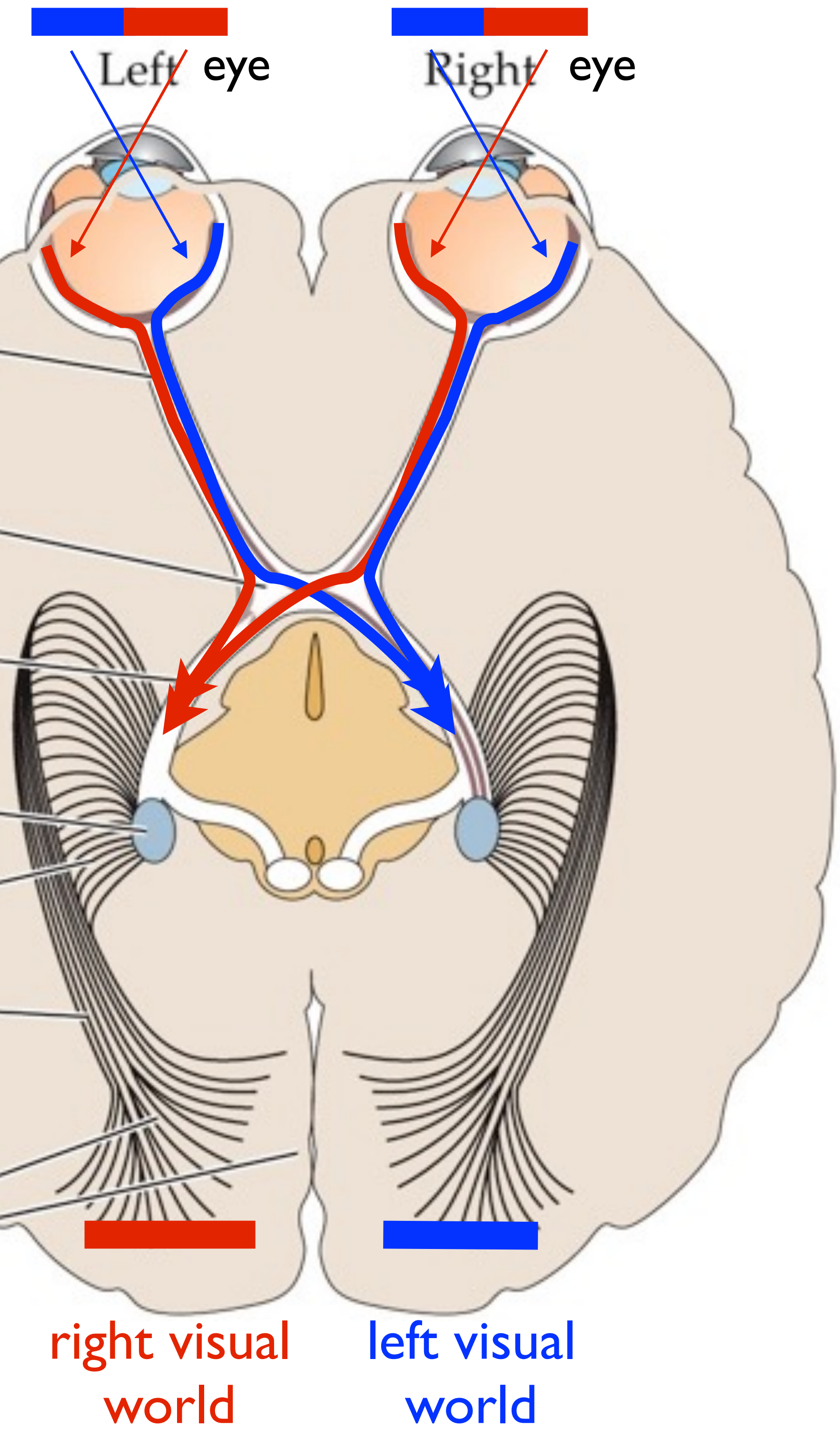
optic tract

thalamus: **lateral geniculate nucleus (LGN)**

optic radiations

cortex: **primary visual cortex ("V1")**

(aka "striate cortex")



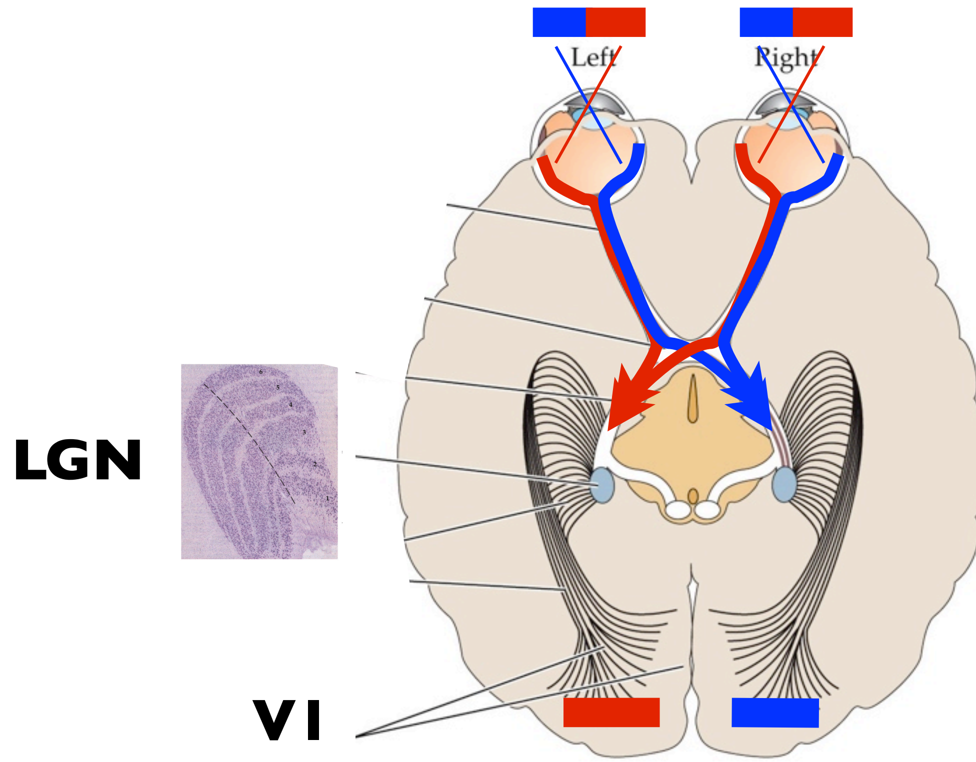
Primary Visual Cortex

- Striate cortex: known as primary visual cortex, or V1
- "Primary visual cortex" = first place in cortex where visual information is processed

  (Previous two stages: retina and LGN are pre-cortical)

Receptive Fields: monocular vs. binocular

- <u>LGN cells</u>: responds to one eye or the other, ***never both***



- <u>V1 cells</u>: can respond to input from **both eyes**

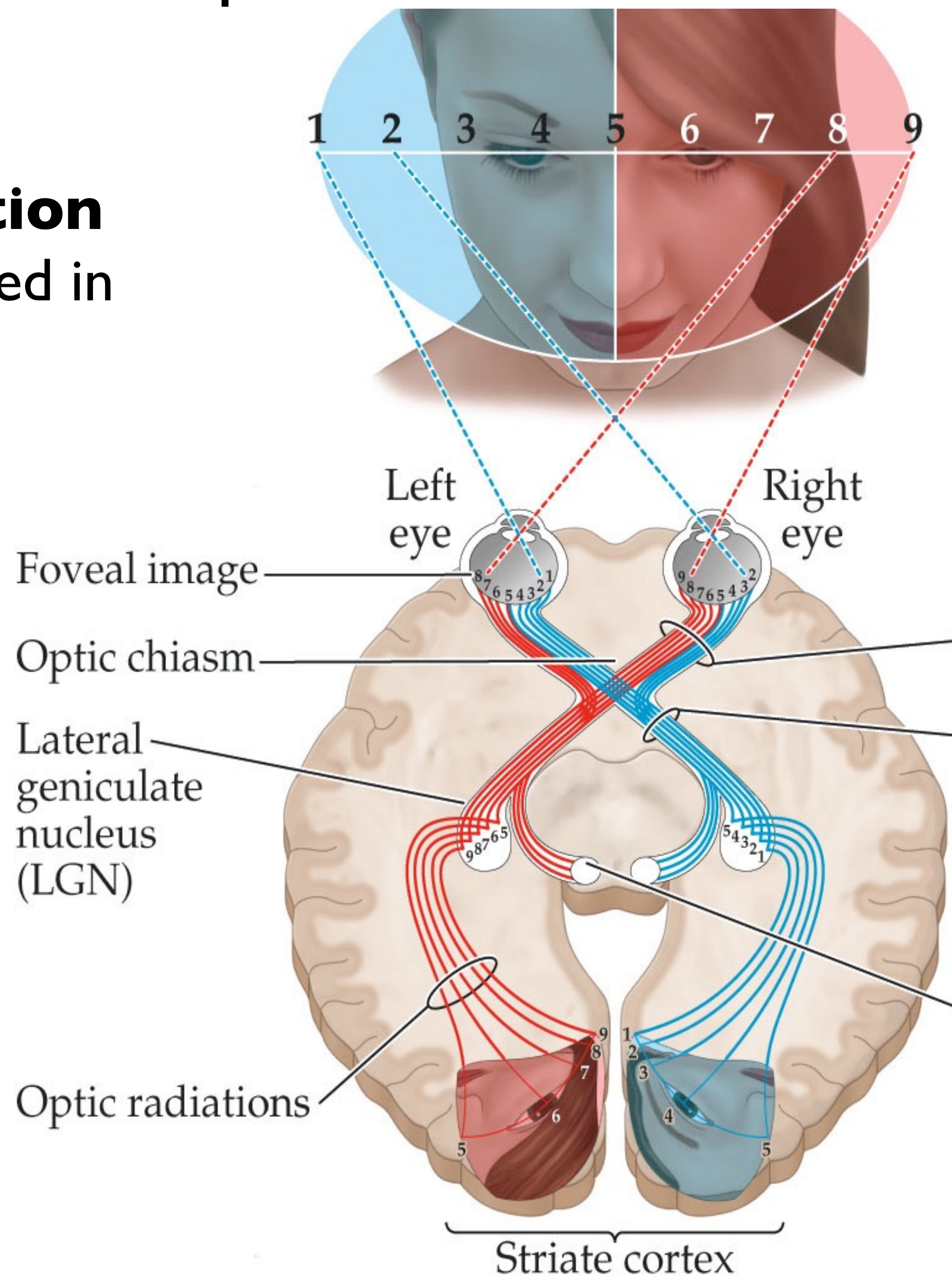
(but V1 neurons still tend to have a ***preferred*** **eye** - they spike more to input from one eye)

Topography: mapping of visual space onto visual cortex

("retinotopic", "visuotopic")

- **contralateral representation**
  - each visual field (L/R) represented in opposite hemisphere

- **cortical magnification**
  - unequal representation of fovea vs. periphery in cortex



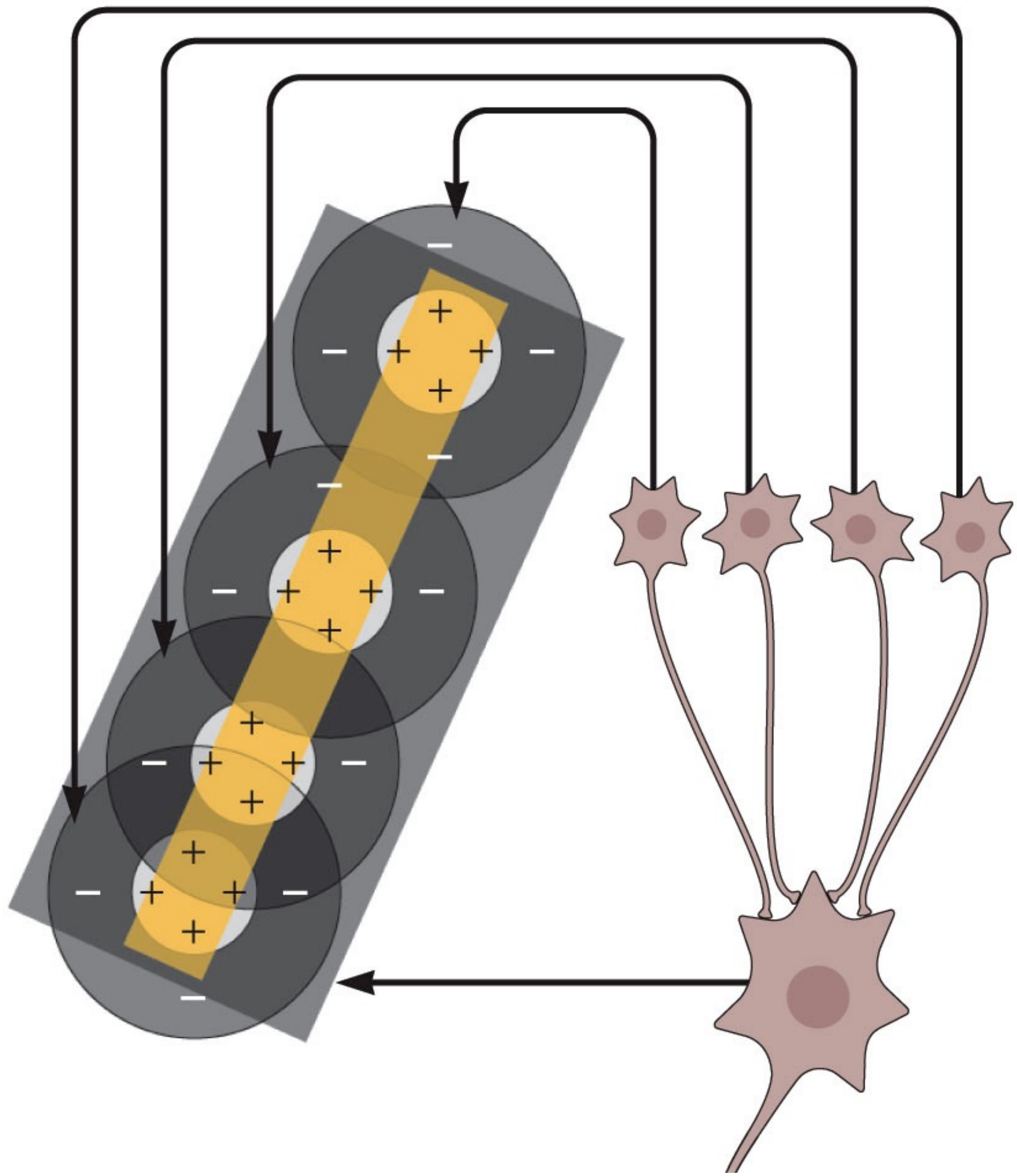
major change in sensory representation in V1

retina & LGN:

- circular RFs
- 1M fibers (from RGCs)

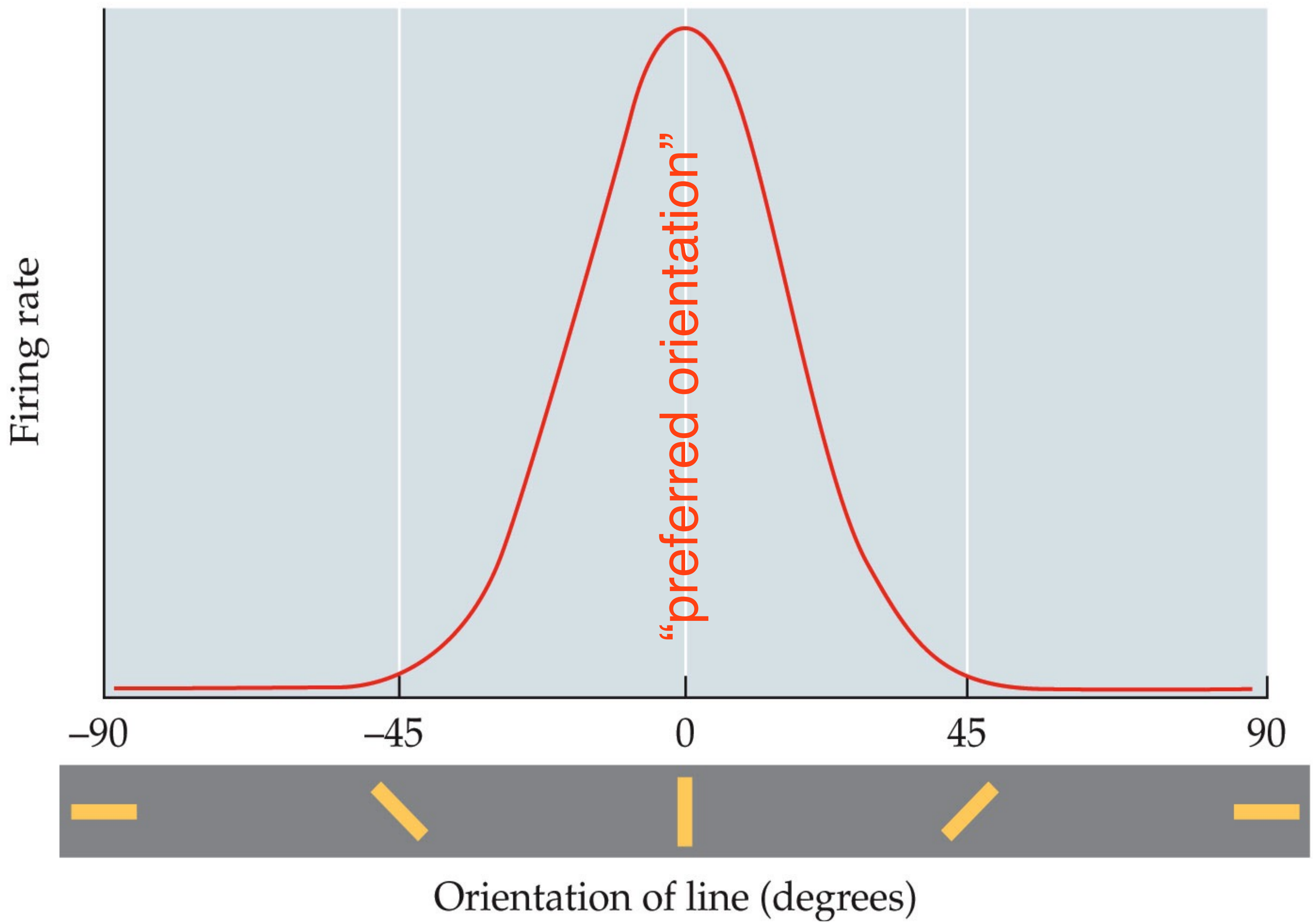
V1

- elongated, oriented RFs
- 200M cells!



Orientation tuning:

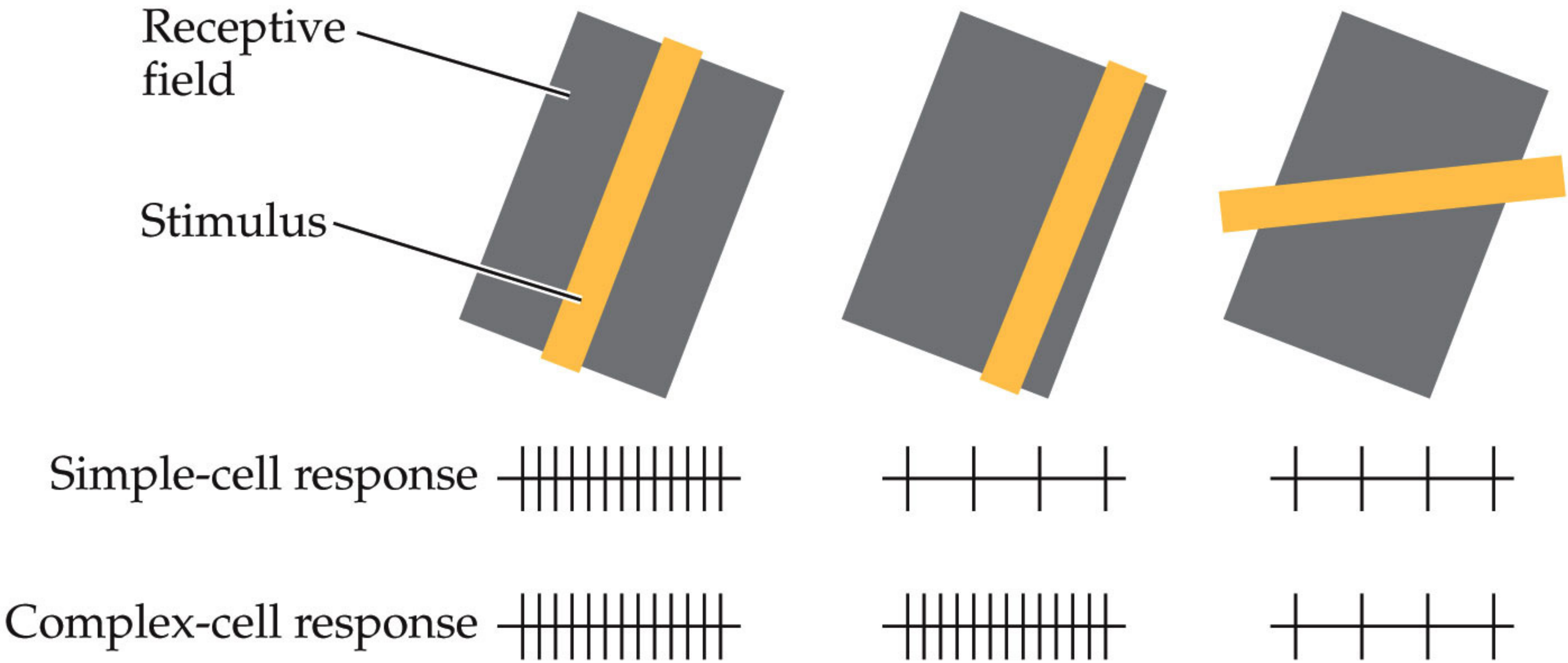
- neurons in V1 respond more to bars of certain orientations
- response rate falls off with difference from preferred orientation



Simple vs. Complex Cells

Cells in V1 respond best to bars of light rather than to spots of light

- **“simple” cells**: prefer bars of light, or prefer bars of dark
- **“complex” cells**: respond to both bars of light and dark

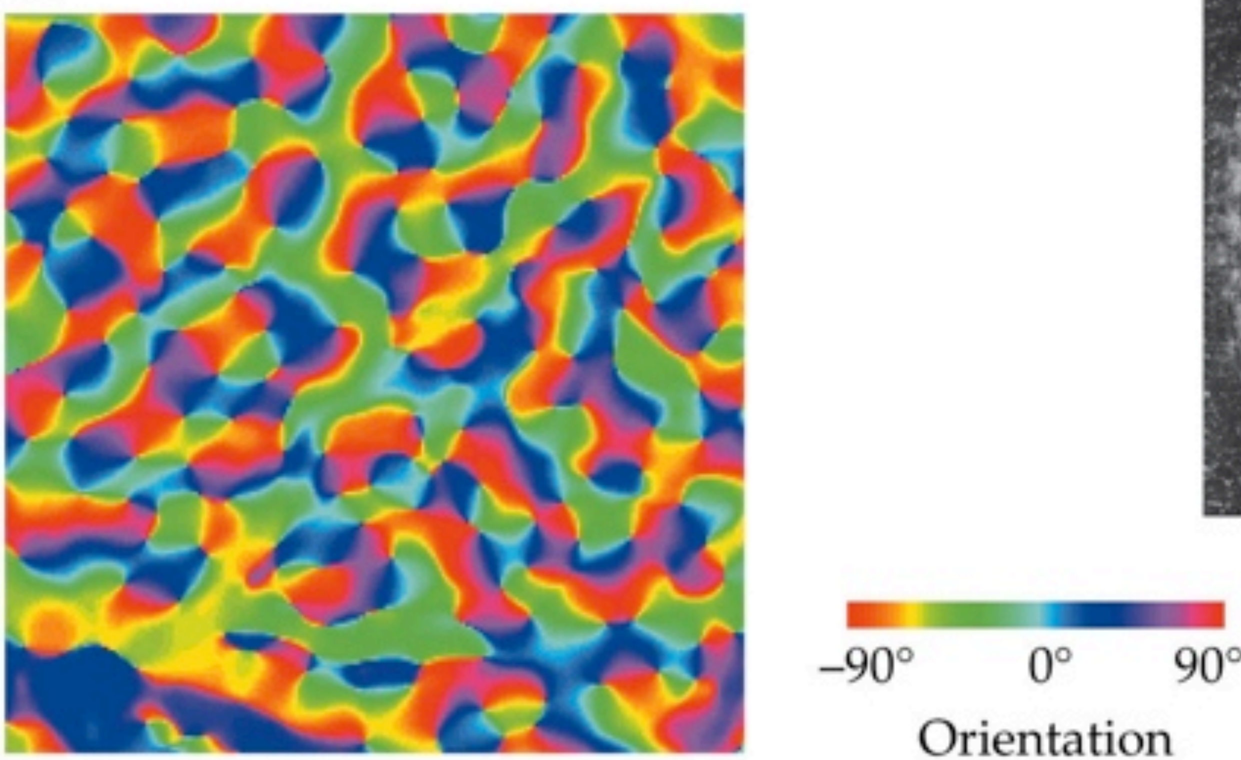


Receptive Fields in V1

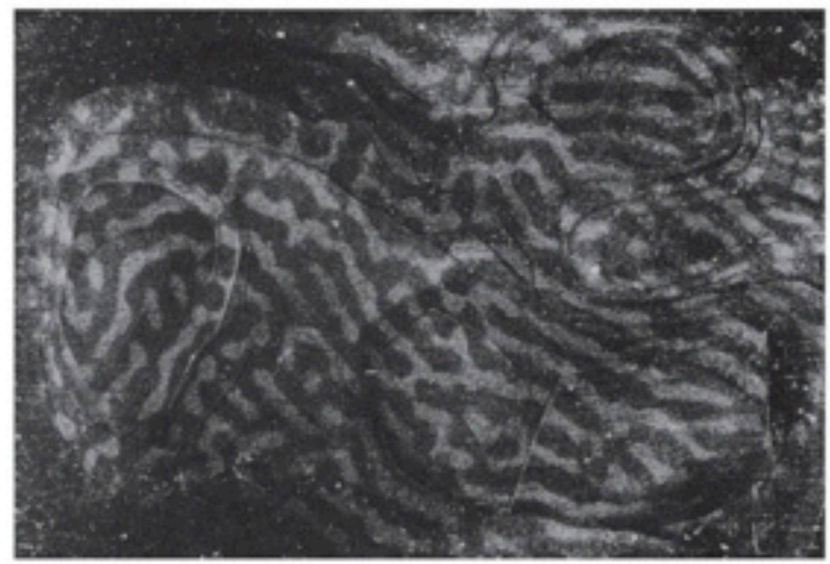
[see link to Hubel & Weisel movie]

<u>Column</u>: a vertical arrangement of neurons

- **orientation column**: for a particular location in cortex, neurons have same preferred orientation

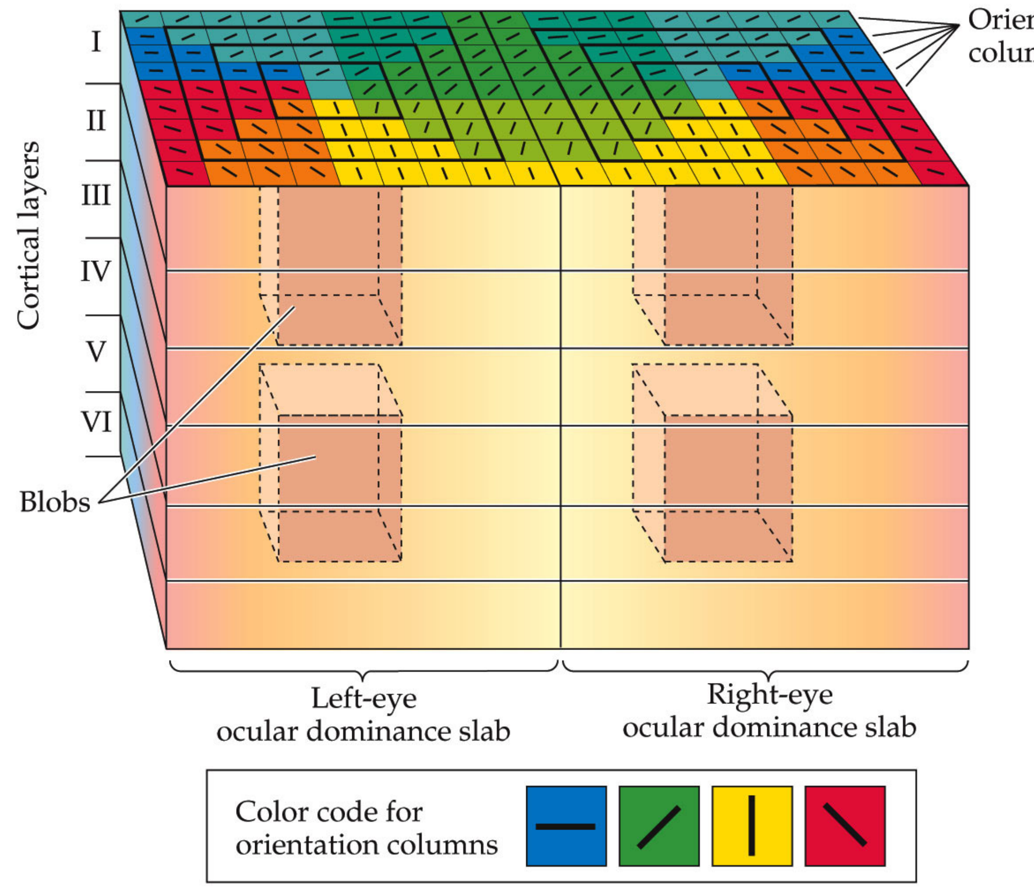


- **ocular dominance column:** for particular location in cortex, neurons have same preferred eye



Hypercolumn - contains all possible columns

- **Hypercolumn**: 1-mm block of V1 containing "all the machinery necessary to look after everything the visual cortex is responsible for, in a certain small part of the visual world" (Hubel, 1982)



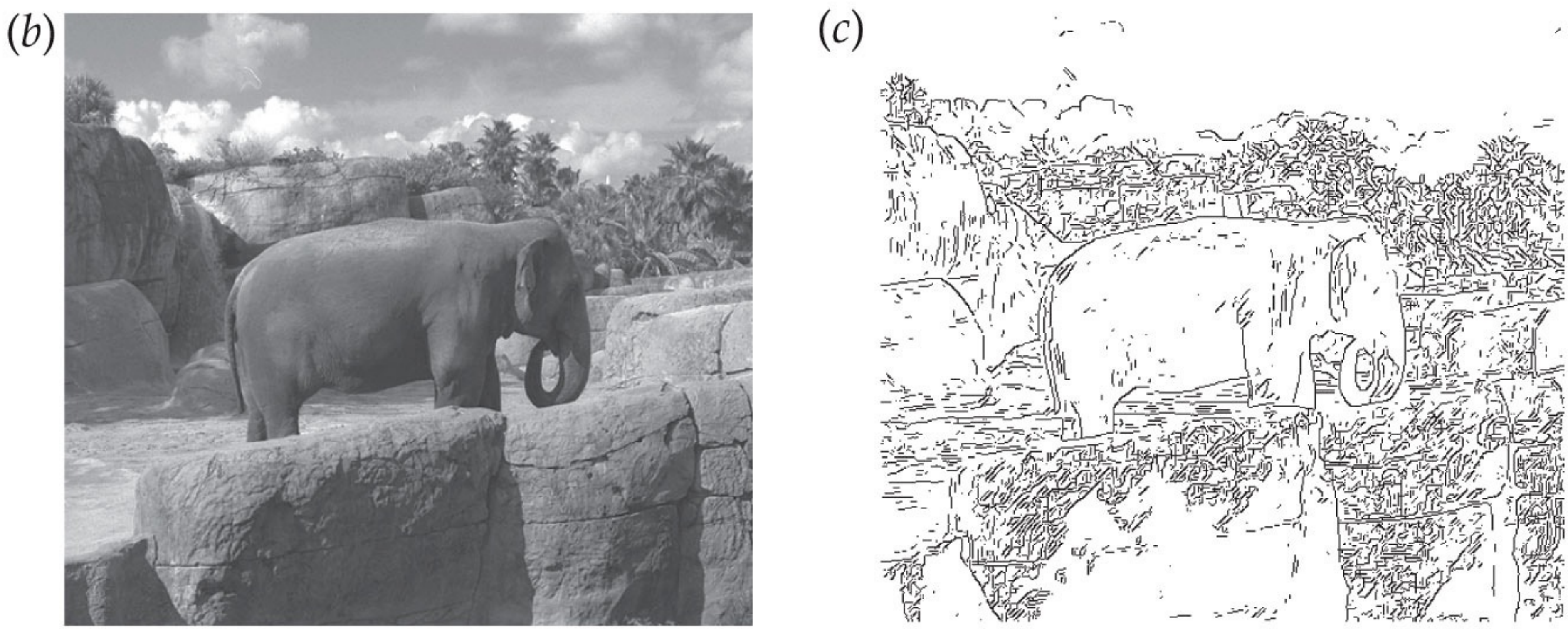
- Each hypercolumn contains a full set of columns
  - has cells responding to every possible orientation, and inputs from left right eyes

Three theories of "What V1 Does"

1. Edge detection
2. Fourier Analysis
3. Sparse coding

“Edge detection” hypothesis for V1:

- The role of V1 is to detect edges
- not obvious that this is a good way to process images
- led to many failed approaches in computer vision



(I might make some disparaging remarks about Marr on this slide if you ask me)

Three theories of "What V1 Does"

1. Edge detection

2. Fourier Analysis

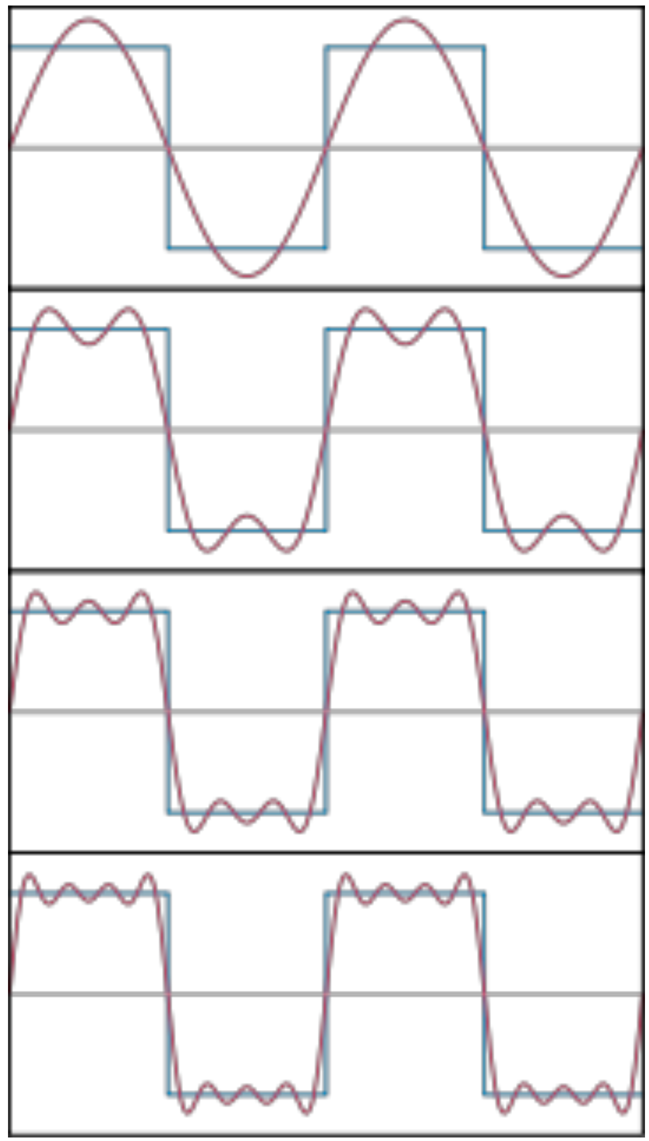
3. Sparse coding

Fourier decomposition

- mathematical decomposition of an image (or sound) into sine waves.

reconstruction:

“image”



1 sine wave

2 sine waves

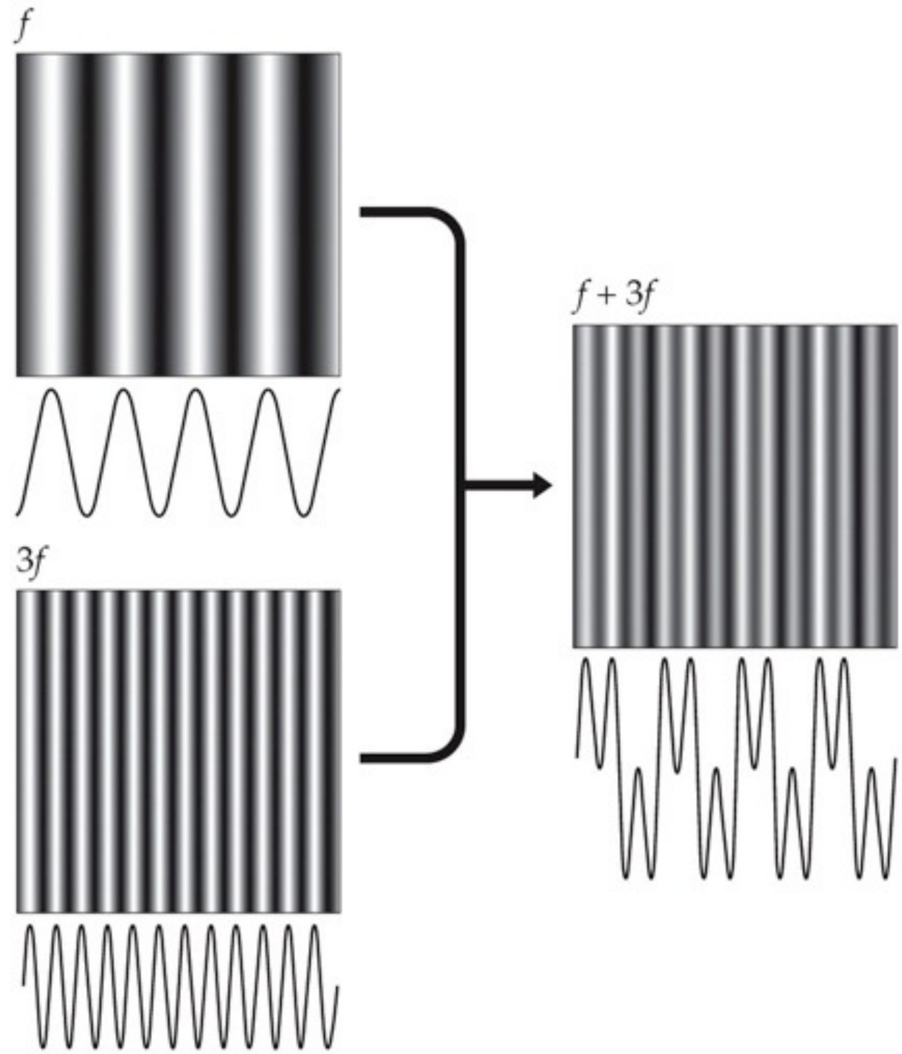
3 sine waves

4 sine waves

“Fourier Decomposition” theory of V1

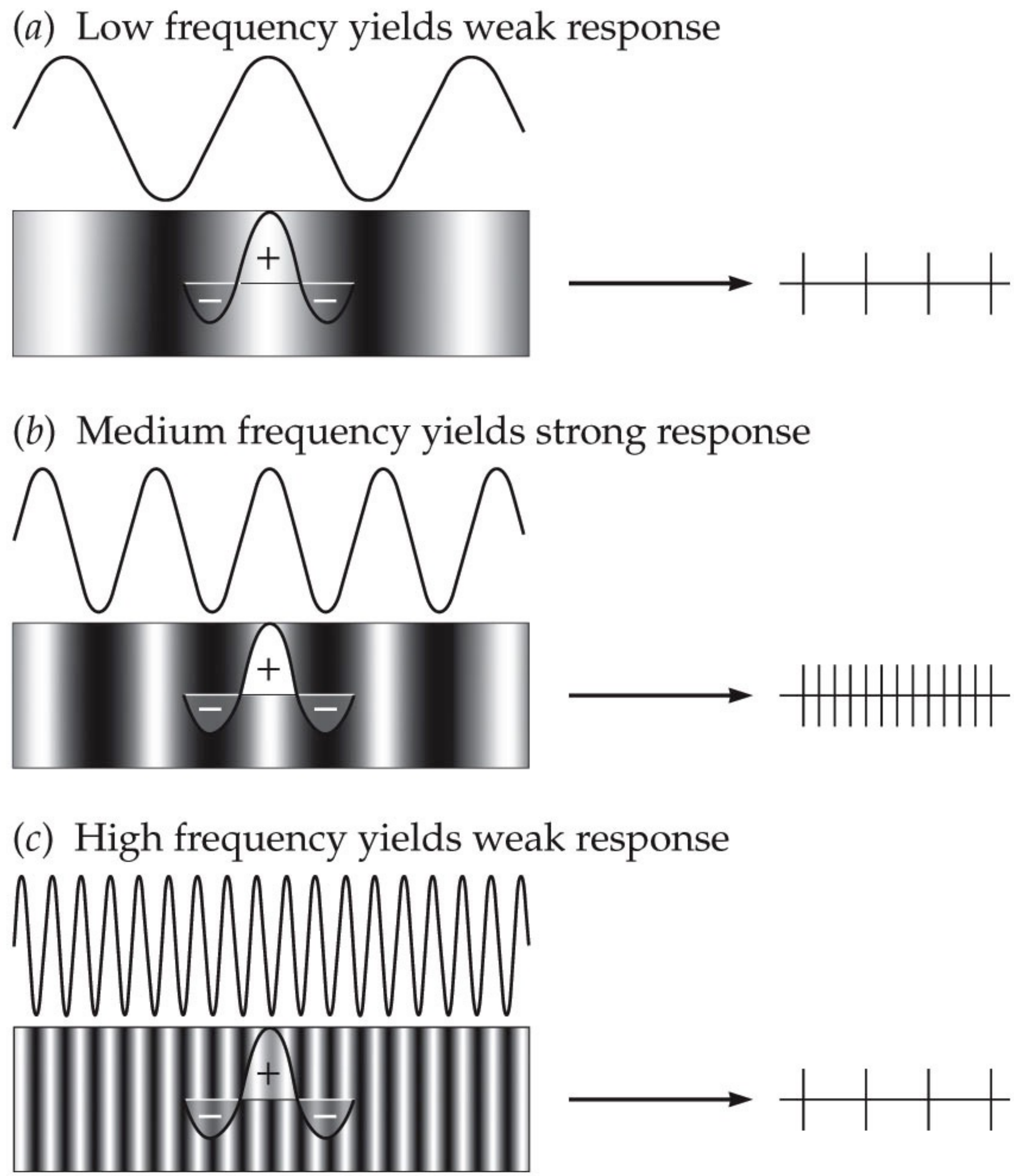
claim: role of V1 is to do “Fourier decomposition”, i.e., break images down into a sum of sine waves

- Summation of two spatial sine waves
- any pattern can be broken down into a sum of sine waves



Retinal Ganglion Cells are tuned to spatial frequency

Response of a ganglion cell to sine gratings of different frequencies



important idea: **spatial frequency channels**

spatial frequency: the number of cycles of a grating per unit of visual angle (usually specified in degrees)

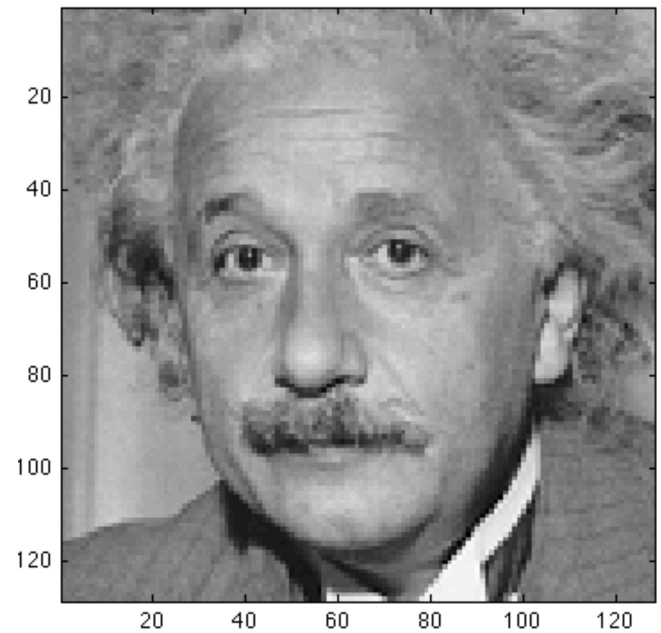
- think of it as: # of bars per unit length

low frequency

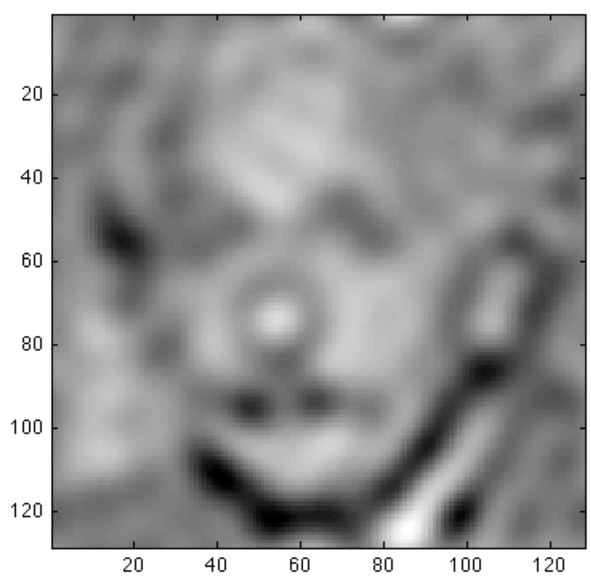
intermediate

high frequency

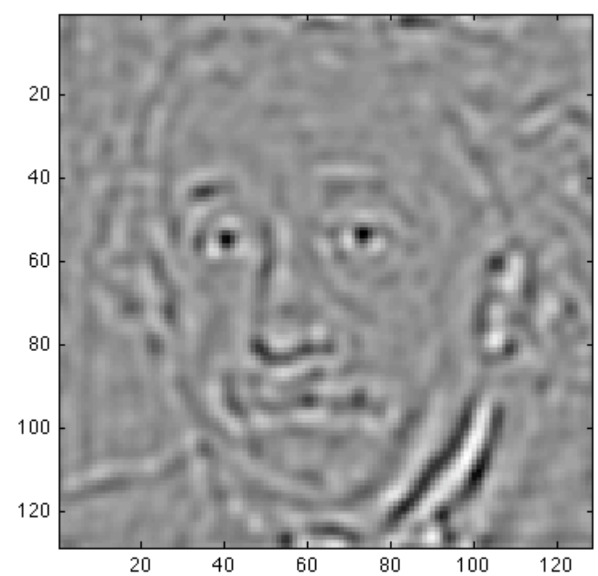
original



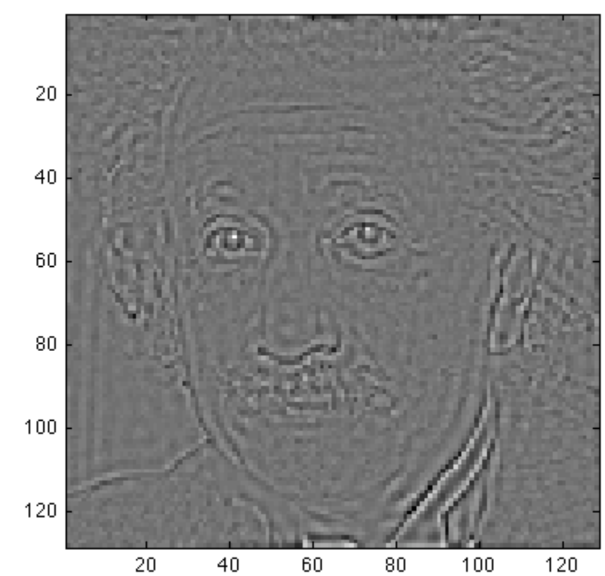
low



medium



high



The contrast sensitivity function

Human contrast sensitivity

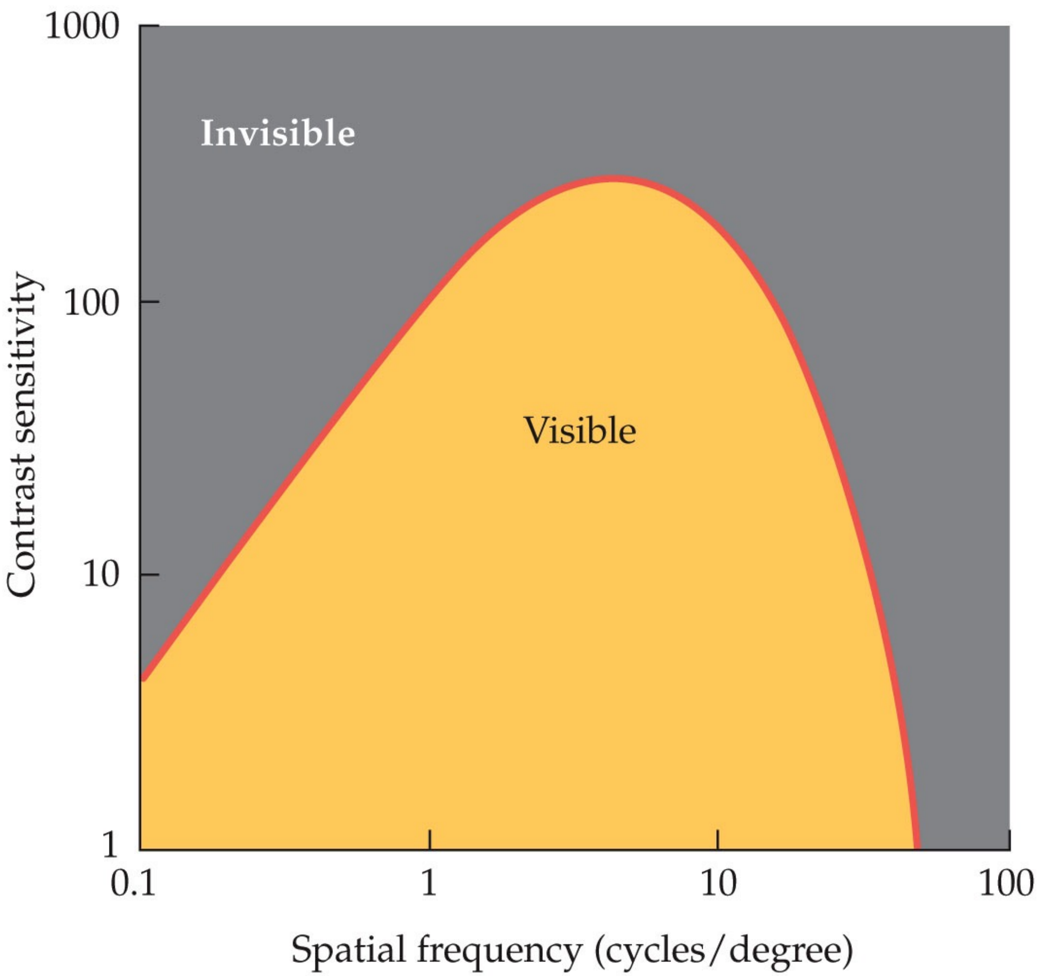


illustration of this sensitivity

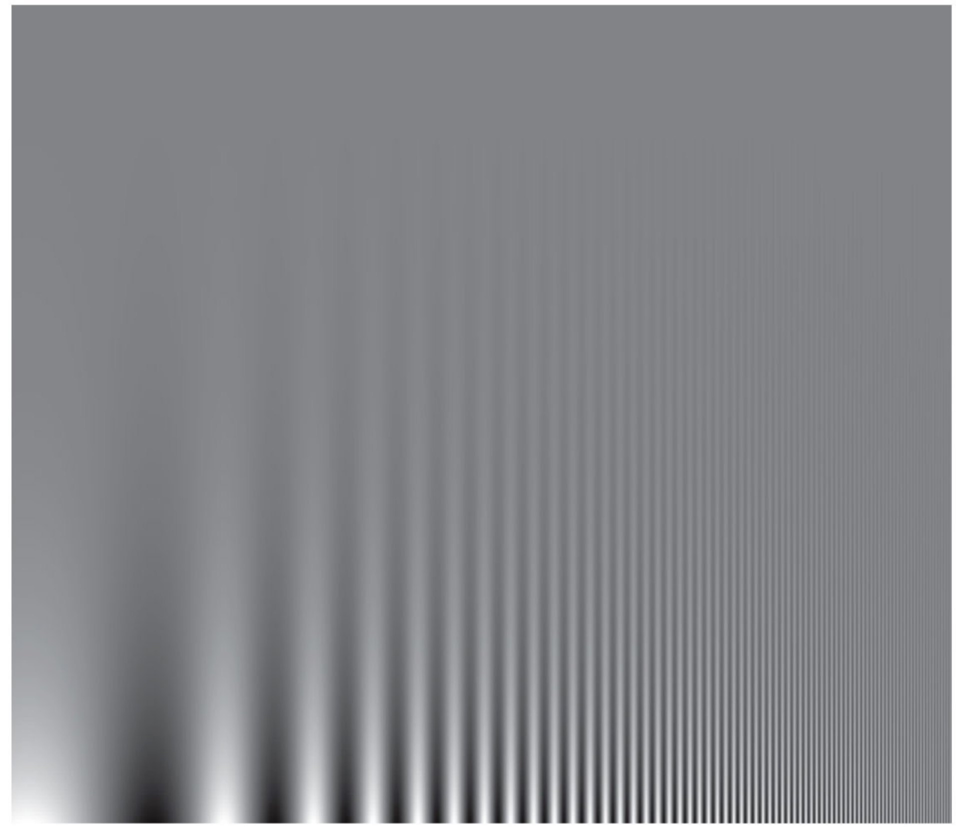


Image Illustrating Spatial Frequency Channels

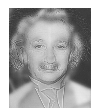
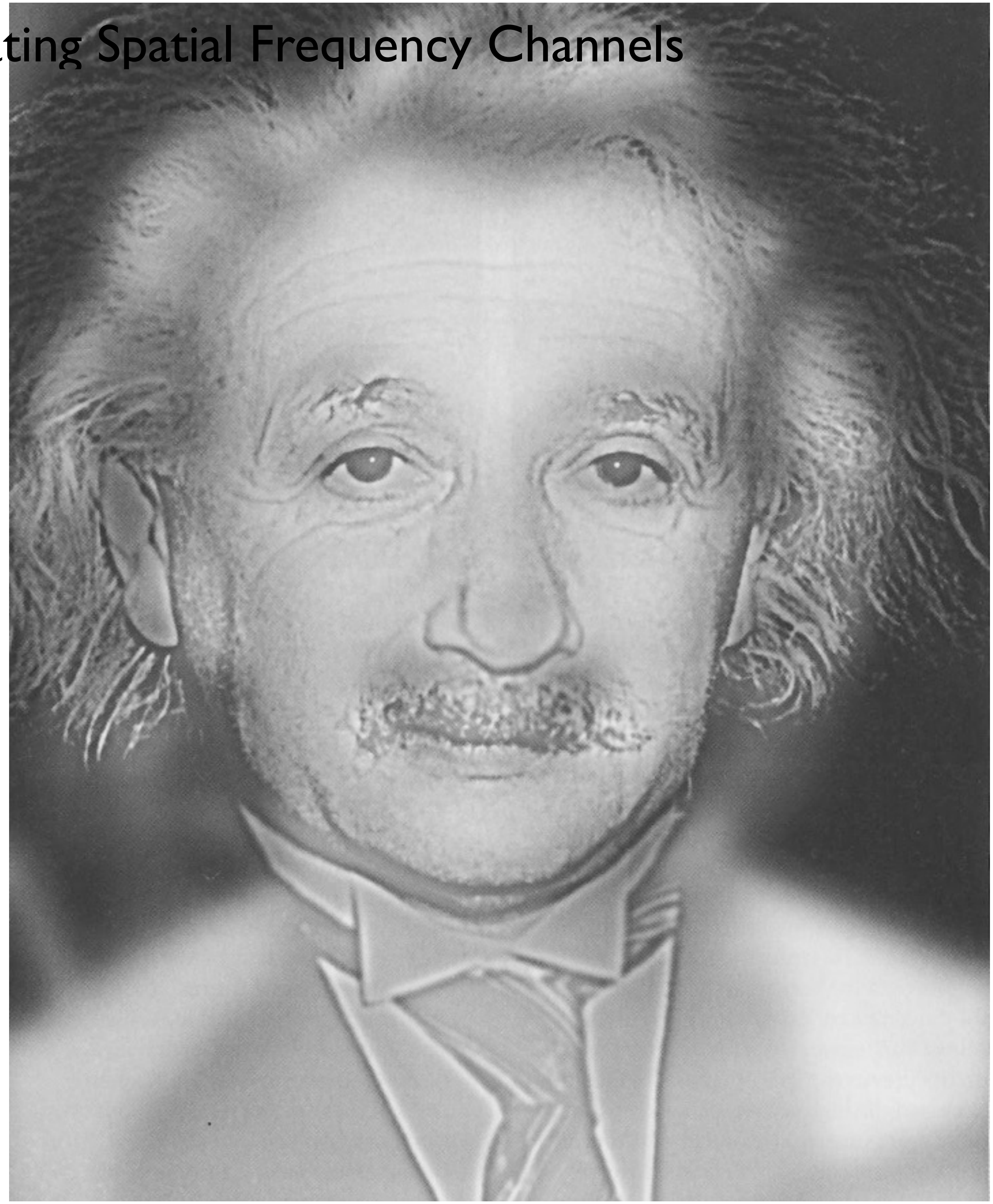
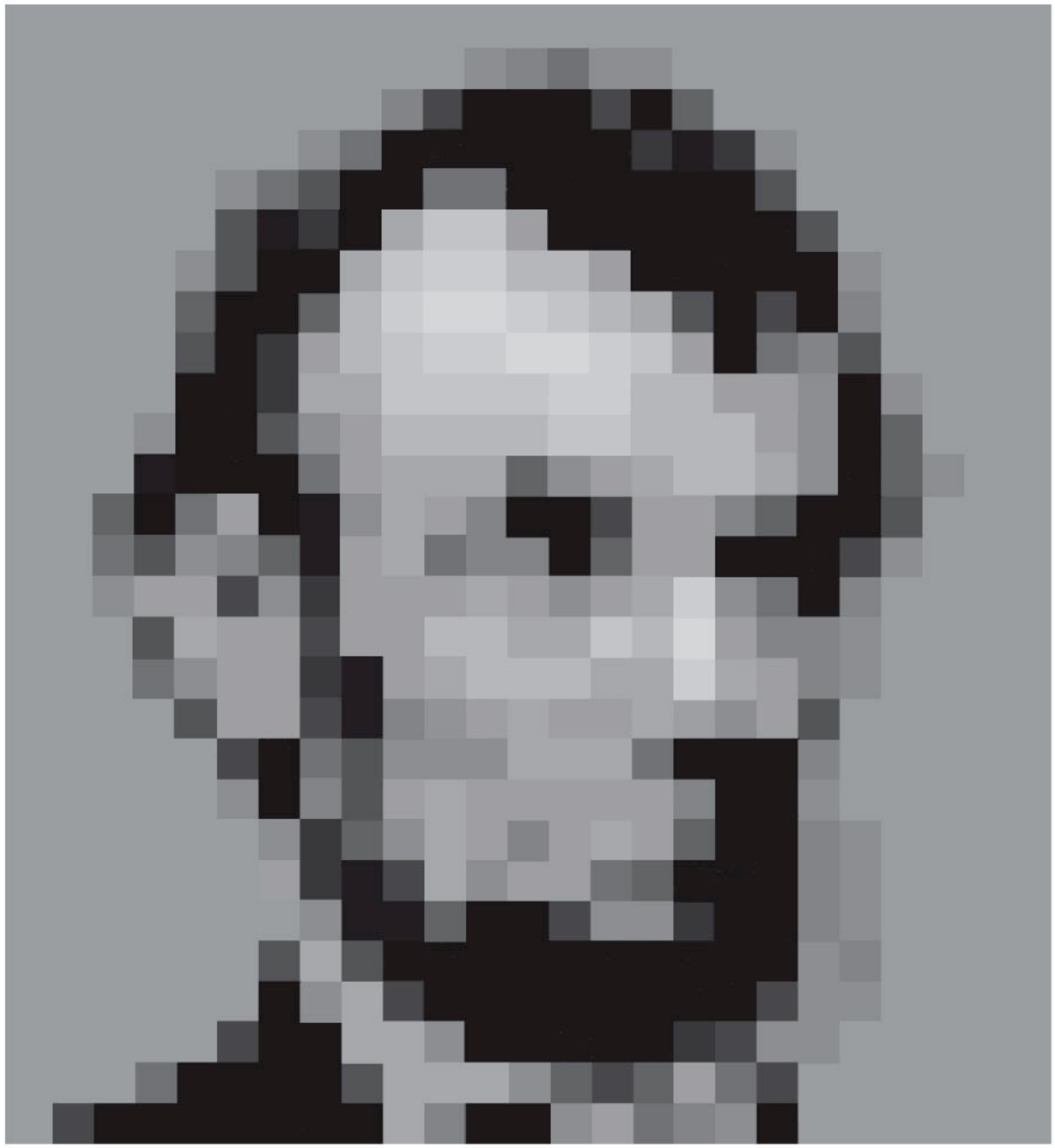


Image Illustrating Spatial Frequency Channels



If it is hard to tell who this famous person is, try squinting or defocusing



“Lincoln illusion” Harmon & Jules 1973

“Gala Contemplating the Mediterranean Sea, which at 30 meters becomes the portrait of Abraham Lincoln (Homage to Rothko)”

- Salvador Dali (1976)

“Gala Contemplating the Mediterranean Sea, which at 30 meters becomes the portrait of Abraham Lincoln (Homage to Rothko)”

- Salvador Dali (1976)

Problems with the Fourier Theory of V1

- Neurons in the visual system are broadly tuned to frequency (unlike Fourier decomposition)
- Fourier decomposition is *linear*. V1 is clearly nonlinear.
- Fourier decomposition is non-local; V1 is local.
  (So "wavelets" are maybe better than sine waves).
- Hasn't shed very much light on visual function
  (e.g., how does Fourier decomposition help explain object recognition?)

Sparse Coding Model

Olshausen & Field 1996

- V1 activity represents inferences under a "sparse" **generative** model of images

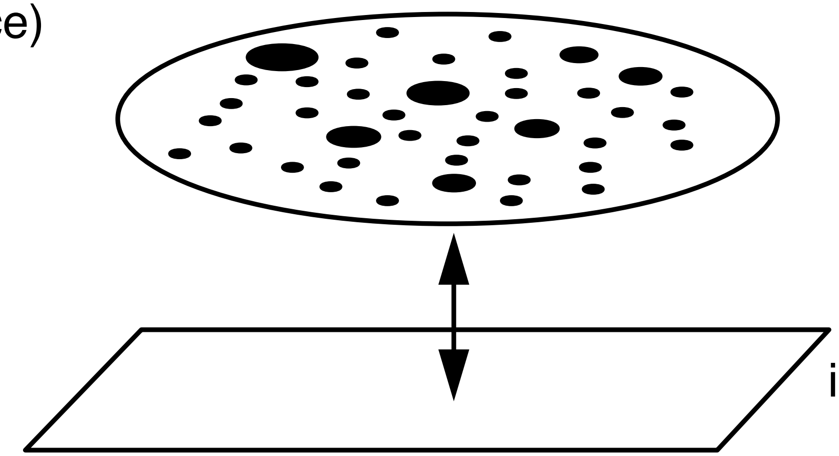
latent variables $y \sim$ sparse (eg., Laplace)

linear basis weights $\{B_i\}$

η noise (Gaussian)

images x

schematic



i'th neuron's "projective field"

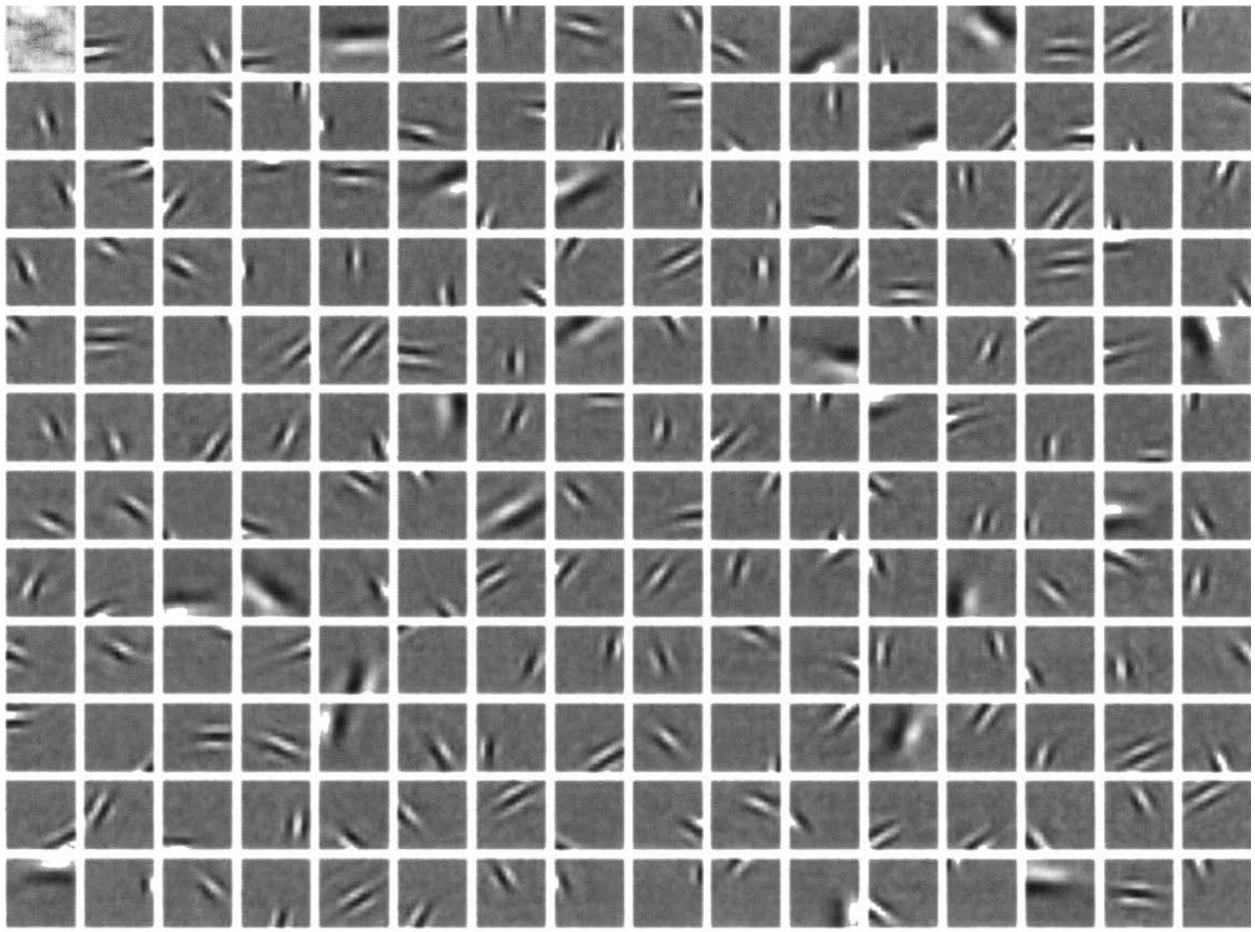
$$\{y_i\} = \arg\min_{\{y_i\}} ||x - \sum B_i y_i||^2 + \lambda \underbrace{\sum |y_i|}_{\text{sparsity penalty}}$$

V1 neural activity

image

i'th neuron's activity

The figure that launched a thousand papers



Olshausen & Field 1996

Projective fields B_i look like V1 receptive fields!

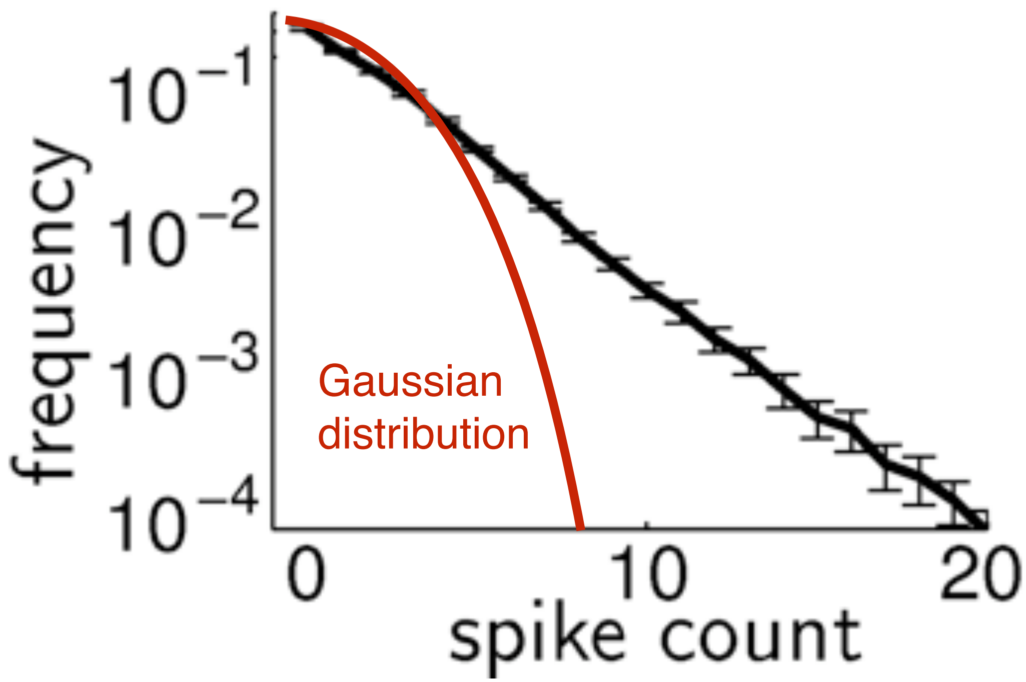
- first normative account of V1 receptive fields

(this doesn't happen if you run PCA on natural images!)

Spike responses are indeed sparse

Sparsity: "large spike counts are rare" OR "distribution is heavy-tailed"

Macaque IT neuron responses
(Baddeley 1997)



(proposed) Advantages of Sparsity

- **efficiency -** uses few spikes to encode any given stimulus (but you need many more neurons!)

- **computational convenience** - easier to decode (you only need to decode a few neurons for readout)

- **learning -** may facilitate learning via local update rules

independent components analysis (ICA)

- deterministic version of Sparse Coding (no noise)
- “complete”: # latent vars = # observed vars

latents $y \sim$ sparse

matrix A

data x

A is a square matrix

Inference is easy:

$$y = A^{-1}x$$

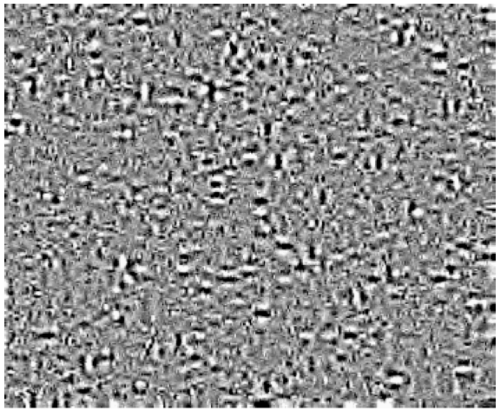
$$x = Ay$$

(so equally a “generative” model as a “recognition” model)

Bell & Sejnowski 1995

Limitations of sparse coding model:

- biology uses a cascade (what happens after V1?)
- why don't responses get sparser after V1?
  (Baddeley et al 97, Chechik et al 06)
- Sparse coding model is a linear generative model: doesn't provide very rich / accurate description of natural images

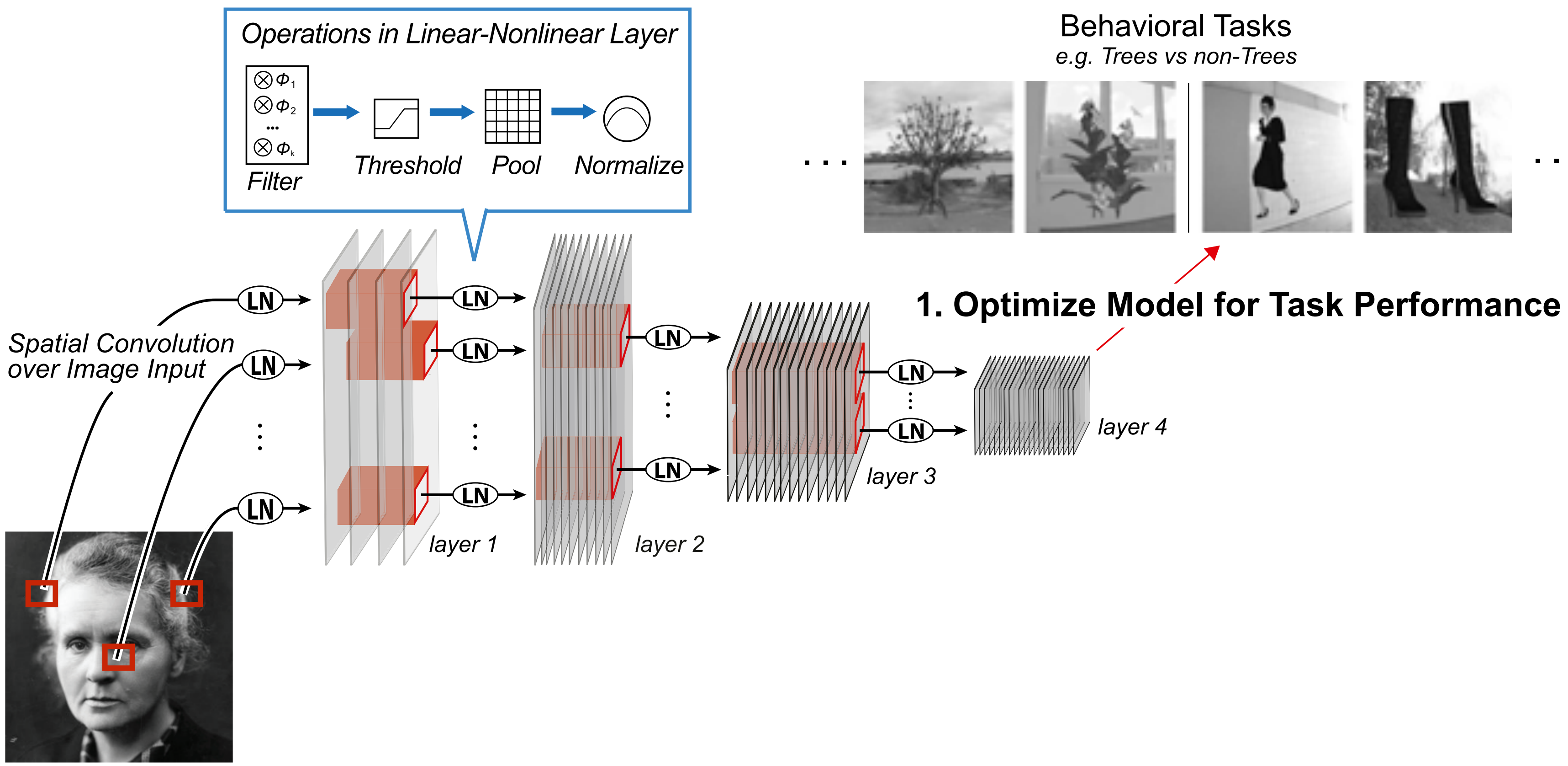


sample

Deep-learning based approaches

"task based" or "goal based" approaches:

1) train a network (CNN / DNN / RNN) to perform the task



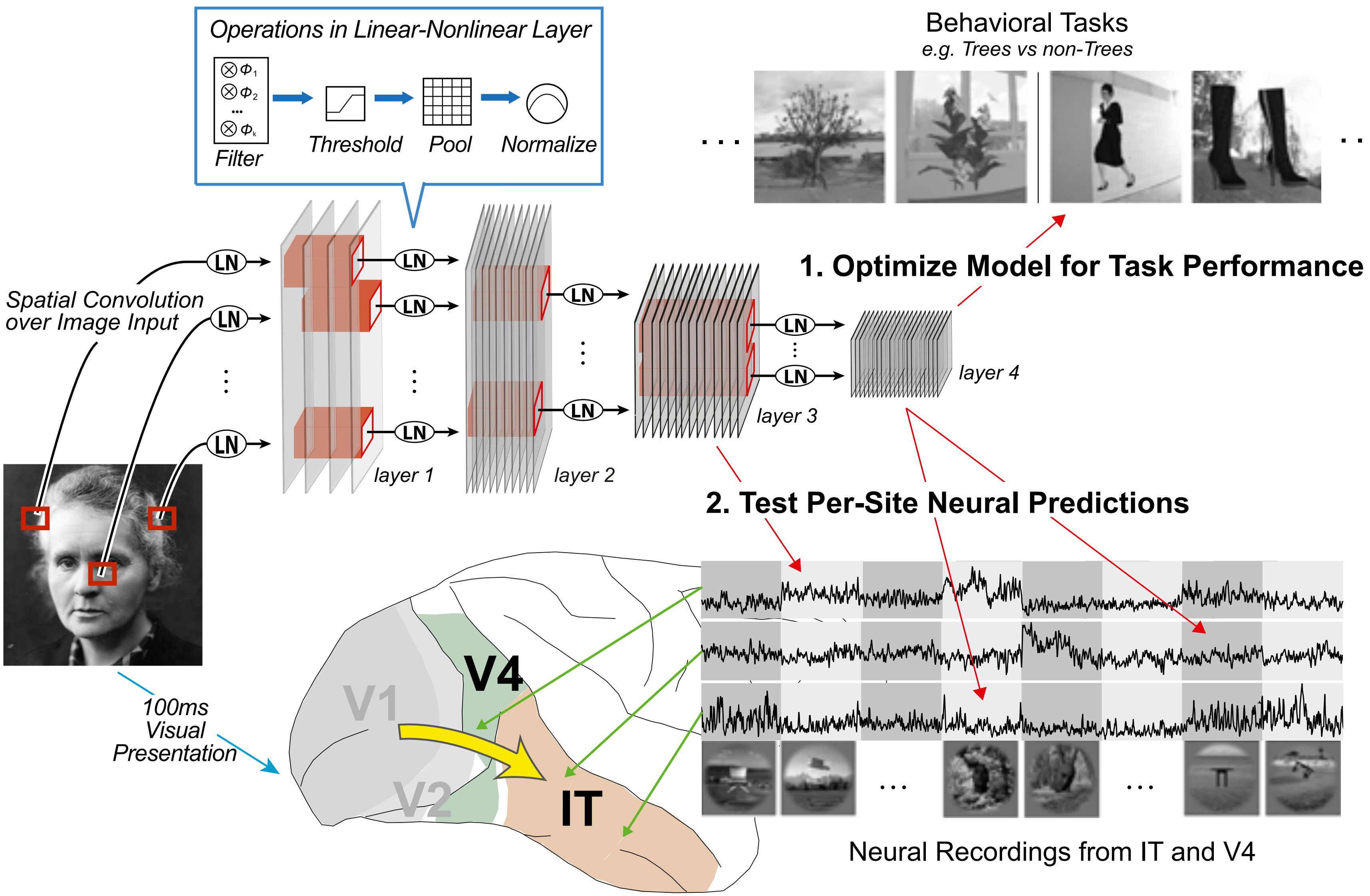
[Yamins et al, PNAS 2014]

Deep-learning based approaches

"task based" or "goal based" approaches:

1) train a network (CNN / DNN / RNN) to perform the task

2) regress units in trained network against neural data



[Yamins et al, PNAS 2014]

Lots of current research directions inspired by deep learning

- early work focused on pre-trained networks (AlexNet, VGG, ResNet)
- recent work shows we can fit models to neural data
- can use DNNs to synthesize images to optimally drive neurons
- use RNNs / LSTMs / GRUs to capture time-course of responses
- can we improve these models by incorporating other ideas from biology (feedback, spiking, divisive normalization, plasticity, etc.)?
- current debate about whether we can ever “understand” V1 (or whether that is even a worthwhile goal)

Summary

- retinal organization: photoreceptors (rods and cones), dark current, bipolar cells, retinal ganglion cells (RGC)
- receptive fields, "ON" and "OFF" receptive fields
- Barlow's efficient coding hypothesis
- Hubel & Weisel: orientation tuning in V1
- ocular dominance
- simple / complex cells
- edge detection
- Fourier analysis, spatial frequency channels
- sparse coding model
- deep-learning based approaches