南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Machine Learning and NeuroEngineering

机器学习与神经工程

Lecture 17 – Human Visual System & CNN

Quanying Liu （刘泉影）

SUSTech, BME department

Email: liuqy@sustech.edu.cn

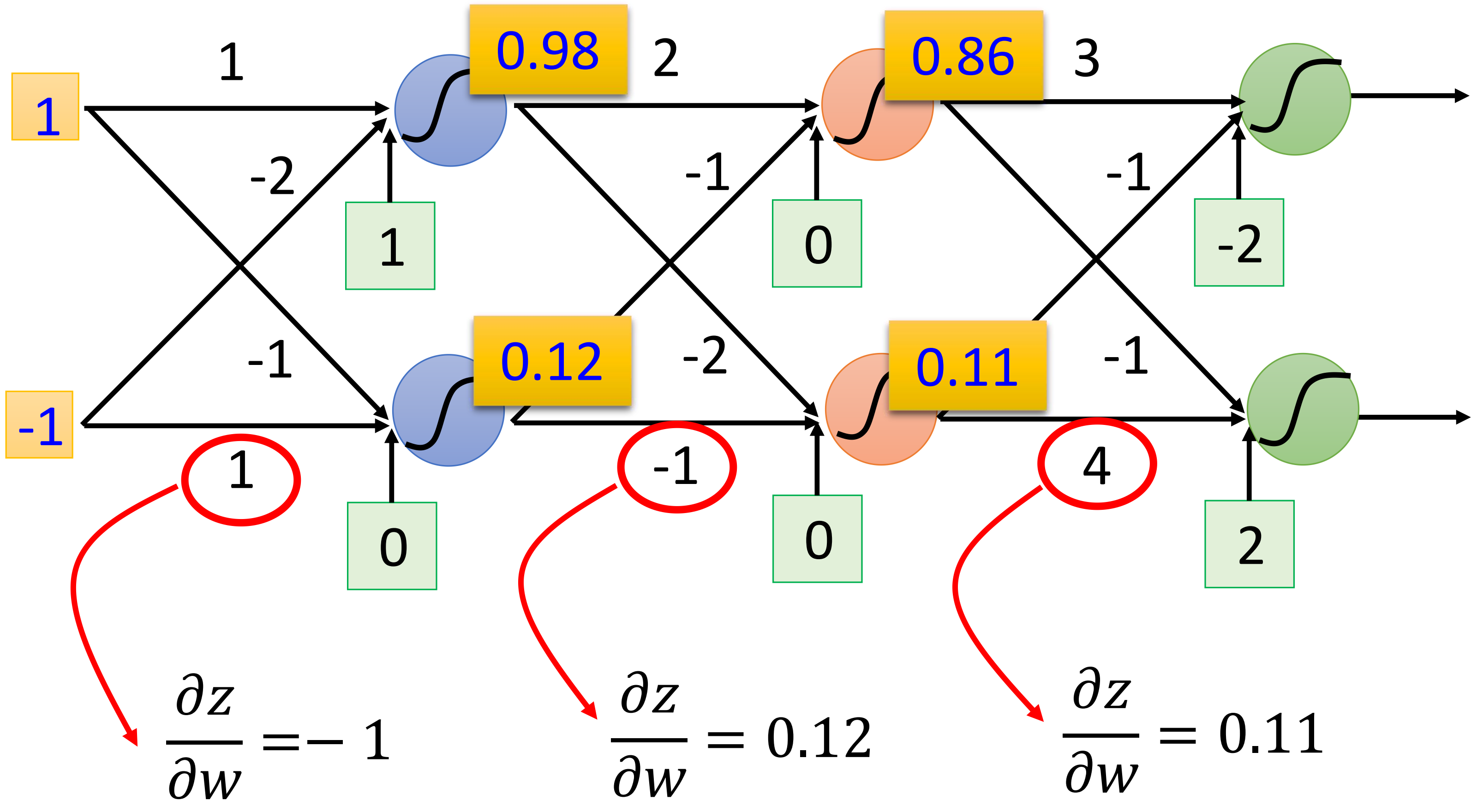
Lecture 15 - Recap

- Gradient Descent (GD)
  - What is Gradient Descent?
  - Gradient Descent to train deep NNs → Error Backpropagation
- Error Back-propagation (BP)
  - Backpropagation
  - Backpropagation – forward pass
  - Backpropagation – backward pass
- The Architecture of CNN
  - Convolution
  - Activation function
  - Pooling
  - Flatten
  - FC
- CNN Hands-on (tensorflow), thanks to 曲由之--> next lecture

Some slides are adjusted from 李宏毅

Backpropagation – Forward pass

Compute $\partial z/\partial w$ for all parameters

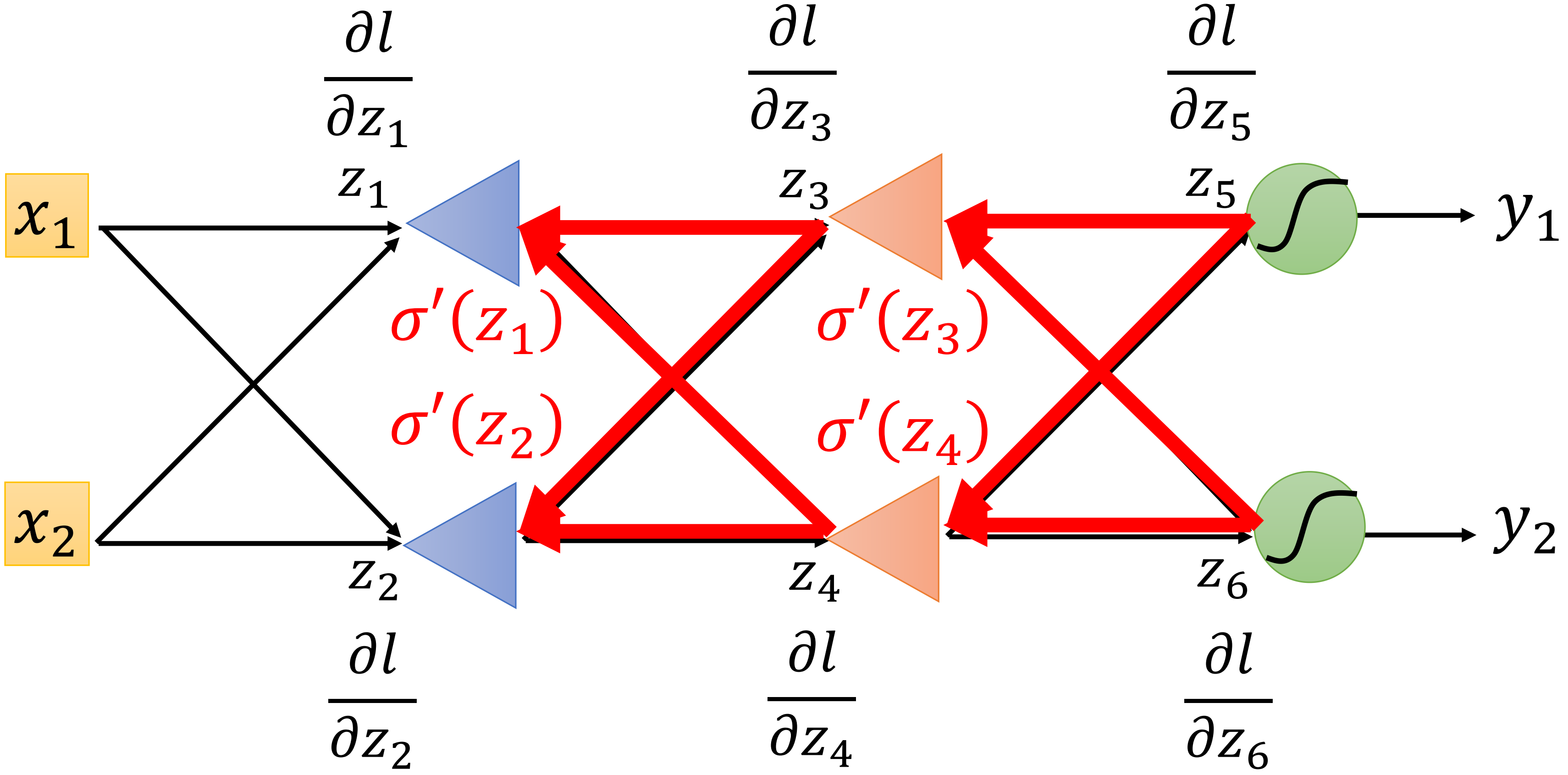


Backpropagation – Backward pass

Compute $\partial l / \partial z$ for all *activation function* inputs z

Compute $\partial l / \partial z$ from the output layer

$$\frac{\partial l}{\partial z} = \sigma'(z)\left[w_3 \frac{\partial l}{\partial z_1} + w_4 \frac{\partial l}{\partial z_2}\right]$$



The Architecture of CNN

Property 1

- Some patterns are much smaller than the whole image

Property 2

- The same patterns appear in different regions.

Property 3

- Subsampling the pixels will not change the object

Convolution

Activation function

Pooling

… …

Flatten

Property 4

Hierarchical architecture

These blocks can repeat many times.

Reverse-engineer the brain

to understand the design principles of brain

The Marr's three level of explanation

Level 1: Computation theory

- What is the problem to be solved?
- What are the inputs and outputs to the computation?
- What is the goal, and what is the logic by which it is carried out?

Level 2: Algorithmic

- How is the information represented and processed to achieve the computational goal?

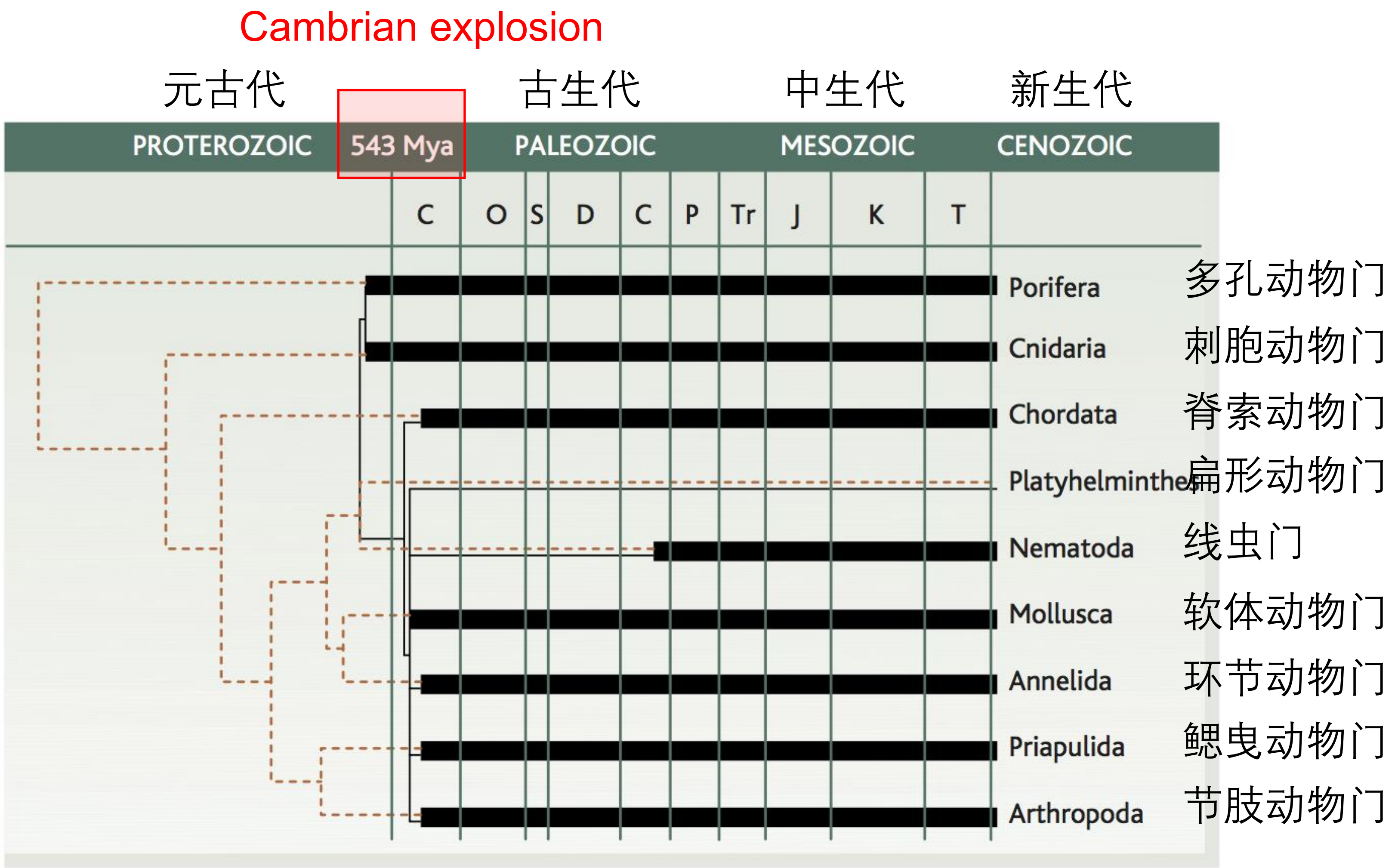
Level 3: Implementation

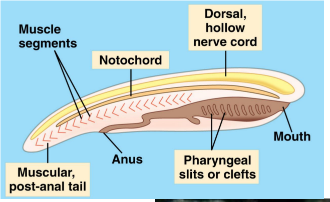
- How is the computation realized in physical or biological hardware?

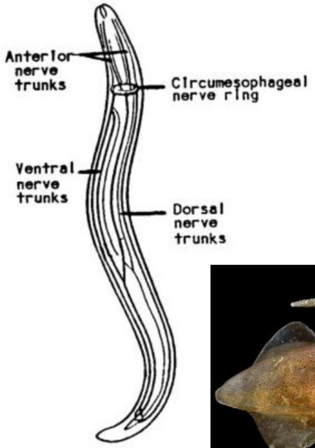
Lecture 17 – Human Visual System & CNN

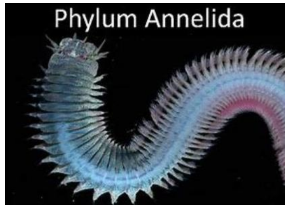
0. Marr's 3 levels of explanation
1. Evolution of the eye
2. Function of the visual system
3. Structure of the eye
4. Photoreceptors
5. Information integration by ganglion cell
6. **Visual pathways**: photoreceptors, interneurons, ganglion cells, LGN, V1, ventral/dorsal pathways
7. **Image process: brain vs. CNN**

Cambrian explosion







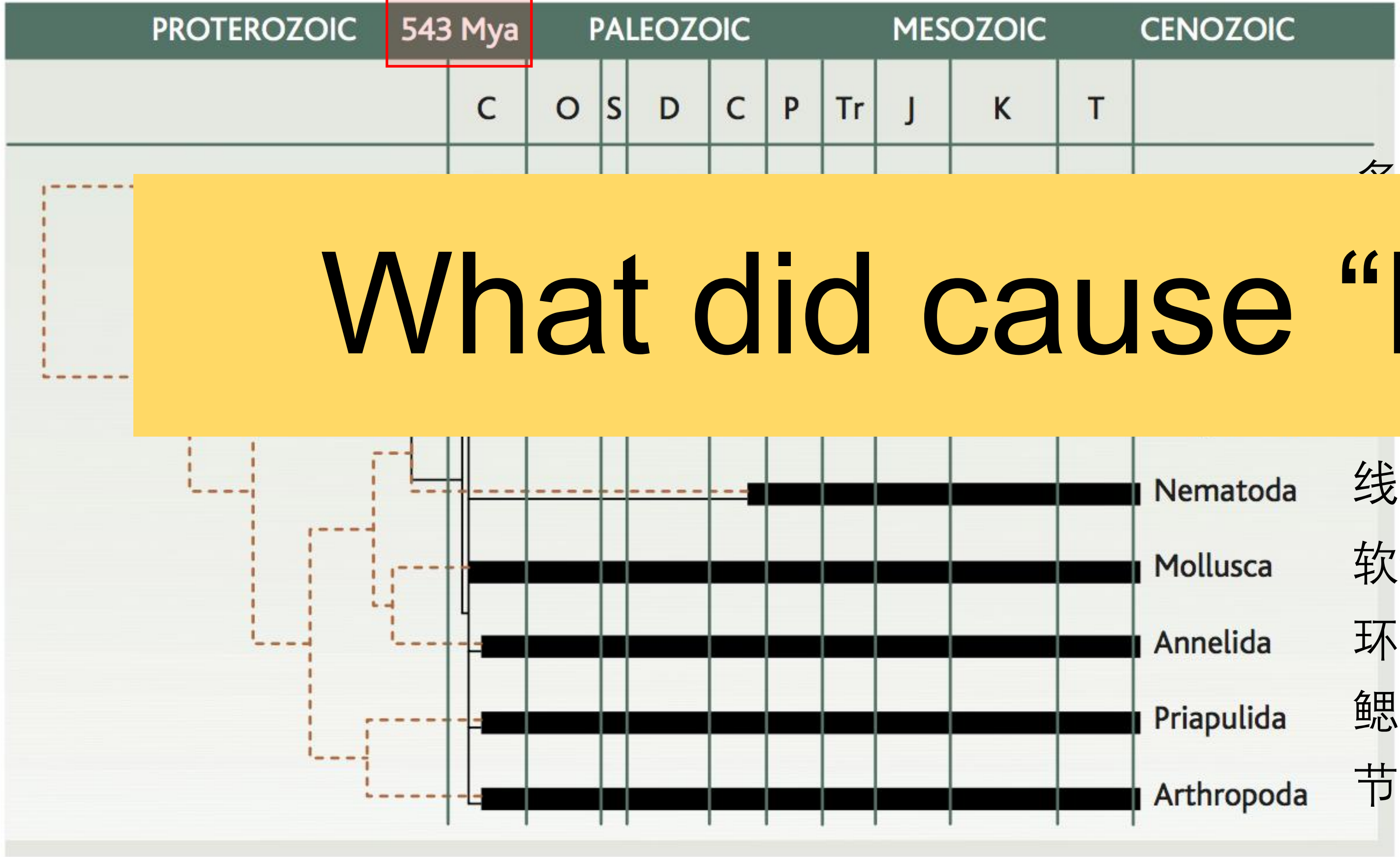


Jermiin, Poladian, Charleston (2005). Is the "Big Bang" in Animal Evolution Real? *Science*

Cambrian explosion

Cambrian explosion

元古代 古生代 中生代 新生代



What did cause "**Big Bang**"?

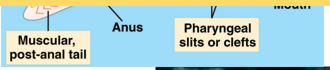
线虫门

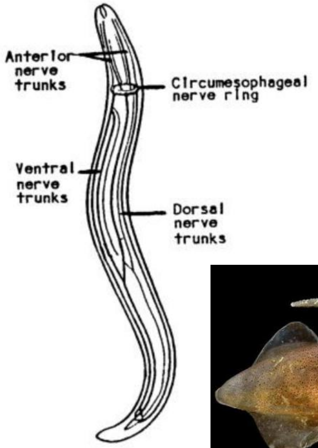
软体动物门

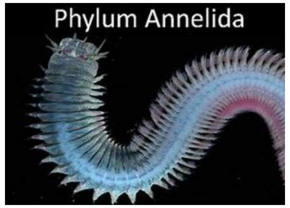
环节动物门

鳃曳动物门

节肢动物门

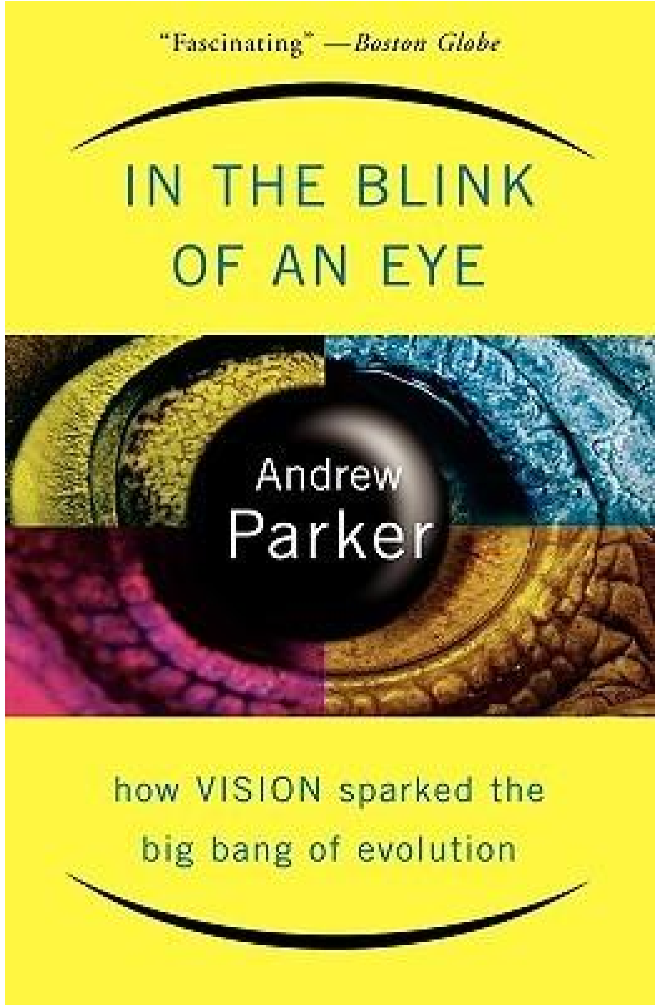






Jermiin, Poladian, Charleston (2005). Is the "Big Bang" in Animal Evolution Real? *Science*

Evolution of the Eye



Andrea Parker
In the Blink of an Eye

Eyes first appeared approximately 543 million years ago during the Cambrian period—the geological period that marks the rapid increase in biodiversity.

Oxford zoologist Andrew Parker proposed "Light Switch Theory'", suggesting that it was the development of **vision** in primitive animals that caused the explosion.

Precambrian creatures were **unable to see**, making it impossible to find friend or foe.

With the evolution of the **eye**, the size, shape, color, and behavior of animals was suddenly revealed for the first time. Once the lights were "turned on," all animals had to either adapt or die, and in a geological instant, the world became a very different place.

Function of the Visual System

The visual system is a part of the central nervous system (CNS) that gives organisms the ability to **detect**, **process**, **interpret** information from **visible light**, with the goal of building a **representation** of the surrounding environment.

NOT a camera:

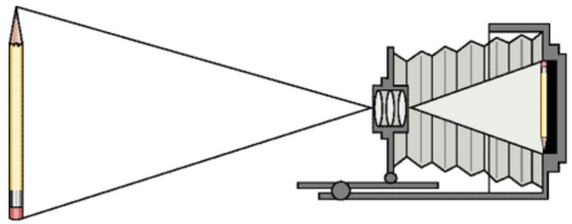
the visual system solves the inverse problem.

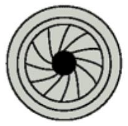
Detect
Process
Interpret

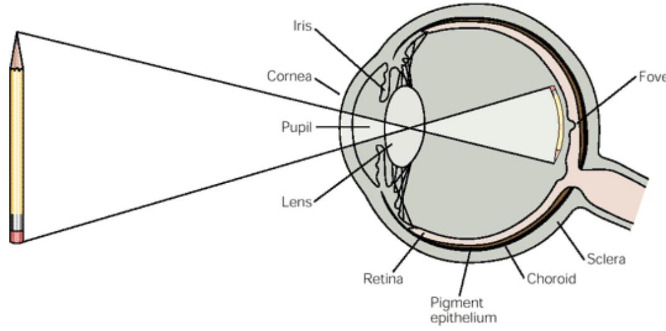
Forward problem

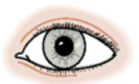
Image formation

Camera







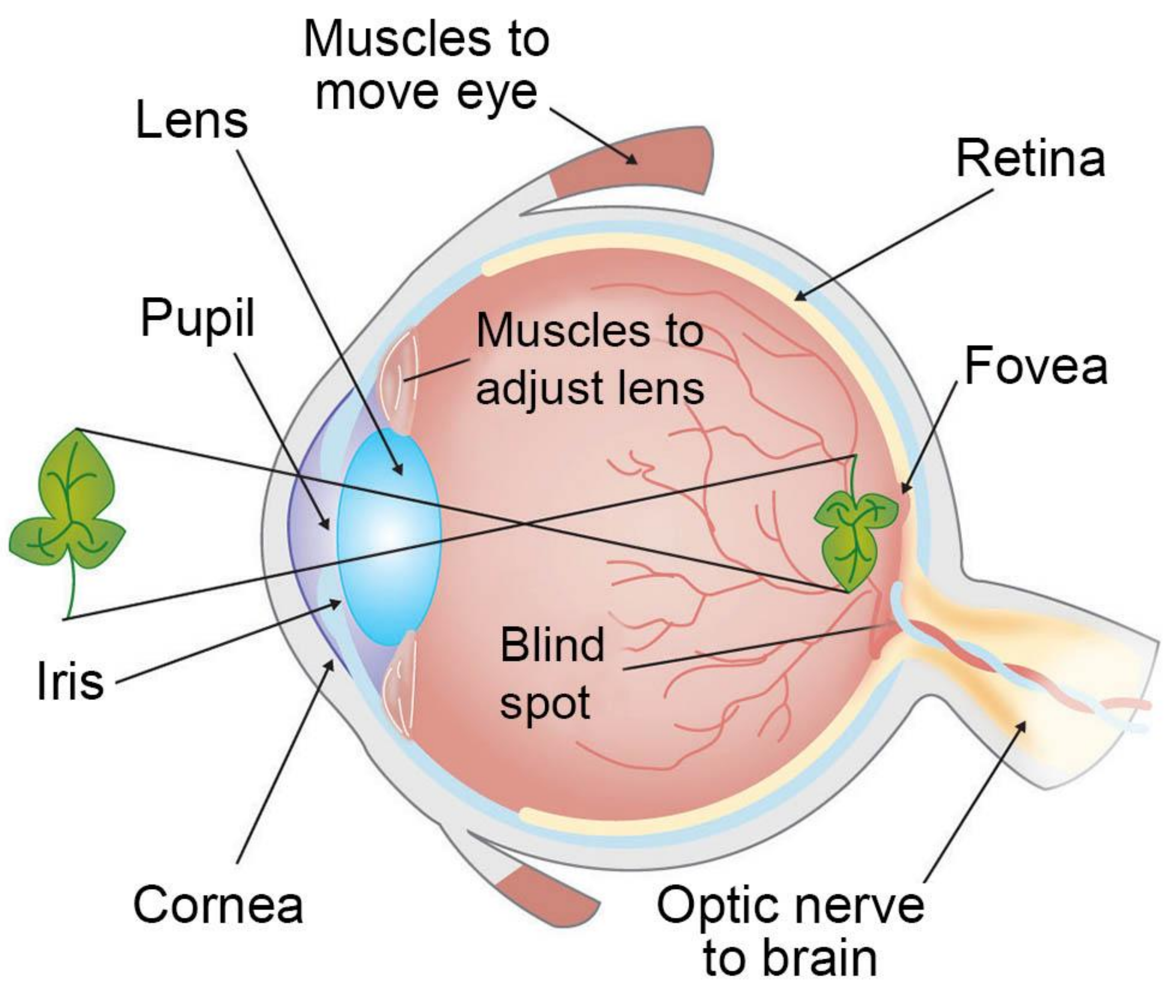


Eye +
LGN +
V1 +
…

Inverse problem

Vision (brain)

Structure of Eye



Lens + Cornea: help focus light onto the eye, like camera lenses.

Photoreceptors (感光细胞) in **Retina**: convert light energy into neuronal activity, like camera sensors.

Pupil: to adjust based on illumination of light, like aperture of the camera.

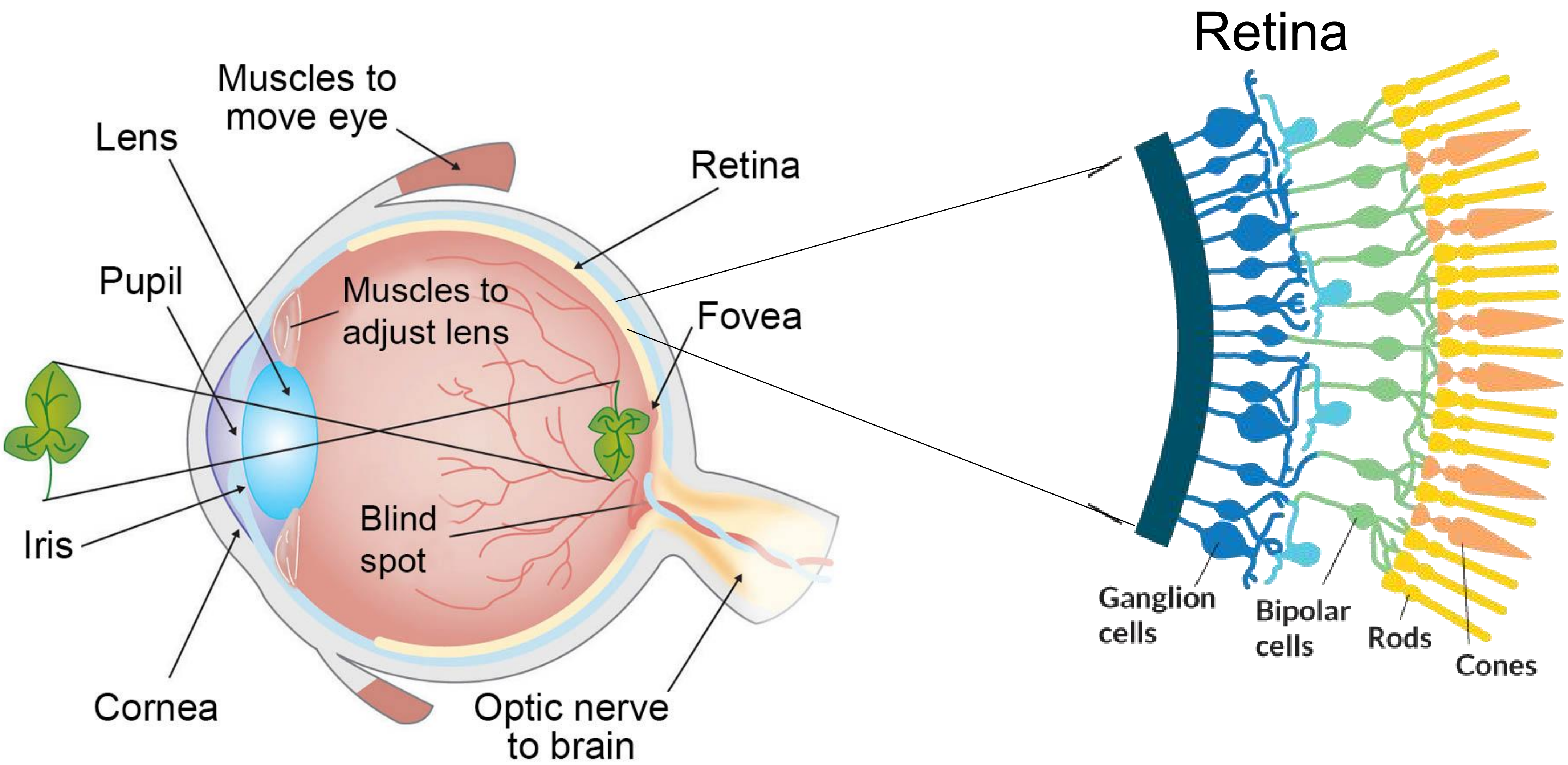
Optic nerve: transmit the image signal to the brain, like camera electronics.

Eye can **clean itself**, by producing tears, blinking.

Eye can even **repair itself** in the event of modest damage to its optic.


including many lenses
arrayed across the eye.

Structure of Eye



Diversity in the cell types

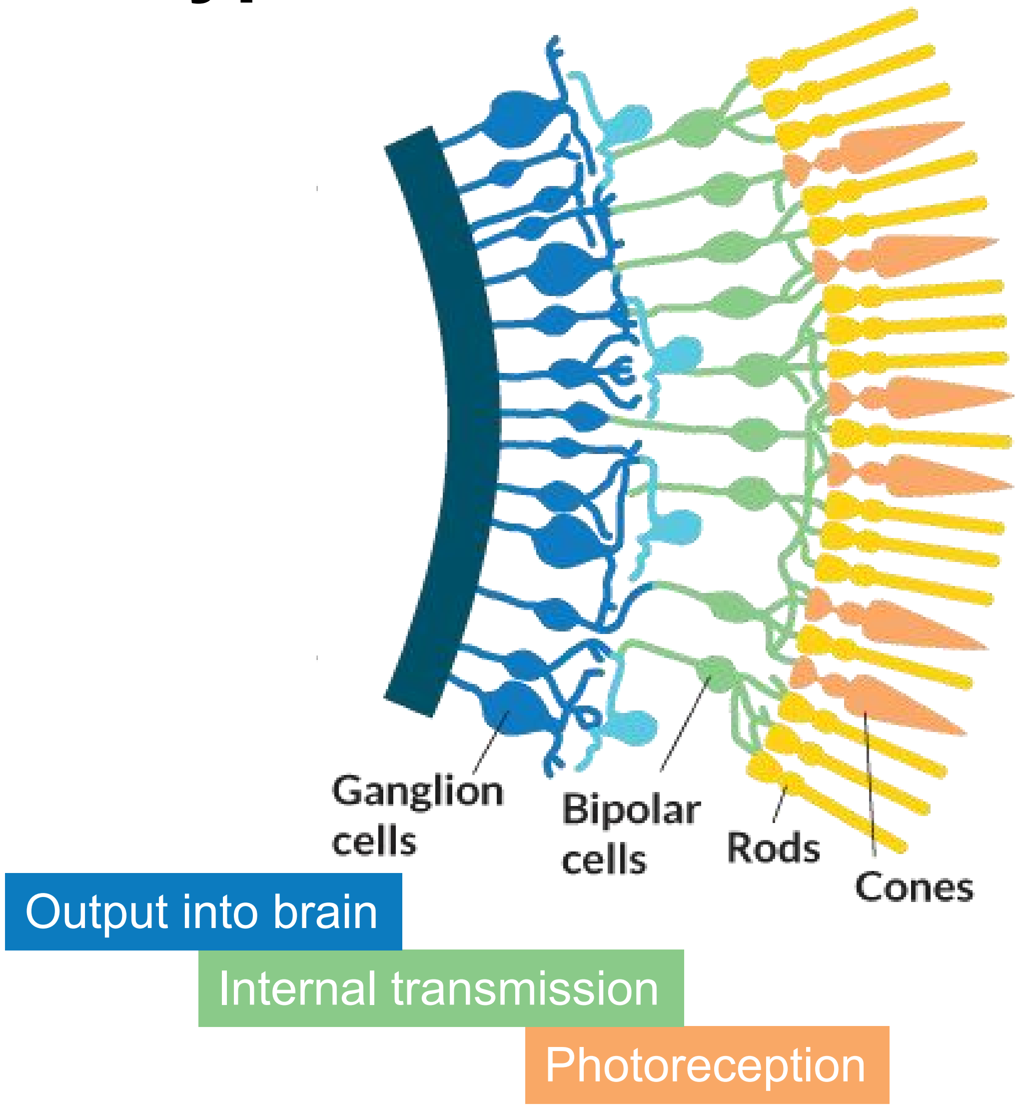
Interneurons: **> 50** types

Retinal ganglion cells: **> 30** types

Photoreceptors: **2** types (rods and cones)

Vote:

Do you think this diversity is well designed by nature, or happens by chance?



Arranged in an inside-out fashion

Photoreceptors

> 50 types of **interneurons**

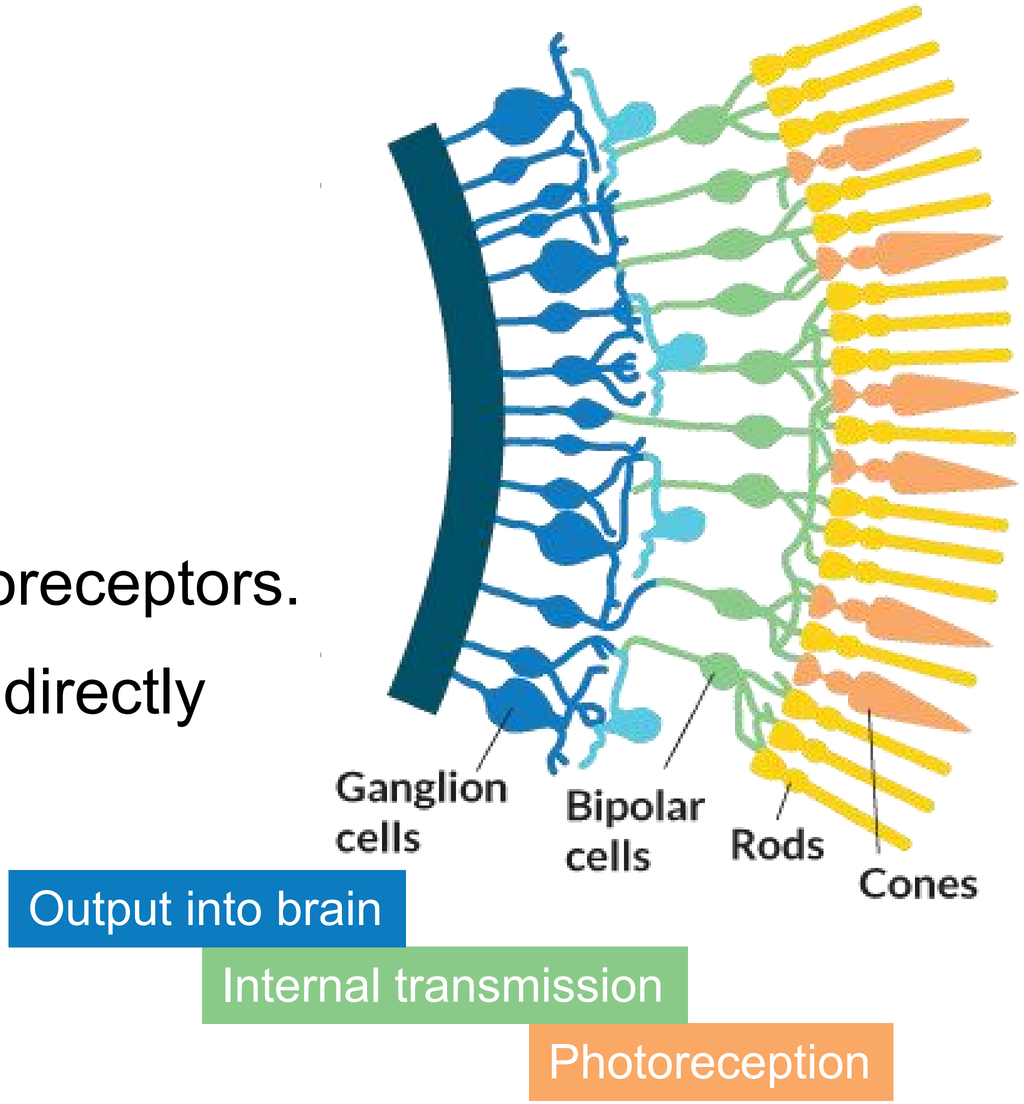
> 30 types of **retinal ganglion cells**

2 types of **photoreceptors**: rods and cones

In 2002, scientists found a **new** type of photoreceptors.

The newly discovered cells turn light energy directly into brain signals.

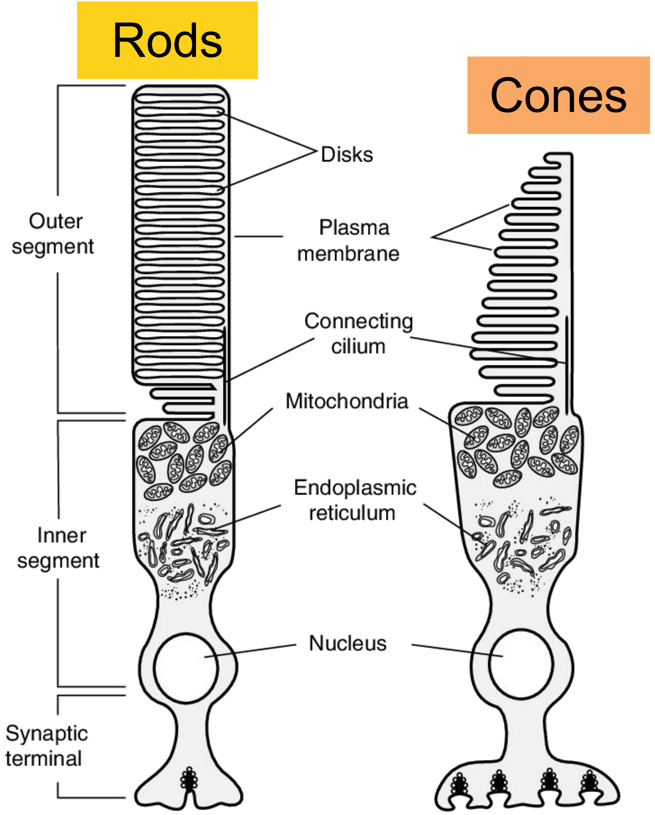
The signals govern the body's 24-hour clock.



Arranged in an inside-out fashion

Berson et al (2002). Phototransduction by Retinal Ganglion Cells That Set the Circadian Clock, *Science*

Rods and Cones



Rods (120 millions): more sensitive to light than the cones, not sensitive to color. (night)

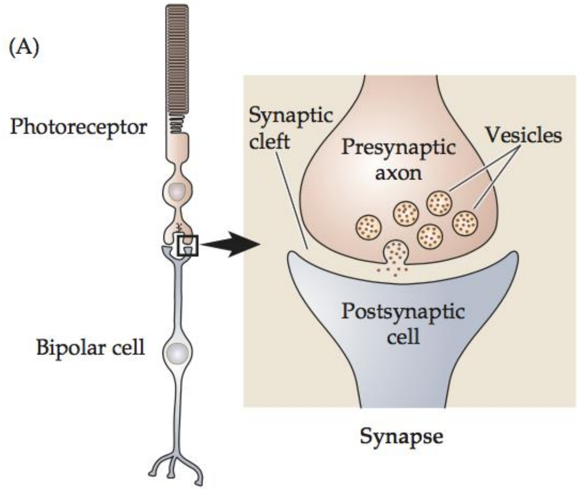
Cones (6~7 millions): color sensitivity, concentrated in the central yellow spot known as the macula. (day)

Signal conversions

Electromagnetic → Chemical: Outer segment

Chemical → Electrical: Inner segment

Electrical → Chemical: Synaptic terminal



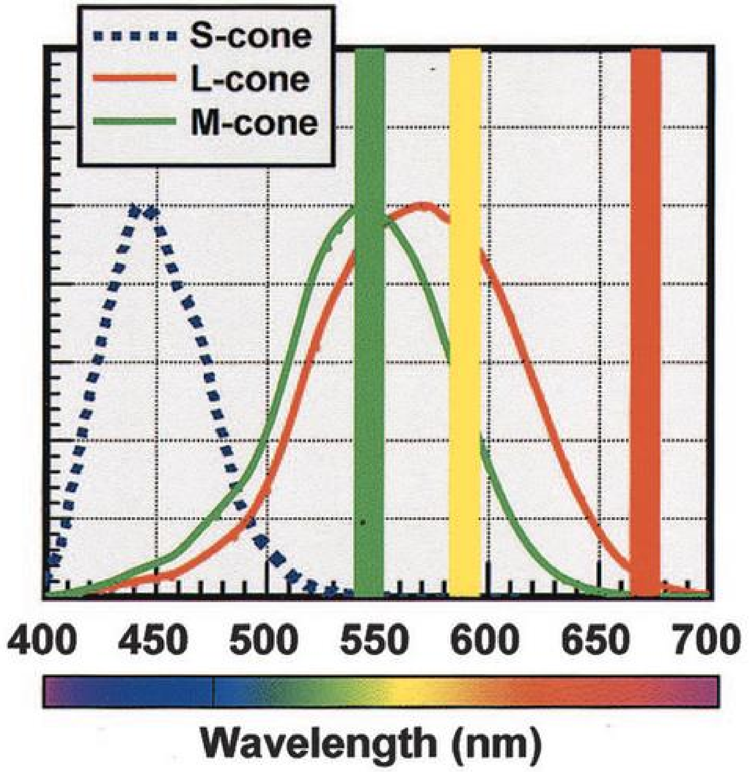
Three Types of Cones

Rods: have a peak sensitivity at 498 nm wavelength.

Short-wave cones: 420-440 nm, detecting blue light

Middle-wave cones: 530-550 nm, detecting green light

Long-wave cones: 565-580 nm, detecting red light



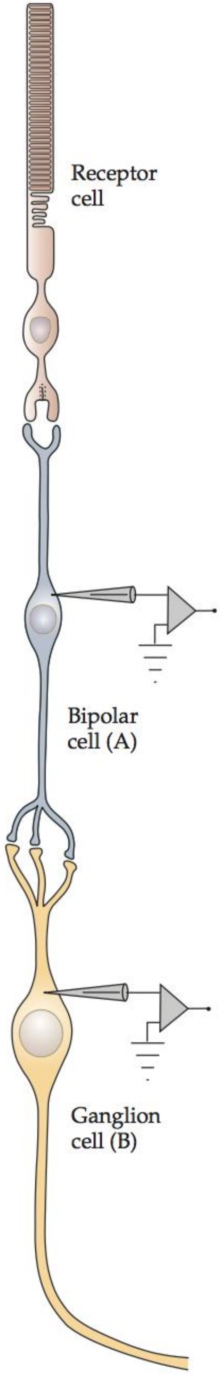
The color represented as a combination of these 3 types of cones, just as the way of RGB color system.

Lack one type of cones → Color blindness

The lack of m-cone lead to red-green blindness.

Fish and birds have 4 types of cones used for vision.

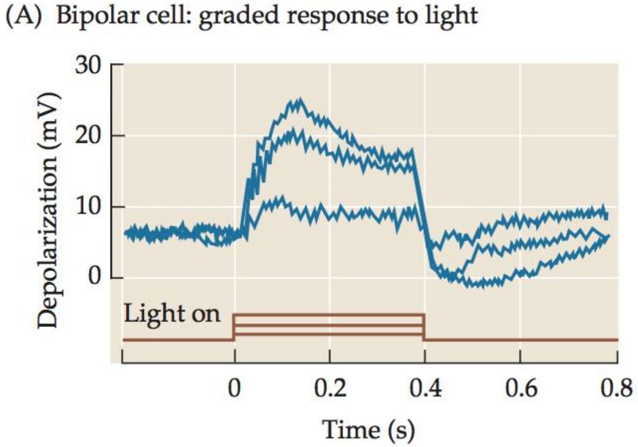
Action Potential generated by Ganglion cells

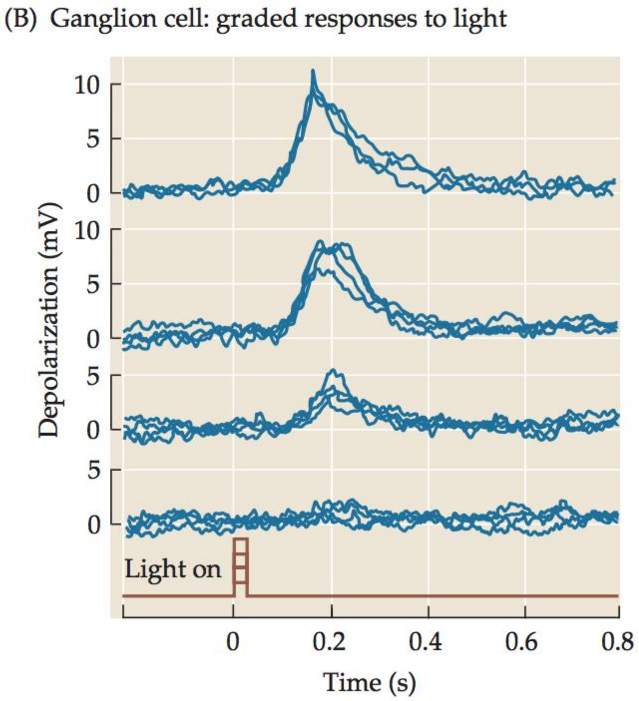


Photoreception

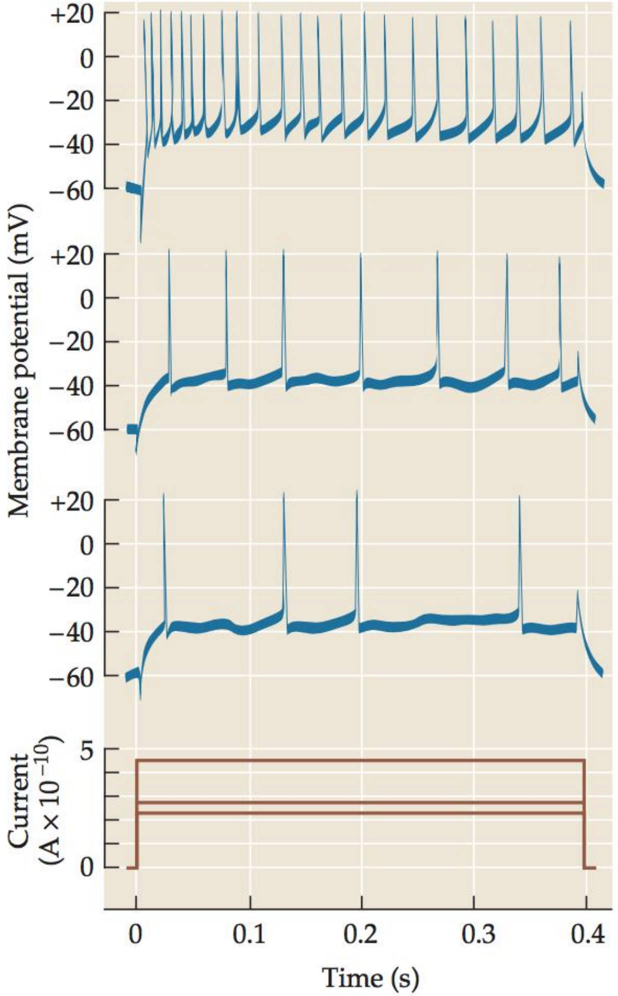
Internal transmission

Output into brain



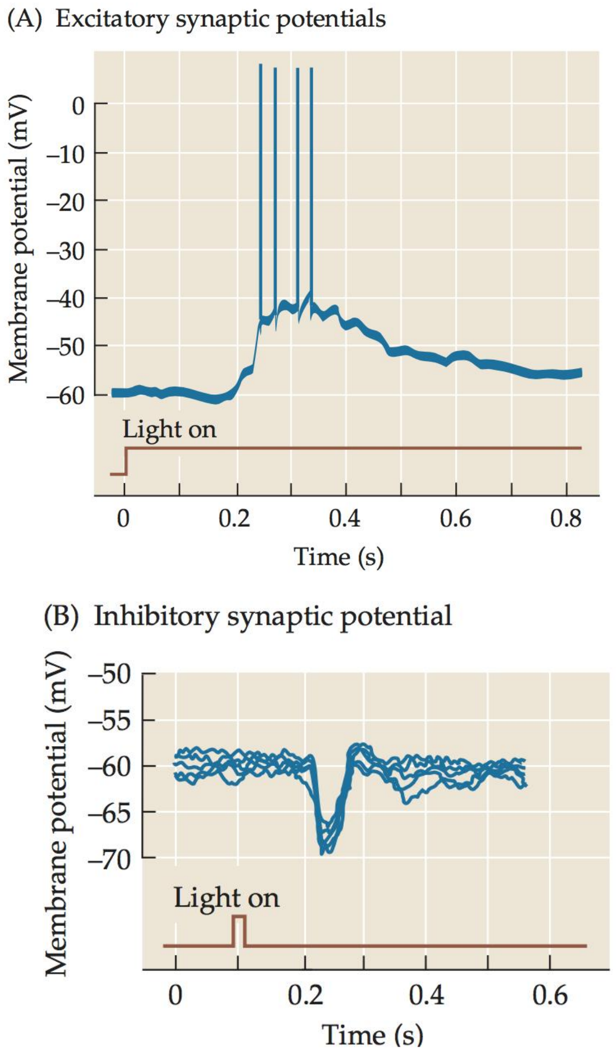


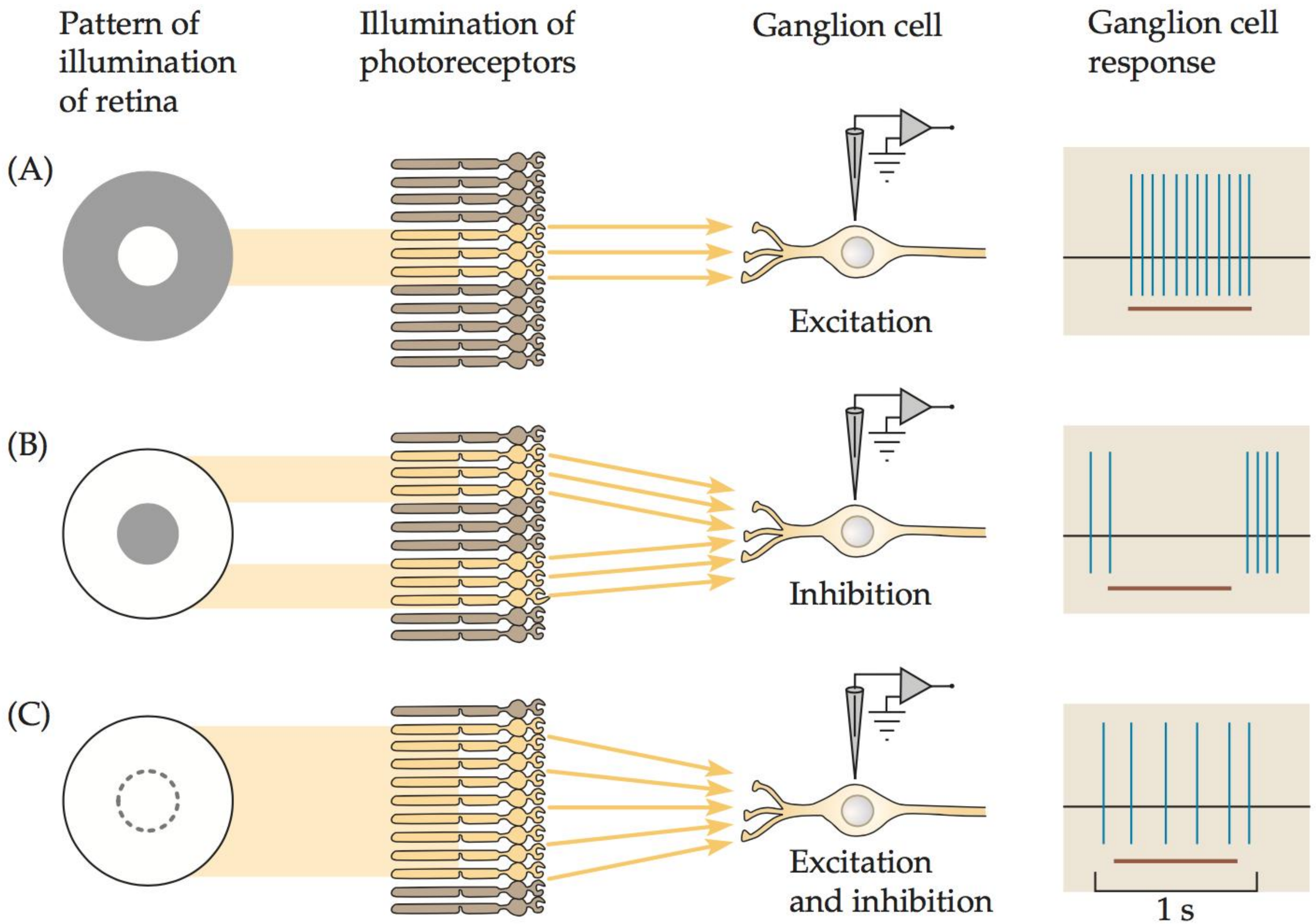
Action potentials are necessary to carry information along the length of a cell.



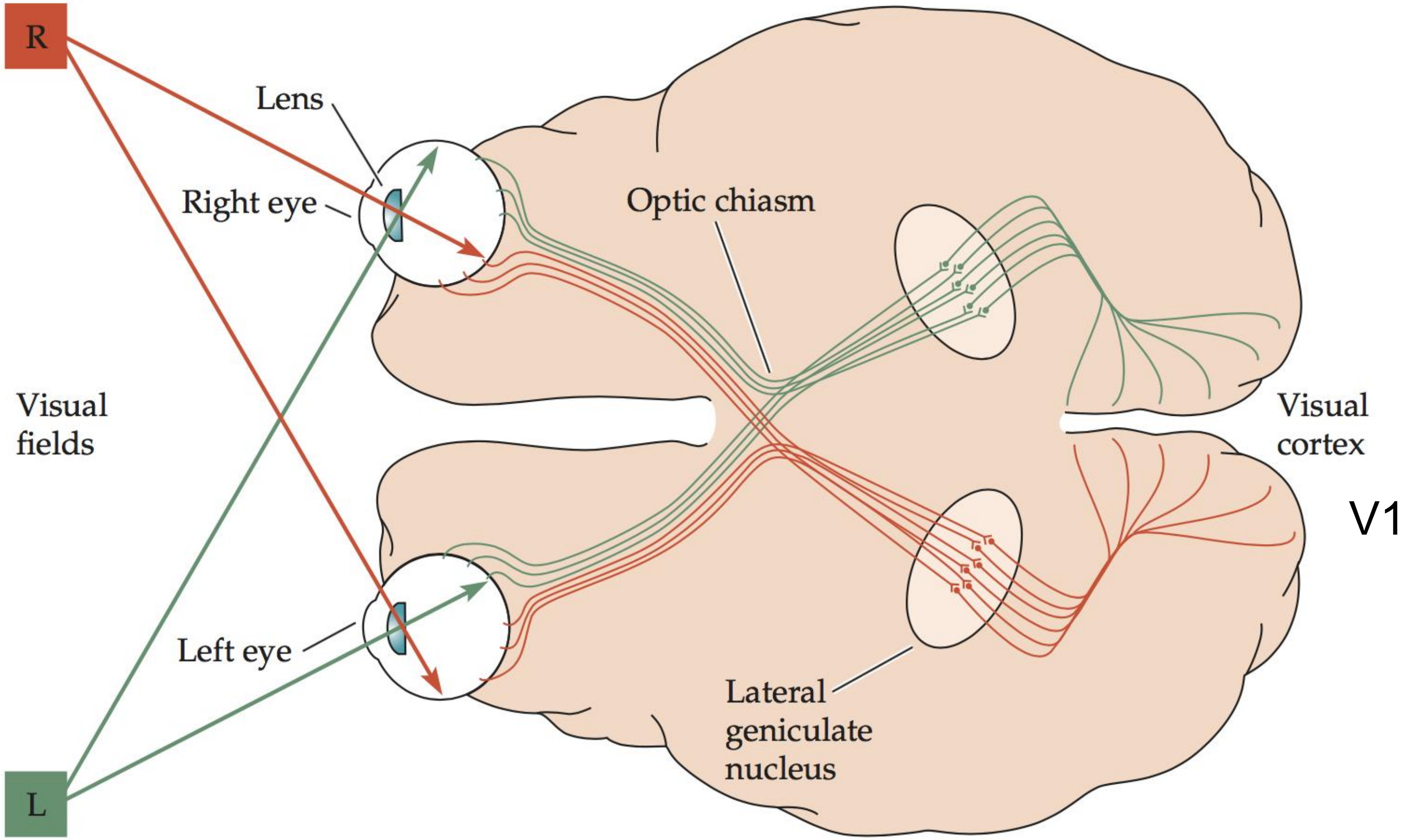
Information Integration by Ganglion Cell

On-center cells





Visual pathways: from the eye to the cerebral cortex



LGN in thalamus

Rods, cones

Interneurons

Ganglion cells

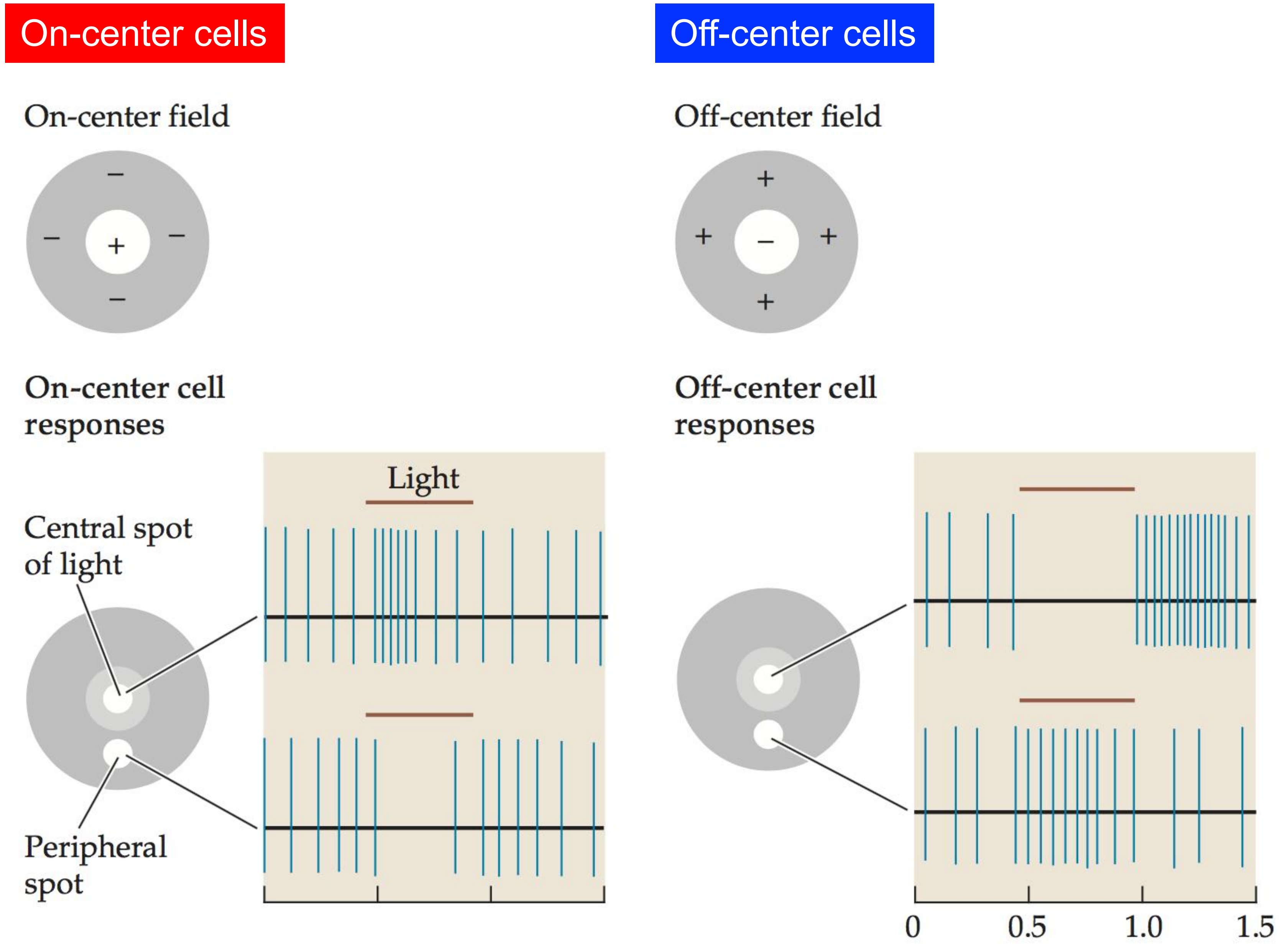
Optic fiber

LGN

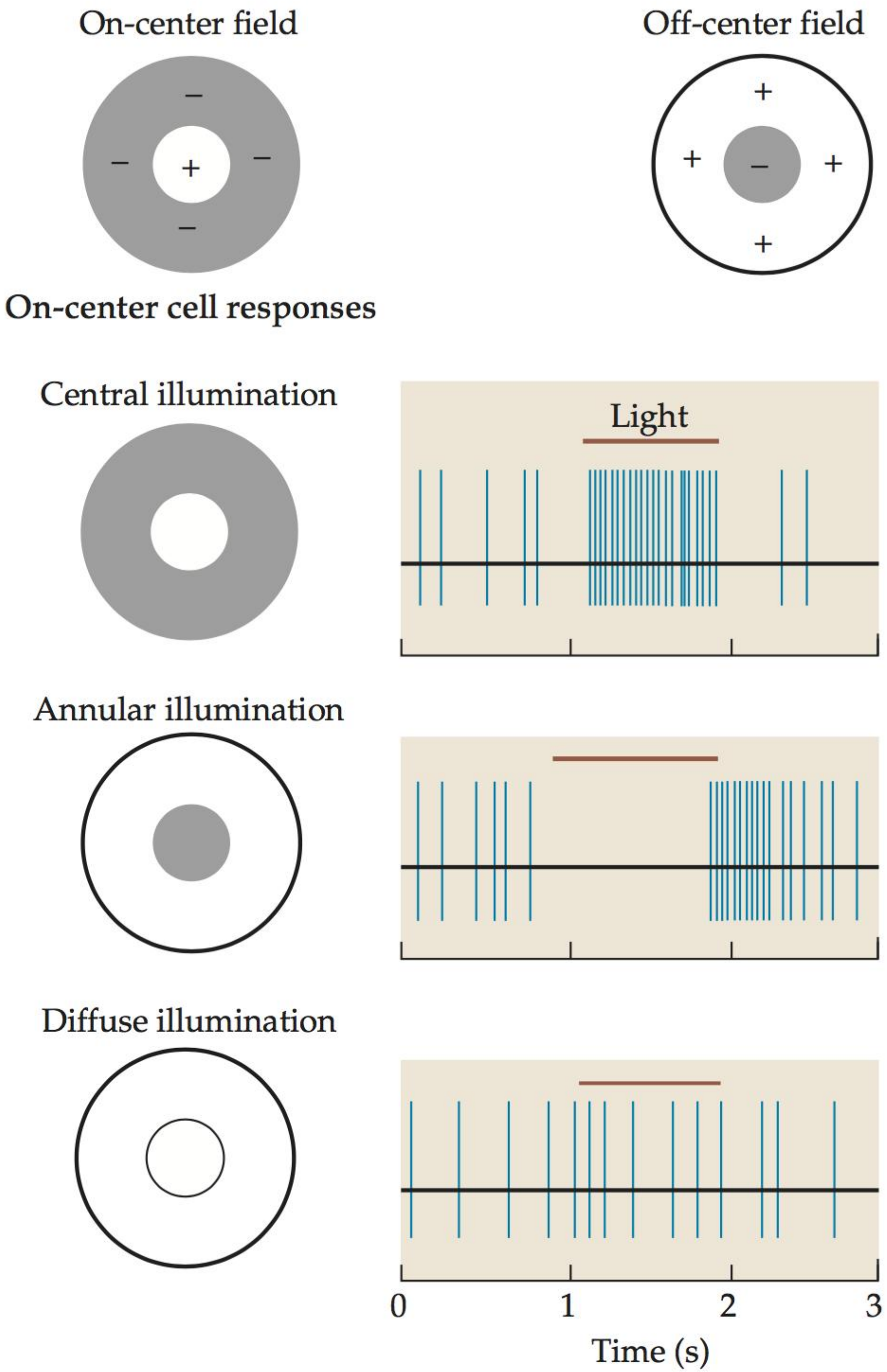
V1

Ventral / dorsal pathways

Ganglion Cells

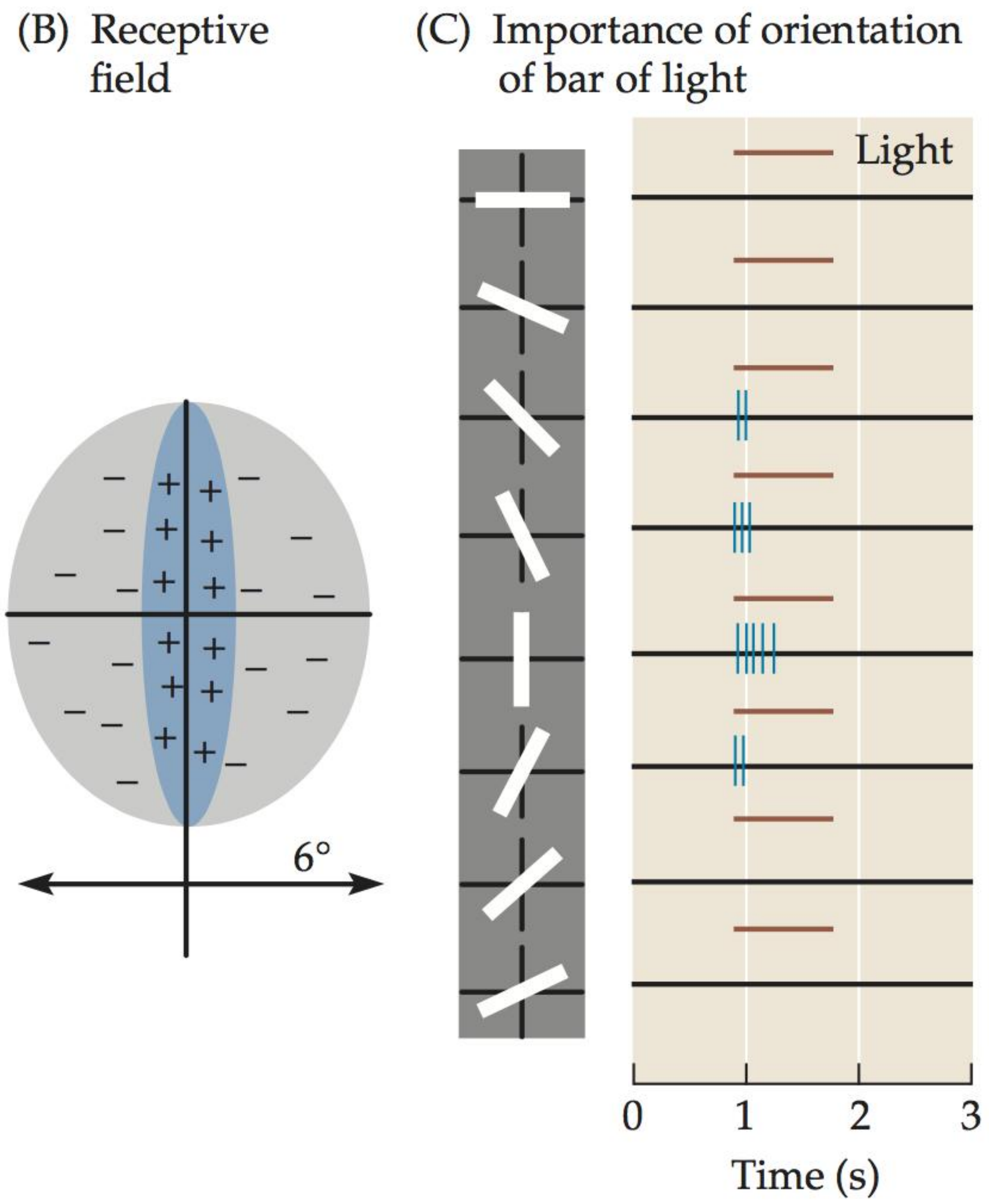


LGN Cells

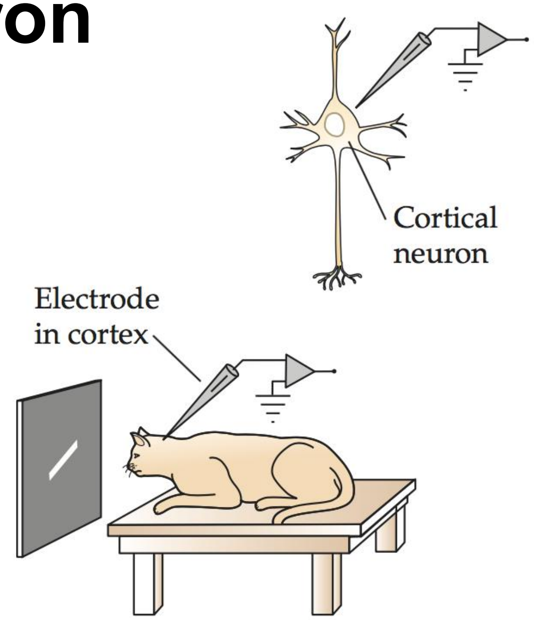


On-center field
Off-center field
On-center cell responses
Central illumination
Light
Annular illumination
Diffuse illumination
0
1
2
3
Time (s)

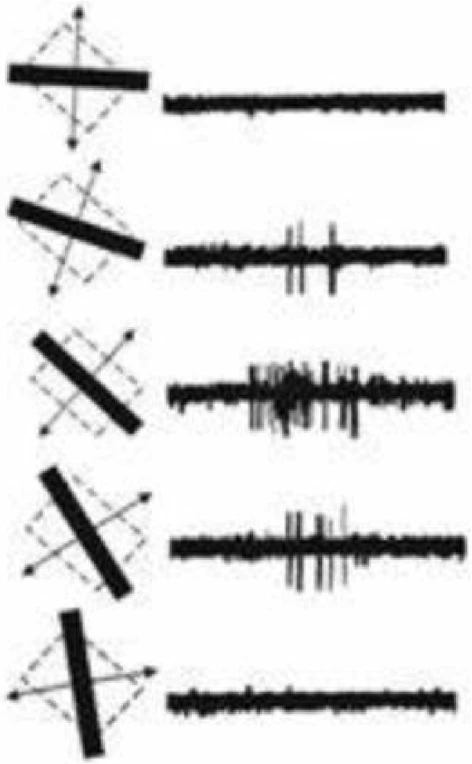
A Simple Cell in Cat Striate Cortex (V1)



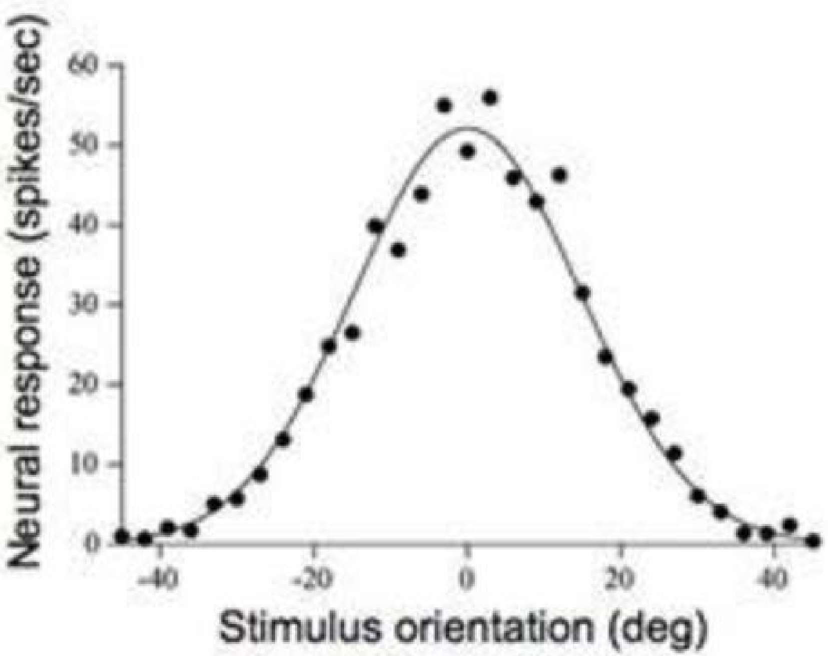
V1 neuron



David H. Hubel &
Torsten N. Wiesel



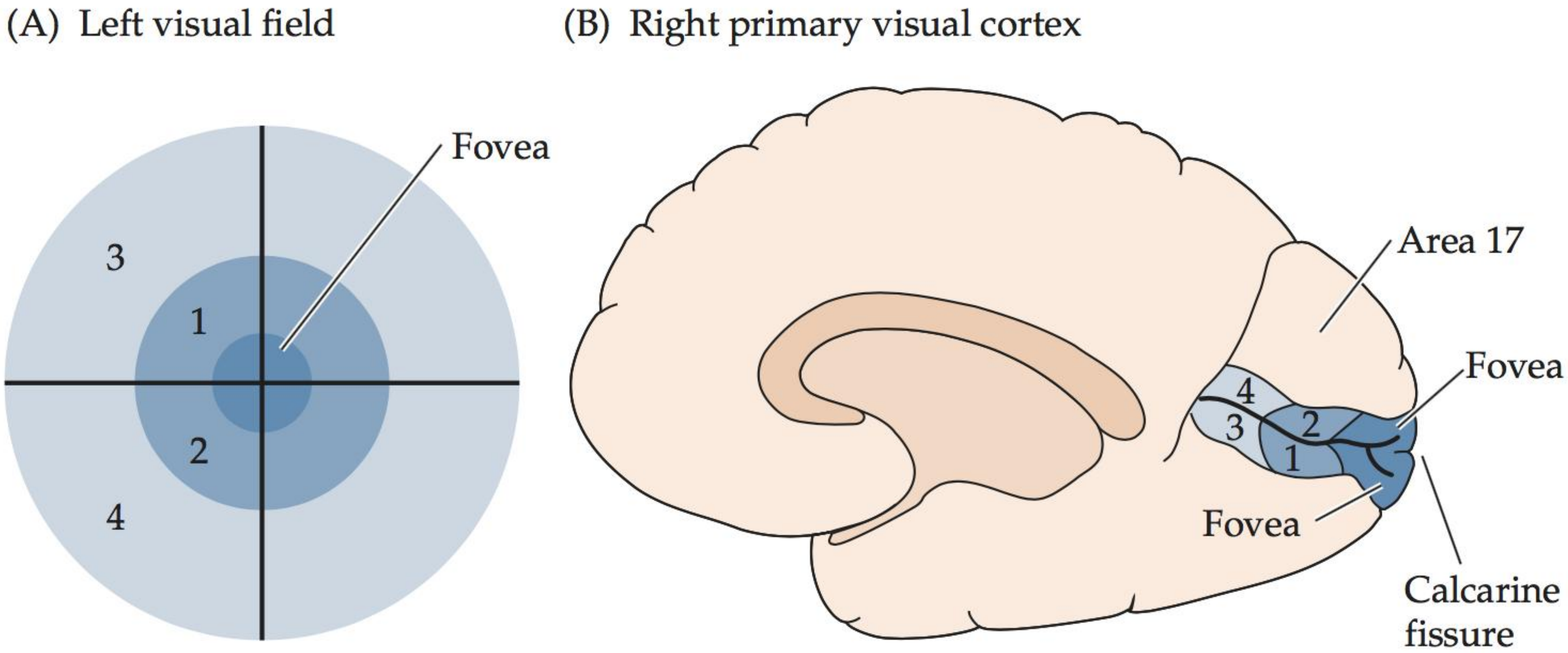
Tuning curve



Nobel Prize for Physiology or Medicine in 1981

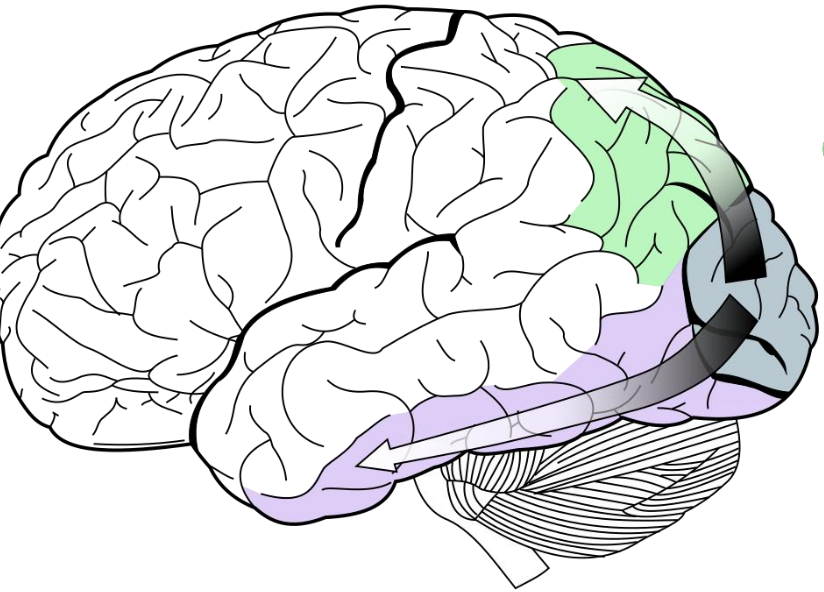
Hubel, D. H. 1982. *Nature* 299: 515–524.
Hubel, D. H., and Wiesel, T. N. 1959. *J. Physiol.* 148: 574–591.
Hubel, D. H., and Wiesel, T. N. 1962. *J. Physiol.* 160: 106–154.
Hubel, D. H., and Wiesel, T. N. 1968. *J. Physiol.* 195: 215–243.

Visual Field Map of the Cortex



Perceptual integration

-- the dorsal and ventral streams



dorsal 'where' pathway

- **the motion & spatial location**
- **V1, V2, V3, MT (V5), MST & inferior parietal cortex**

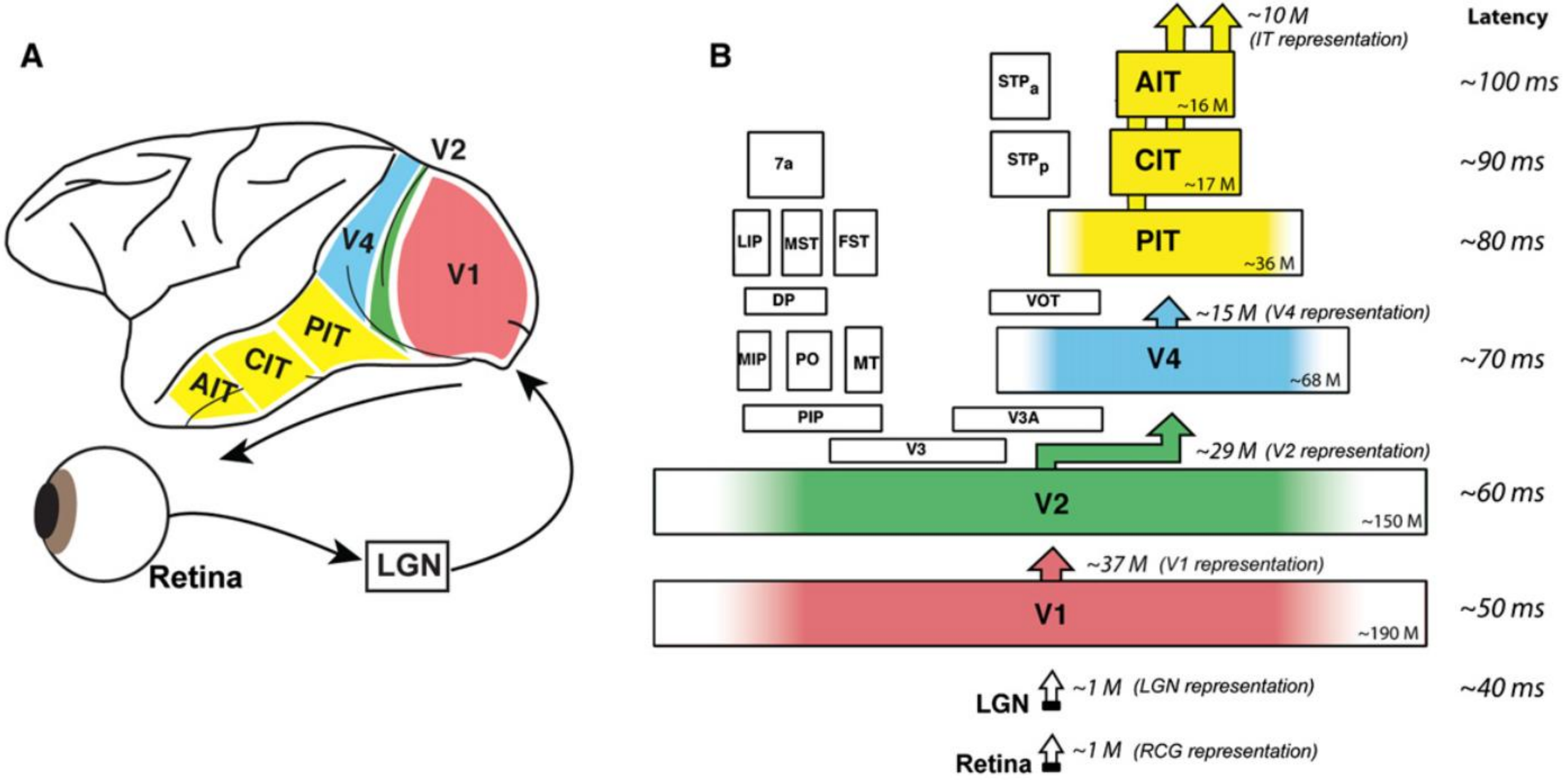
&

ventral 'what' pathway

- **the detailed features, form & object identity**
- **V1, V2, V4 & inferior temporal areas**

Goodale MA, Milner AD (1992). Separate visual pathways for perception and action. *Trends Neurosci*

The Ventral Visual Pathway: for object recognition

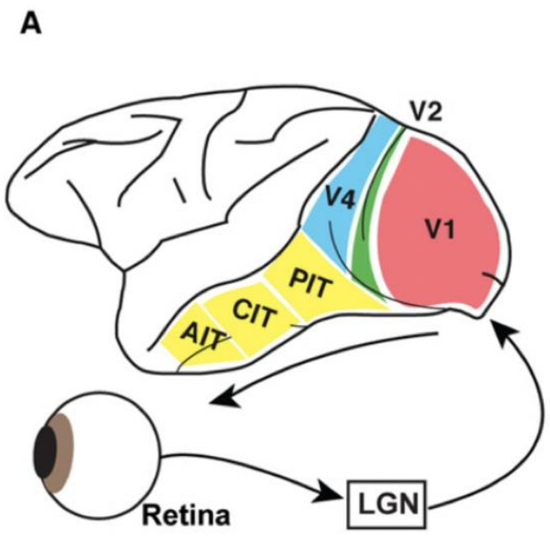


Anterior Inferior Temporal

Lateral Geniculate Nucleus

(A) Ventral stream cortical area locations in the macaque monkey brain, and flow of visual information from the retina.
(B) Each area is plotted so that its size is proportional to its cortical surface area (Felleman and Van Essen, 1991). Approximate total number of neurons (both hemispheres) is shown in the corner of each area (M = million). The approximate dimensionality of each representation (number of projection neurons) is shown above each area, based on neuronal densities (Collins et al., 2010), layer 2/3 neuronal fraction (O'Kusky and Colonnier, 1982), and portion (color) dedicated to processing the central 10 deg of the visual field (Brewer et al., 2002). Approximate median response latency is listed on the right (Nowak and Bullier, 1997; Schmolesky et al., 1998).

Along the Ventral Visual Pathway, feature extraction from simple to complex.



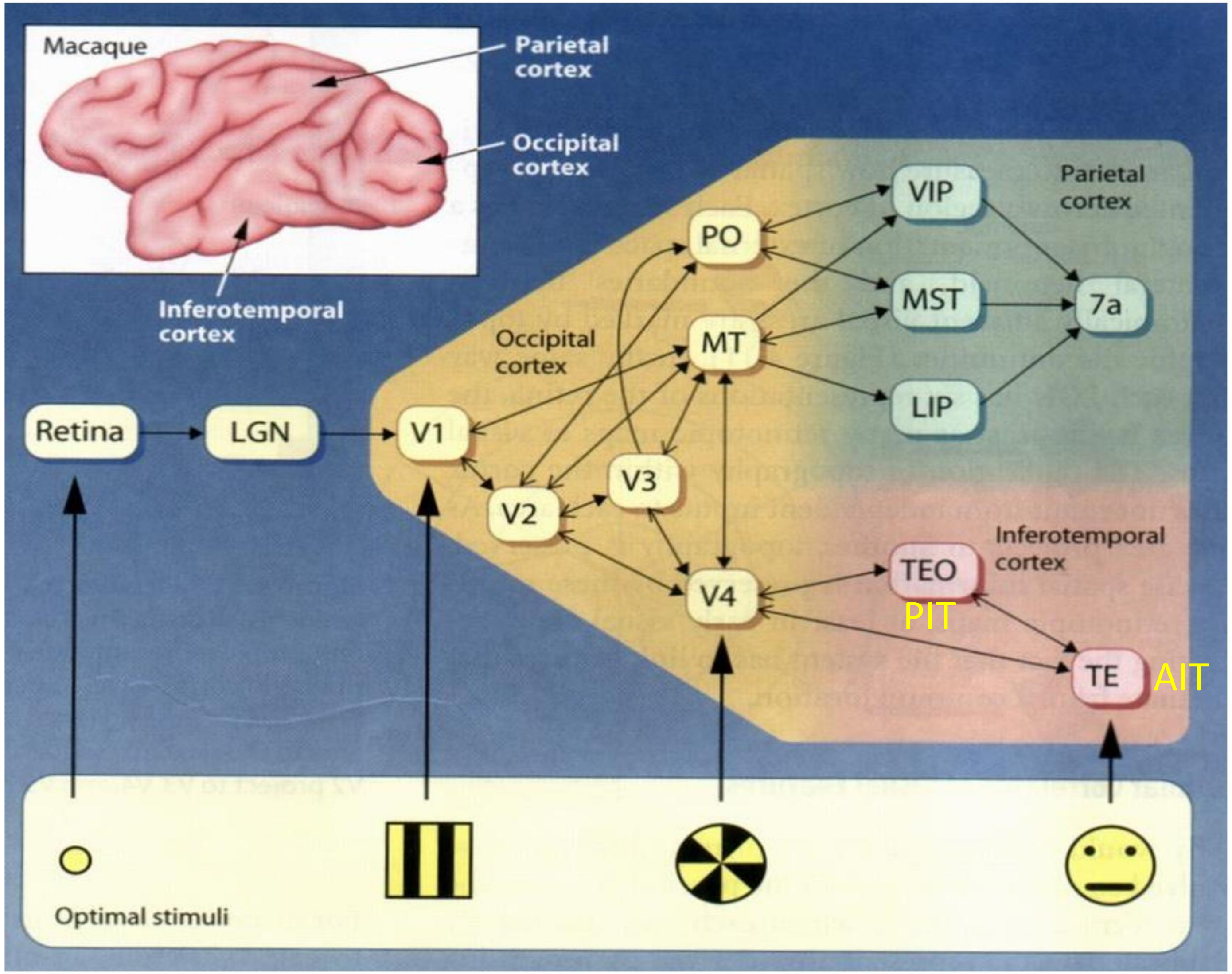
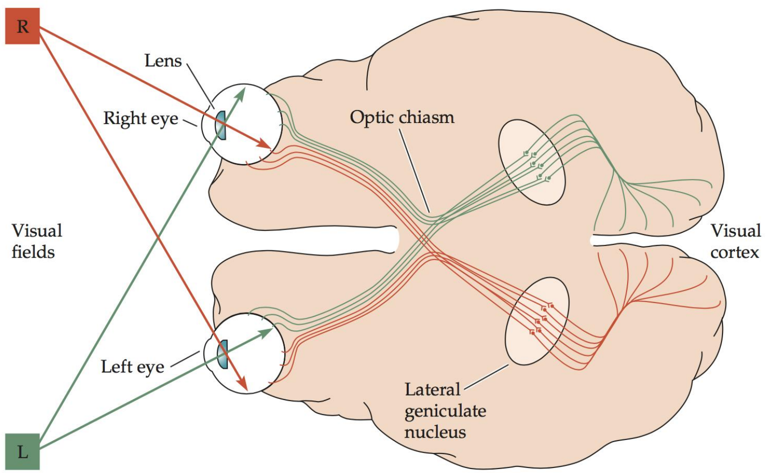
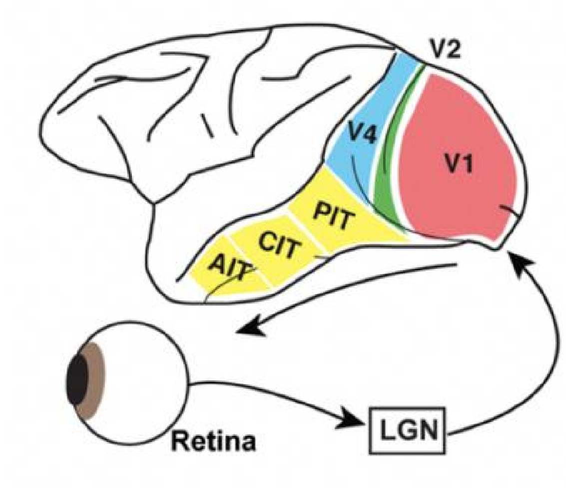


Image process: brain vs. CNN

Both the visual system and CNN are layered and hierarchical.



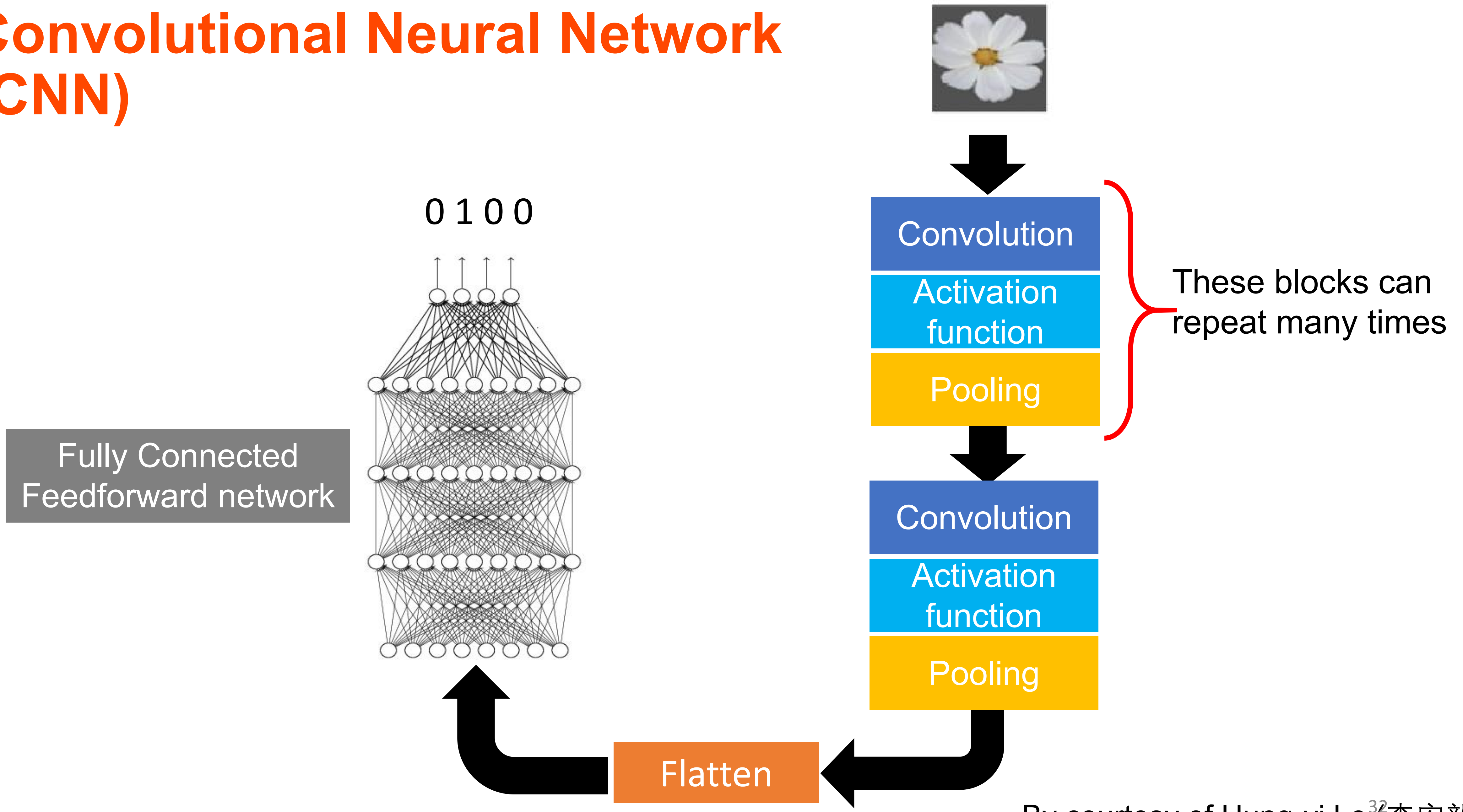


- Rods, cones
- Interneurons
- Ganglion cells
- Optic fiber
- LGN
- V1
- Ventral / dorsal pathways

Along the Visual Pathway, feature extraction from simple to complex.

➡ Automatically learn the hidden features in the image

Convolutional Neural Network (CNN)

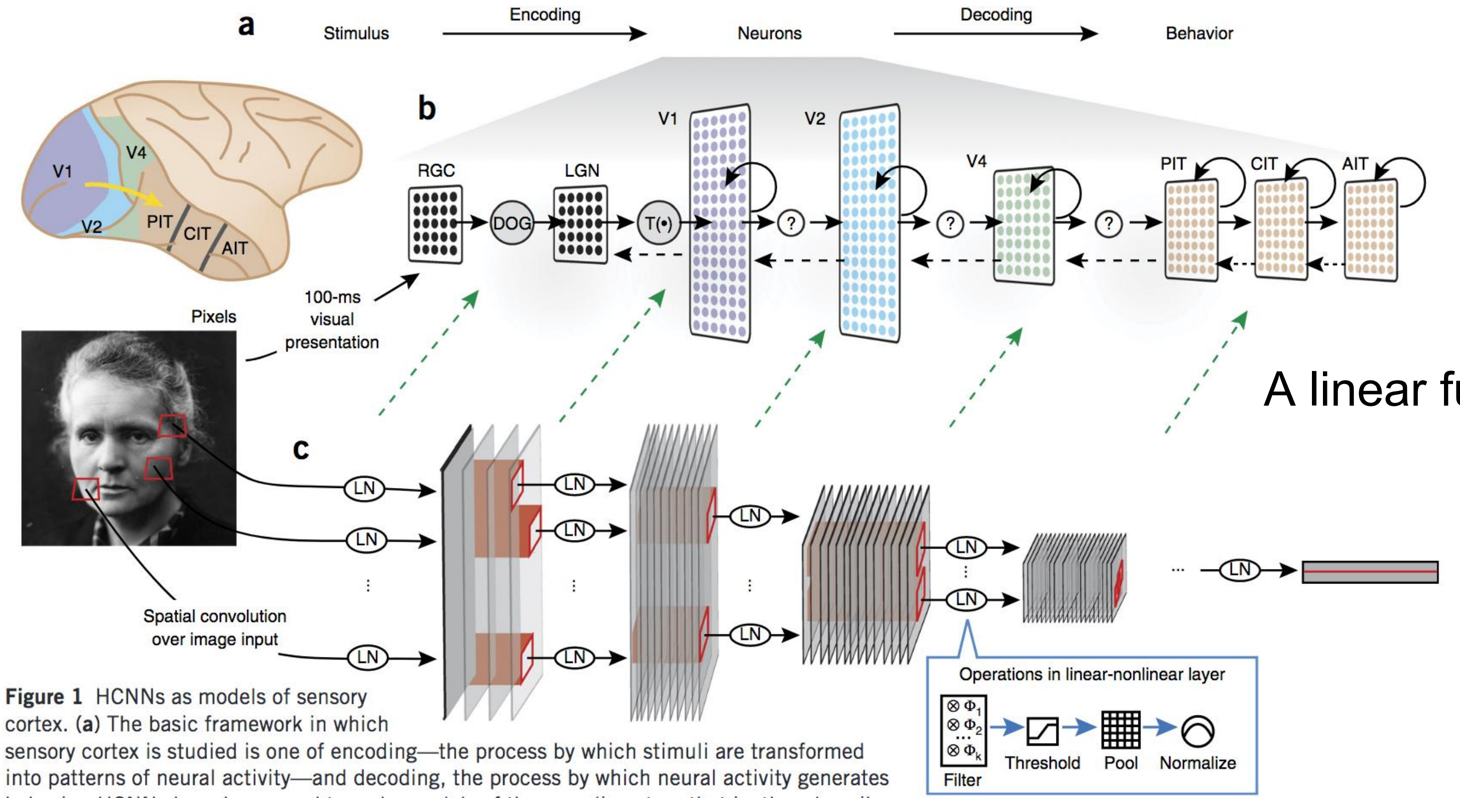


By courtesy of Hung-yi Le (李宏毅)

Q1:

Whether can we predict the neural signals based on the input image?

Image → neural activity



A linear function

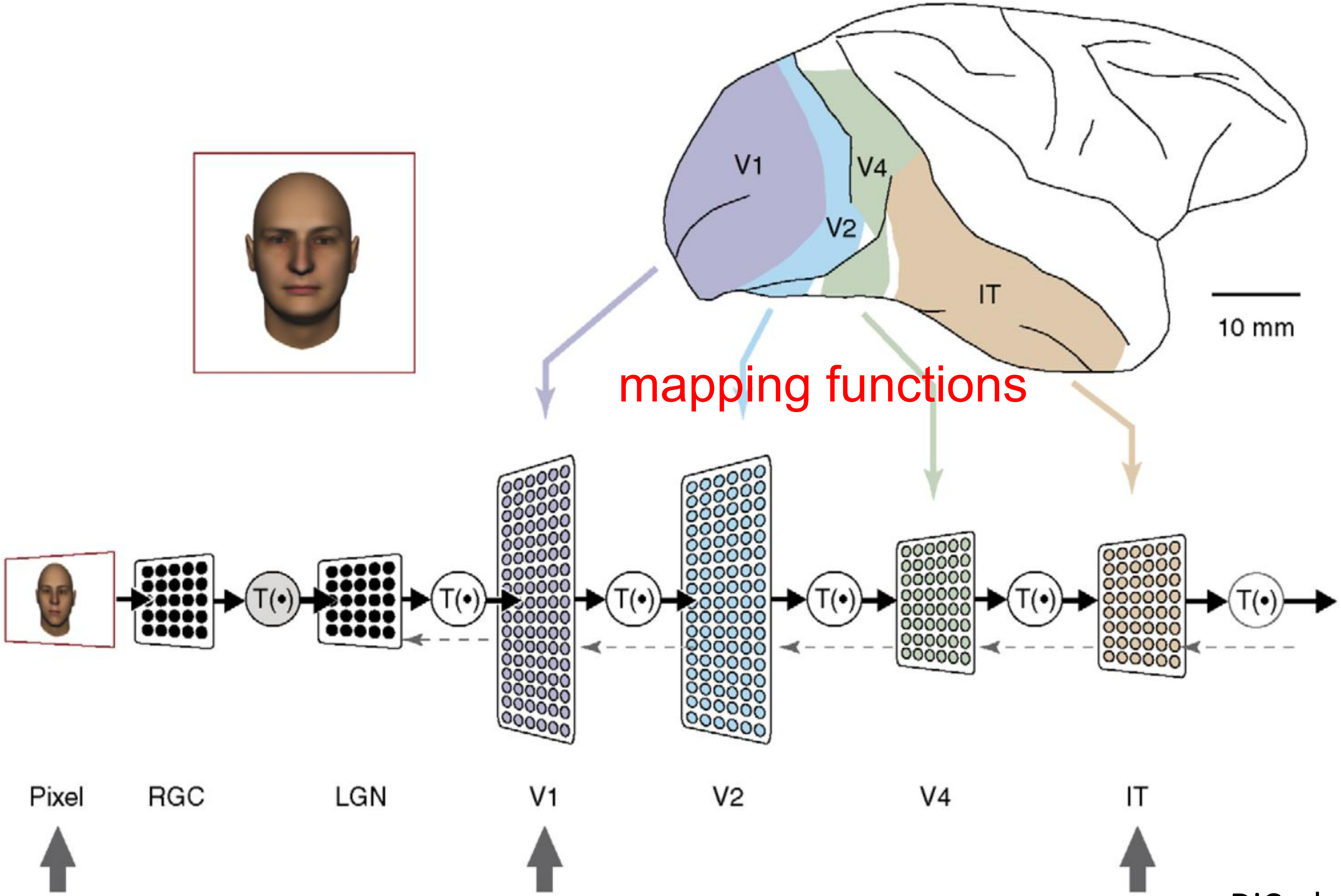
Figure 1 HCNNs as models of sensory cortex. (**a**) The basic framework in which sensory cortex is studied is one of encoding—the process by which stimuli are transformed into patterns of neural activity—and decoding, the process by which neural activity generates behavior. HCNNs have been used to make models of the encoding step; that is, they describe the mapping of stimuli to neural responses as measured in brain. (**b**) The ventral visual pathway is the most comprehensively studied sensory cascade. It consists of a series of connected cortical brain areas (macaque brain shown). PIT, posterior inferior temporal cortex; CIT, central; AIT, anterior; RGC, retinal ganglion cell; LGN, lateral geniculate nucleus. DoG, difference of Gaussians model; T(•), transformation. (**c**) HCNNs are multilayer neural networks, each of whose layers are made up of a linear-nonlinear (LN) combination of simple operations such as filtering, thresholding, pooling and normalization. The filter bank in each layer consists of a set of weights analogous to synaptic strengths. Each filter in the filter bank corresponds to a distinct template, analogous to Gabor wavelets with different frequencies and orientations; the image shows a model with four filters in layer 1, eight in layer 2, and so on. The operations within a layer are applied locally to spatial patches within the input, corresponding to simple, limited-size receptive fields (red boxes). The composition of multiple layers leads to a complex nonlinear transform of the original input stimulus. At each layer, retinopy decreases and effective receptive field size increases. HCNNs are good candidates for models of the ventral visual pathway. By definition, they are image computable, meaning that they generate responses for arbitrary input images; they are also mappable, meaning that they can be naturally identified in a component-wise fashion with observable structures in the ventral pathway; and, when their parameters are chosen correctly, they are predictive, meaning that layers within the network describe the neural response patterns to large classes of stimuli outside the domain on which the models were built.

Yamins (2016) Nature Neuroscience

Is there any similarity between the neural representation in brain and AI?

The answer is yes.
Just require a mapping function

Biological neural representation resembles to the artificial neural representation.



Technical limits

Low-dimensional neuron recordings

easy

hard

High-dimensional CNN feature maps

DiCarlo and Cox, TiCS, 2007

Brain score: how well existing models explain the neural data

http://www.brain-score.org/#leaderboard

Sort by average score

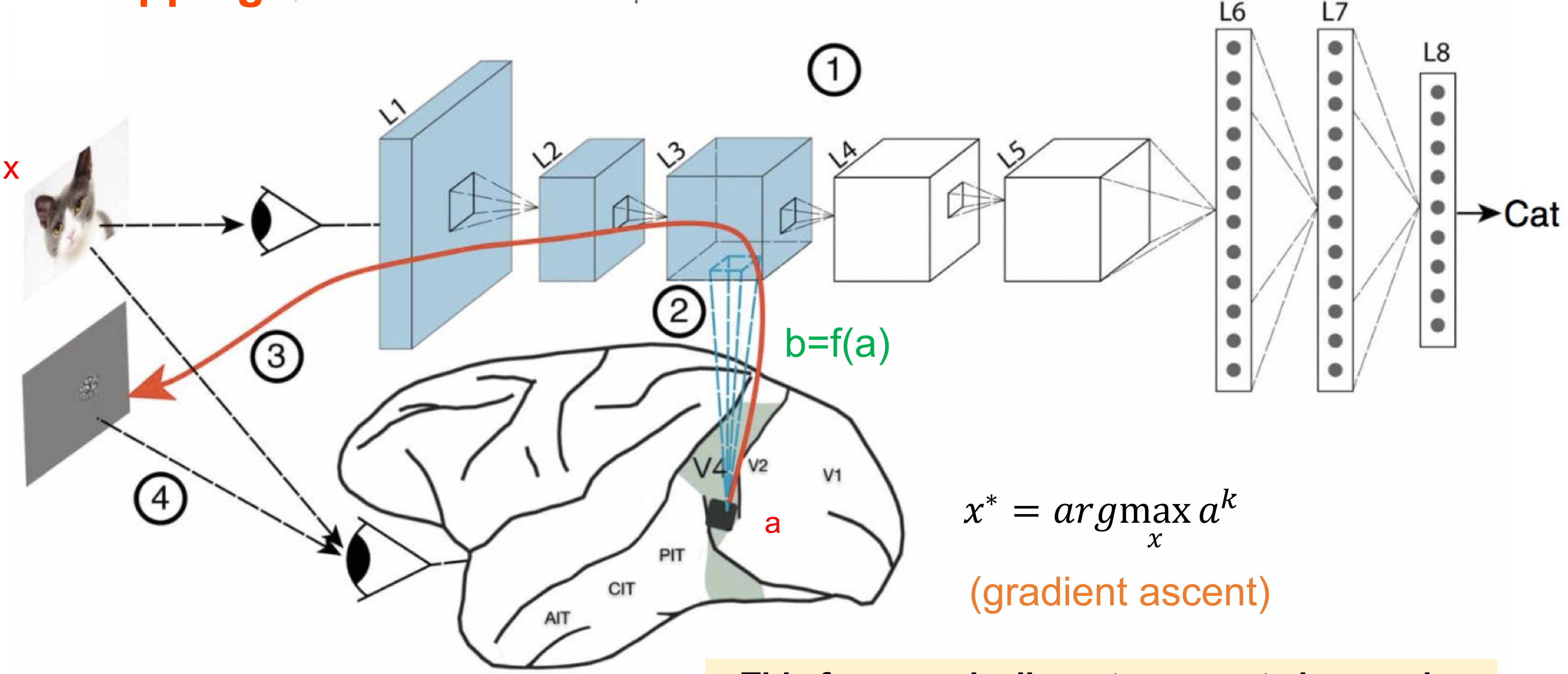
Rank	Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
1	efficientnet-b0 *Tan et al., 2019*	.442	.215	.317	.556	.547	.573		
2	efficientnet-b6 *Tan et al., 2019*	.435	.263	.295	.563	.541	.513		
3	efficientnet-b2 *Tan et al., 2019*	.434	.213	.317	.569	.547	.526		
4	efficientnet-b4 *Tan et al., 2019*	.434	.228	.286	.575	.543	.535		
5	CORnet-S *Kubilius et al., 2018*	.417	.294	.242	.581	.423	.545	.747	.747
6	vgg-19 *Simonyan et al., 2014*	.408	.347	.341	.610	.248	.494	.711	.711
7	resnet-50-robust *Santurkar et al., 2019*	.408	.378	.365	.537	.243	.515		
8	resnet-101_v1 *He et al., 2015*	.407	.266	.341	.590	.274	.561	.764	.764
9	vgg-16 *Simonyan et al., 2014*	.406	.355	.336	.620	.259	.461	.715	.715
10	resnet-152_v1 *He et al., 2015*	.405	.282	.338	.598	.277	.533	.768	.768

Sort by V4 score

Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
vgg-16 *Simonyan et al., 2014*	.406	.355	.336	.620	.259	.461	.715	.715
vgg-19 *Simonyan et al., 2014*	.408	.347	.341	.610	.248	.494	.711	.711
xception *Chollet et al., 2016*	.384	.245	.306	.610	.249	.508	.790	.790
densenet-169 *Huang et al., 2016*	.404	.281	.322	.601	.274	.543	.759	.759
resnet-50-pytorch *He et al., 2015*	.399	.289	.317	.600	.259	.528	.752	.752
resnet-101_v2 *He et al., 2015*	.404	.274	.332	.599	.263	.555	.774	.774
resnet50-SIN_IN *Geirhos et al., 2019*	.404	.282	.324	.599	.276	.541	.746	.746
densenet-201 *Huang et al., 2016*	.402	.277	.325	.599	.273	.537	.772	.772
resnet-152_v1 *He et al., 2015*	.405	.282	.338	.598	.277	.533	.768	.768
resnet50-SIN_IN_IN *Geirhos et al., 2019*	.397	.275	.321	.596	.273	.523	.767	.767

A mapping function from L3 to V4

x



b=f(a)

a

$$x^* = \underset{x}{argmax}\, a^k$$

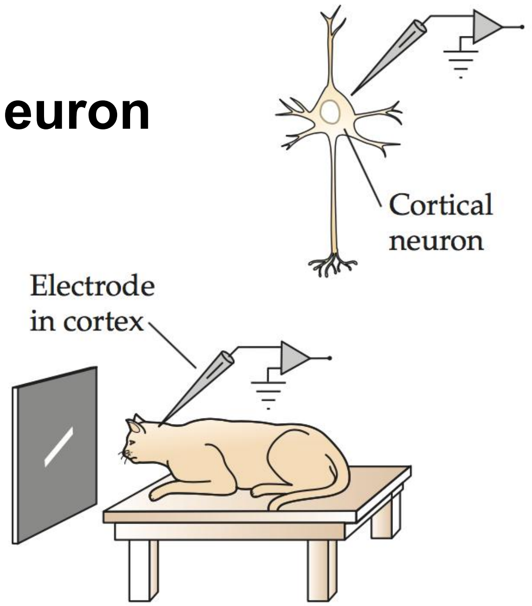
(gradient ascent)

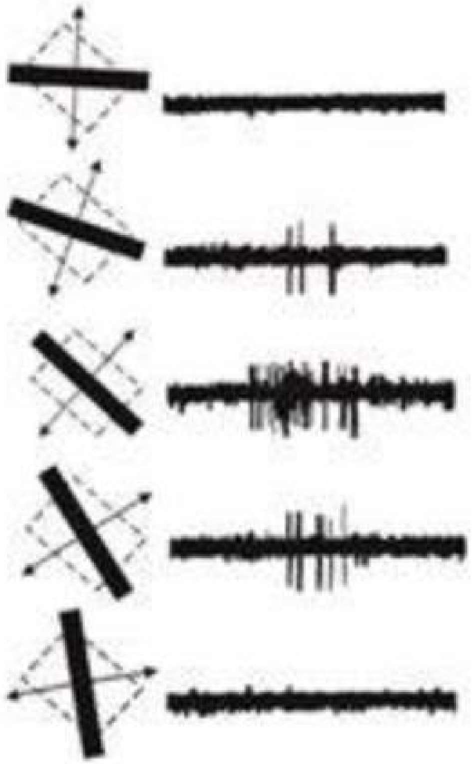
This framework allows to generate images by gradient ascent in ANN to maximize biological neural population activity.

Yamins, nature neurosci, 2016
Bashivan, Science, 2019

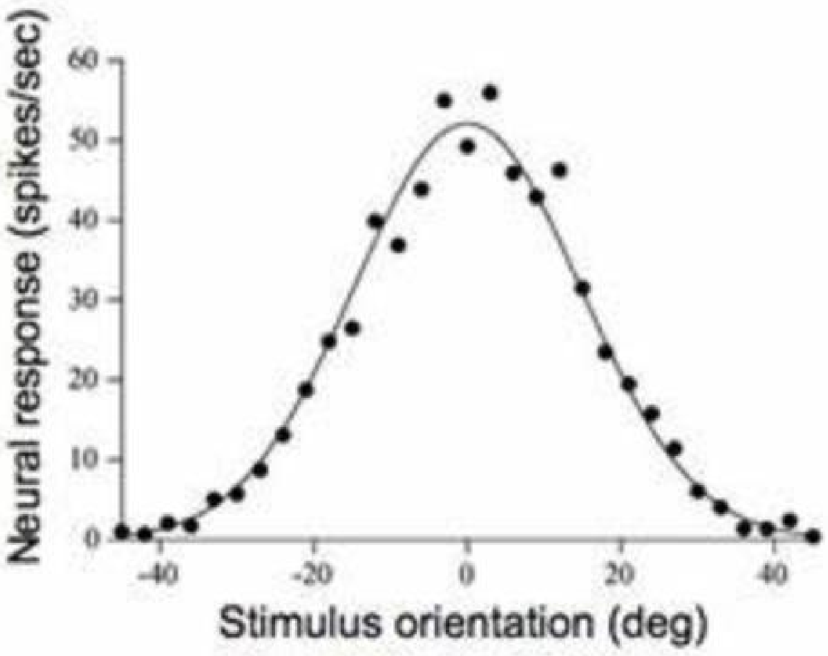
Why to find the image to maximize neuronal activity?

V1 neuron





Tuning curve



Hubel, D. H. 1982. *Nature* 299: 515–524.

Hubel, D. H., and Wiesel, T. N. 1959. *J. Physiol.* 148: 574–591.

Hubel, D. H., and Wiesel, T. N. 1962. *J. Physiol.* 160: 106–154.

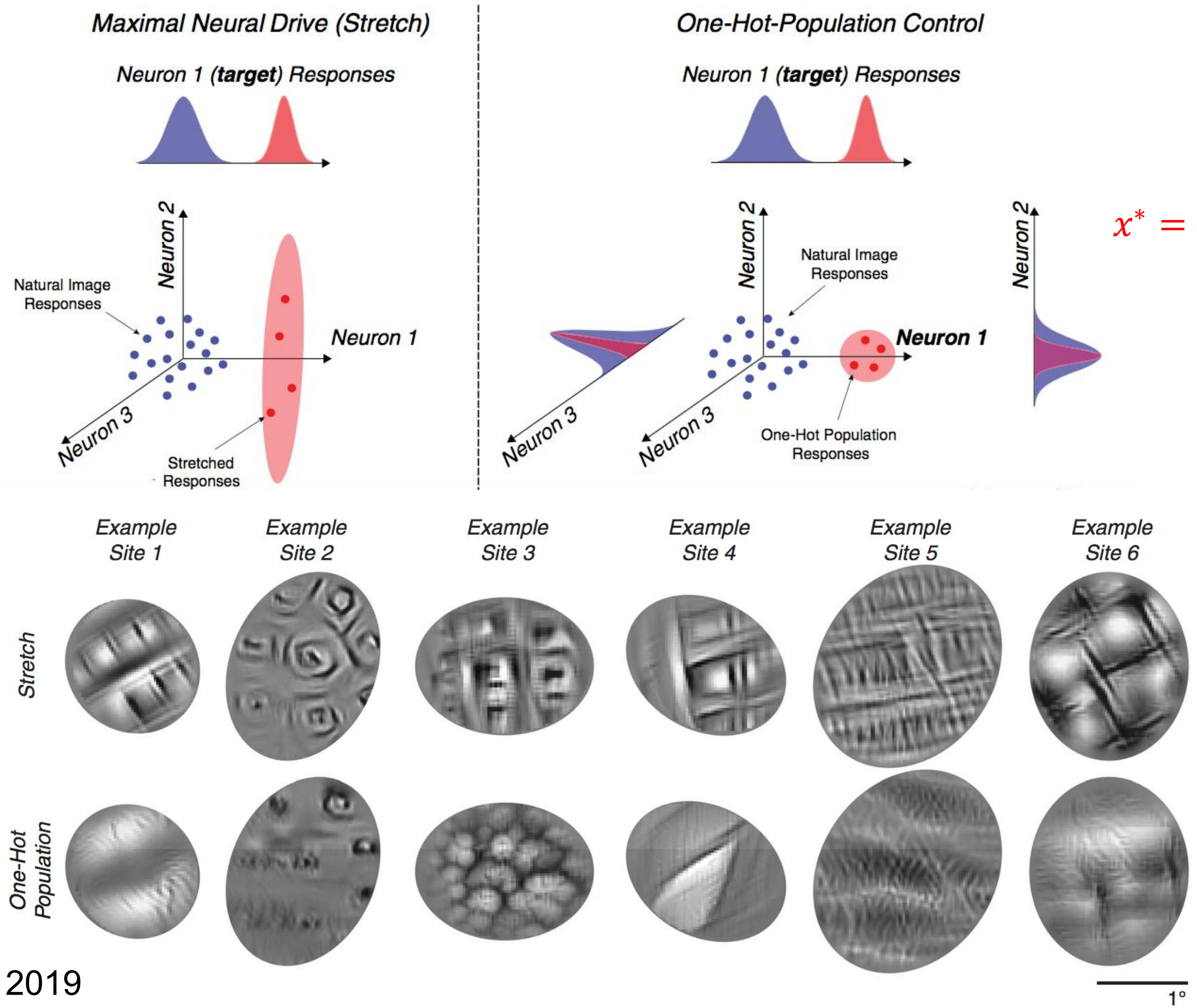
Hubel, D. H., and Wiesel, T. N. 1968. *J. Physiol.* 195: 215–243.

David H. Hubel & Torsten N. Wiesel

Nobel Prize for Physiology or Medicine in 1981

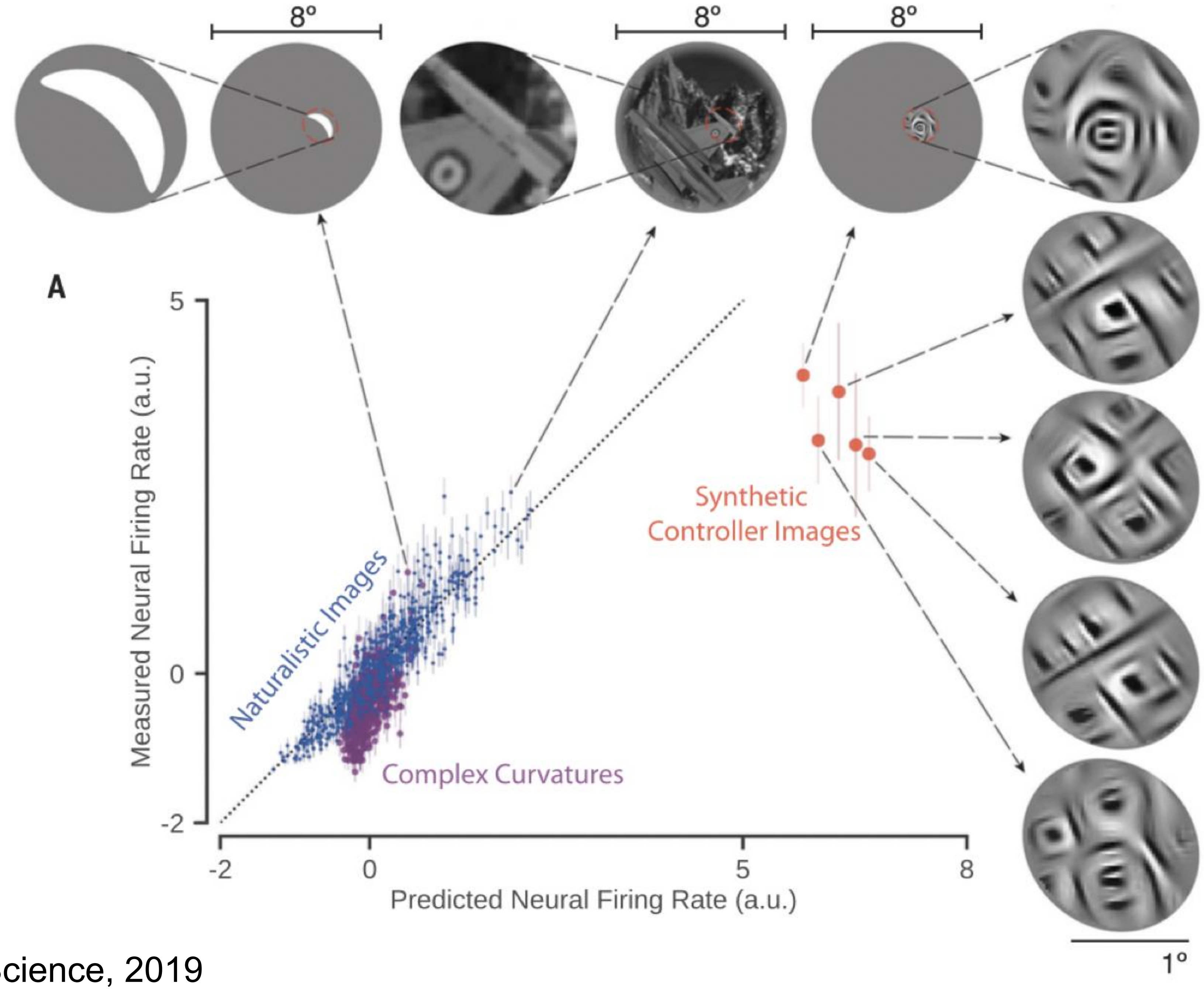
How about V4 and IT?

$$x^* = \arg\max_x a^k$$



$$x^* = \arg\max_x \frac{e^{a^k}}{\sum_k e^{a^k}}$$

Bashivan, Science, 2019



$$x^* = \underset{x}{\mathrm{arg\,max}}\, a^k$$

(gradient ascent)

Bashivan, Science, 2019

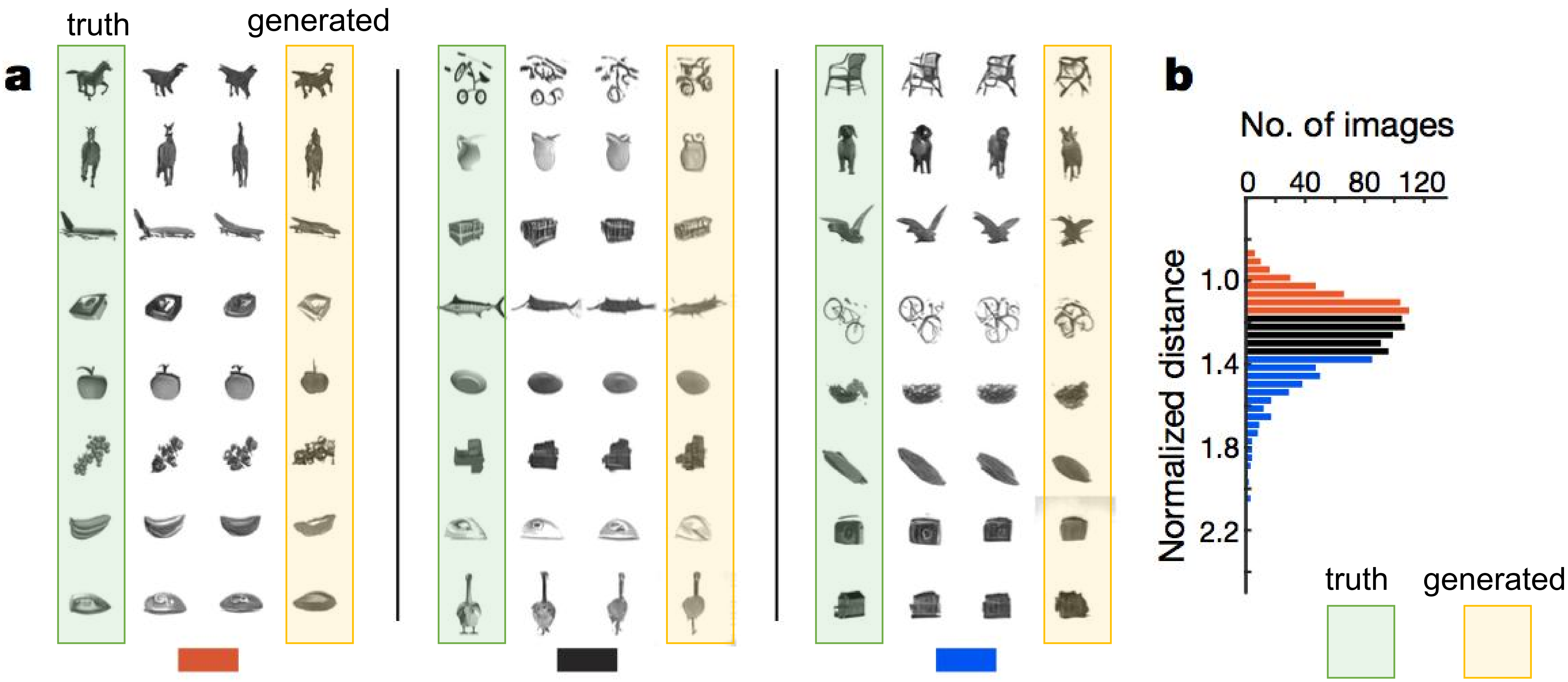
Q2:

Whether can we reconstruct the image based on neural signals?

neural activity → image

We will show you some detailed studies/algorithms in the following weeks.

Reconstruct images based on IT neuronal activity via GAN



Pinglei Bao, Liang She, Mason McGill & Doris Y. Tsao (2020). A map of object space in primate inferotemporal cortex. *Nature*

The Marr's three level of explanation

Level 1: Computation theory

- What is the problem to be solved?
- What are the inputs and outputs to the computation?
- What is the goal, and what is the logic by which it is carried out?

Level 2: Algorithmic

- How is the information represented and processed to achieve the computational goal?

Level 3: Implementation

- How is the computation realized in physical or biological hardware?

Reverse-engineer the brain

to understand the design principles of brain

Design eye-like camera

Design AI for computer vision

Explain visual phenomenon

Explain/Treat visual disease

Human vision enhancement

Decode/encode neural signals

Summary of Lecture 17 – Visual System

0. Marr's 3 levels of explanation
1. Evolution of the eye
2. Function of the visual system
3. Structure of the eye
4. Photoreceptors
5. Information integration by ganglion cell
6. **Visual pathways**: photoreceptors, interneurons, ganglion cells, LGN, V1, ventral/dorsal pathways
7. Image process: Brain vs CNN

Recommended materials

Some materials are from the textbook

- From Neuron to Brain (Ch1-Ch3)

Recommended book

- Principles of Neural Design, by Peter Sterling and Simon Laughlin

Course Project

- Find your teammates, **form a team**, discuss about your course project, and read 4 papers relevant to your course project (until May 10)
- **Do** the course project (until May 24)
- **Finalize** the course project (until May 26)
- The final presentation will be a **poster** presentation next to 大榕树 (on June 1)