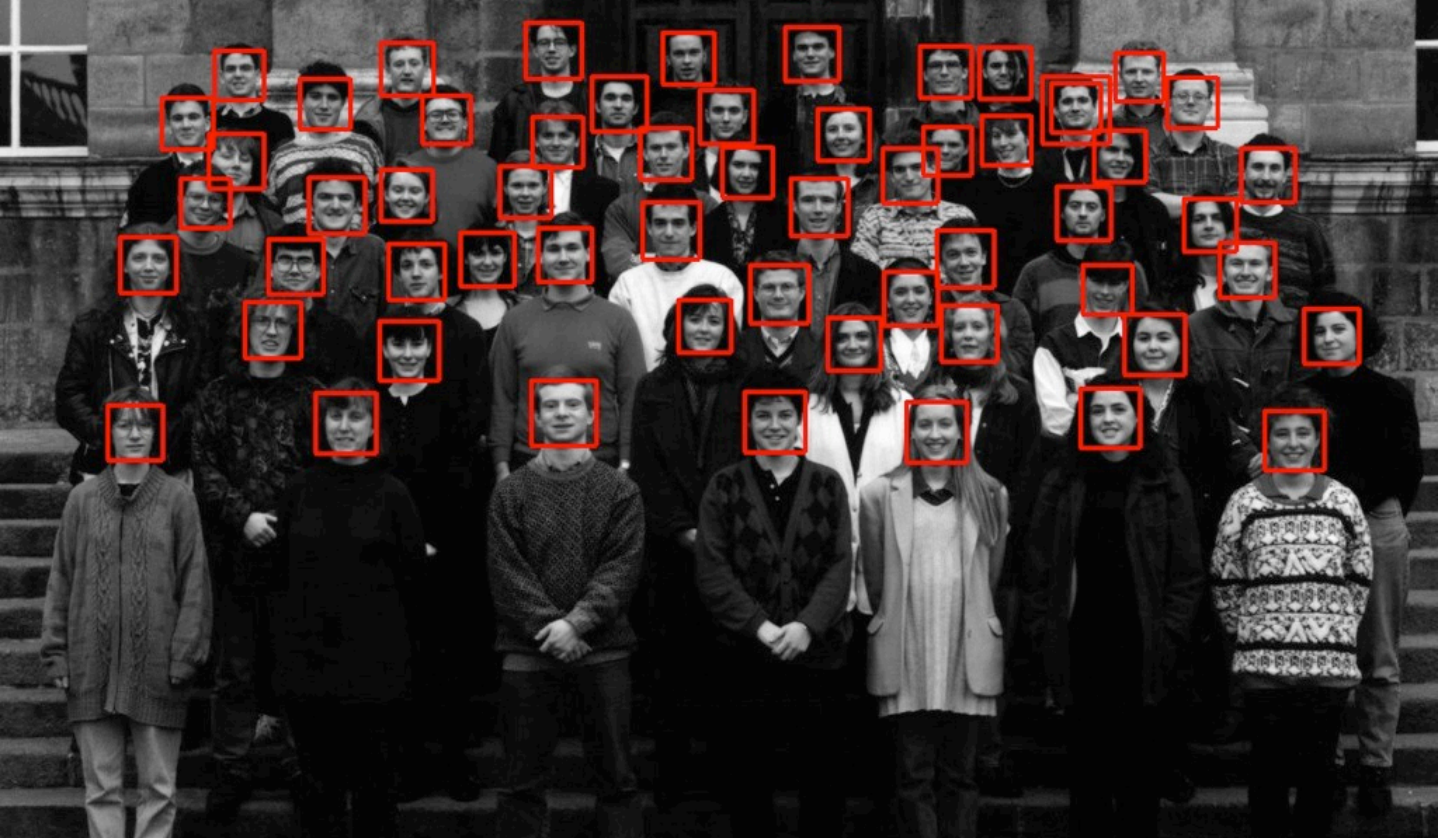
Deep learning in the visual cortex

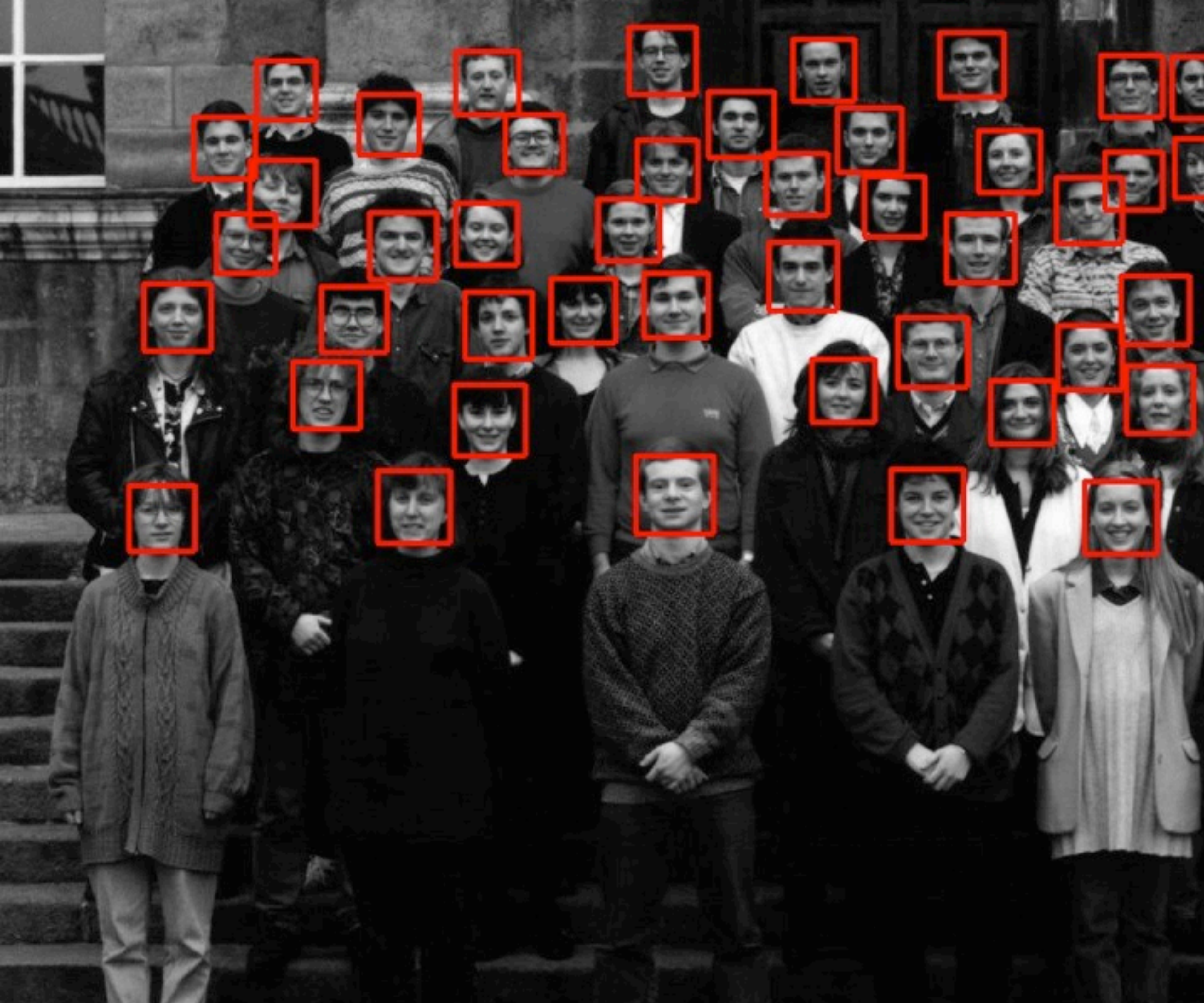
Thomas Serre
Brown University

Invariant recognition in natural images



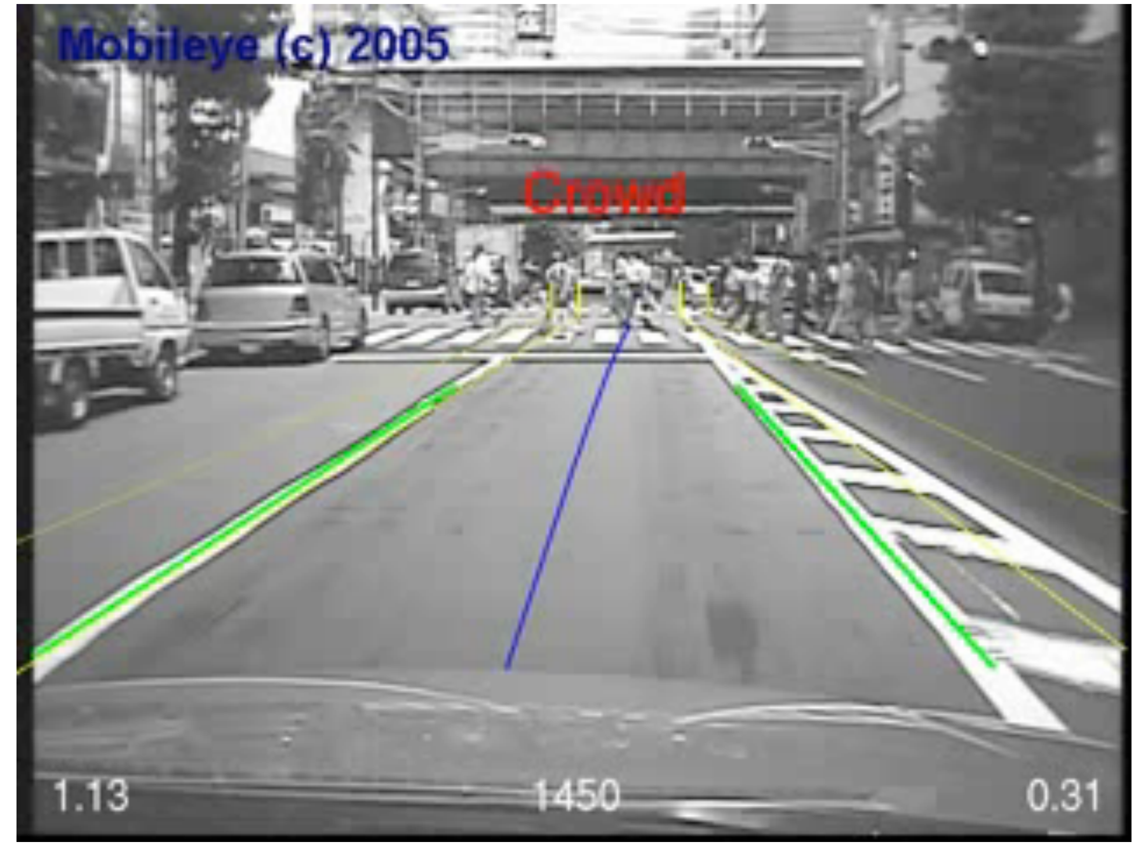
Computer vision successes

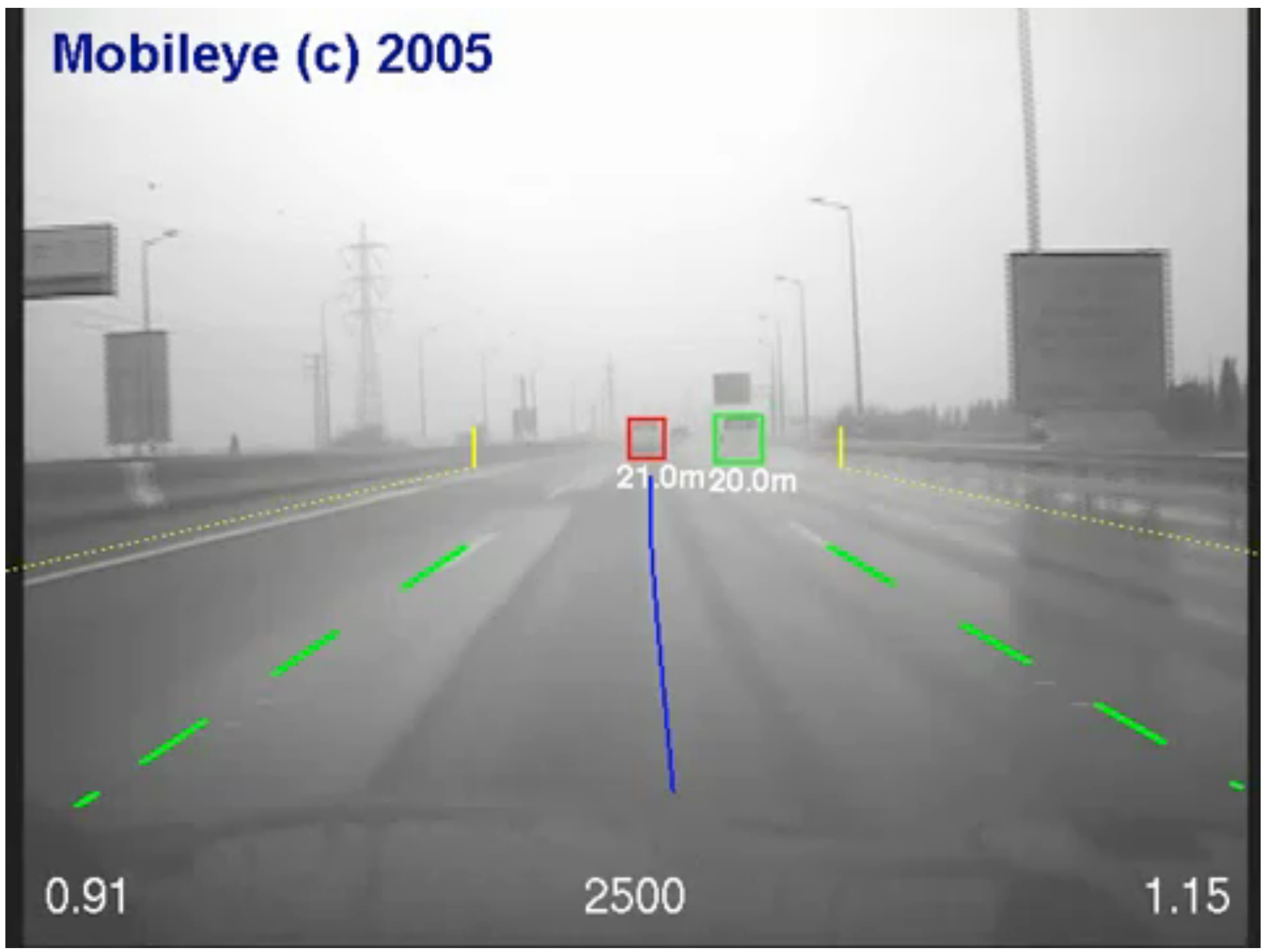
Face detection



Computer vision successes

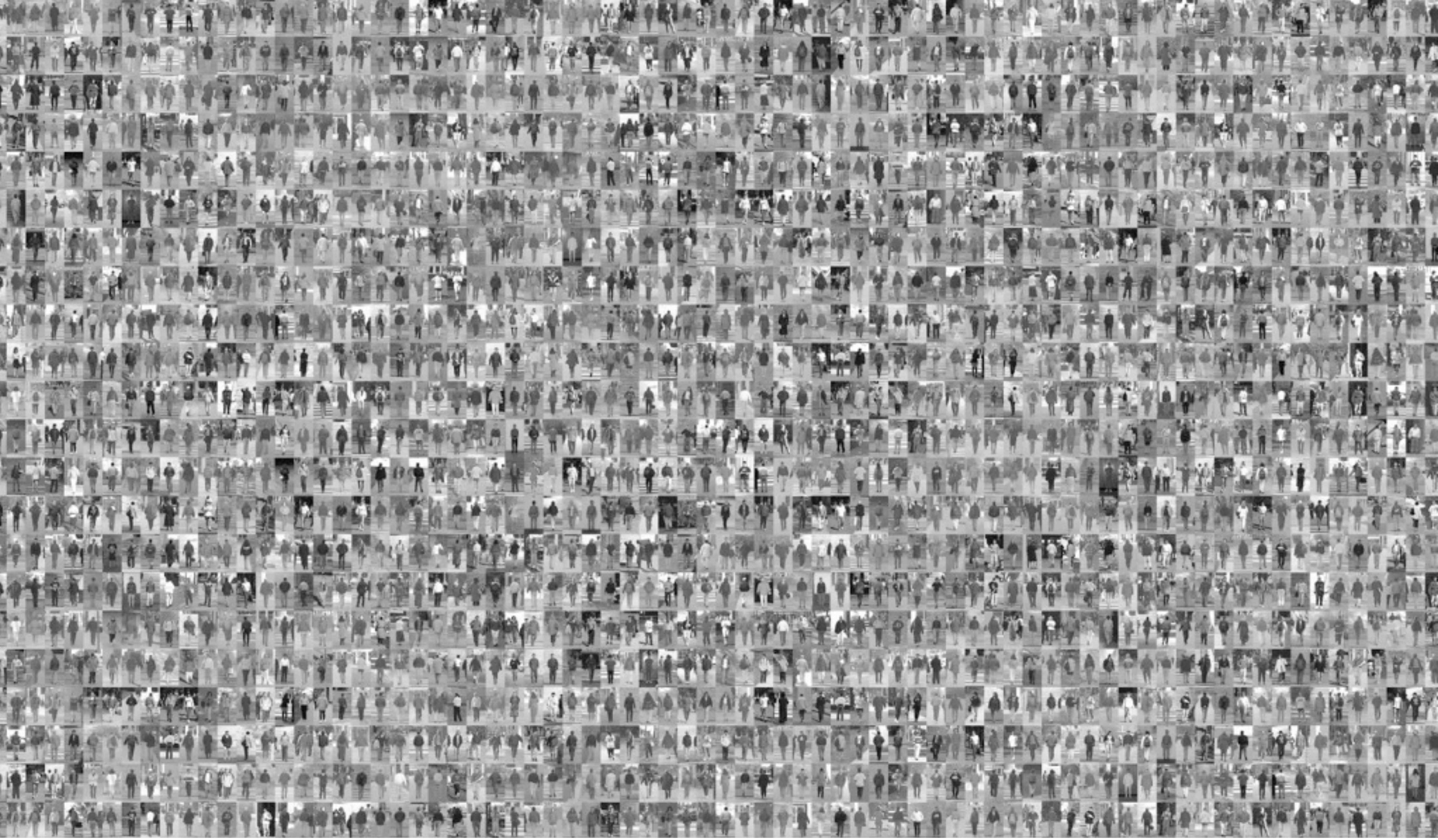
Face detection





Mobileye system

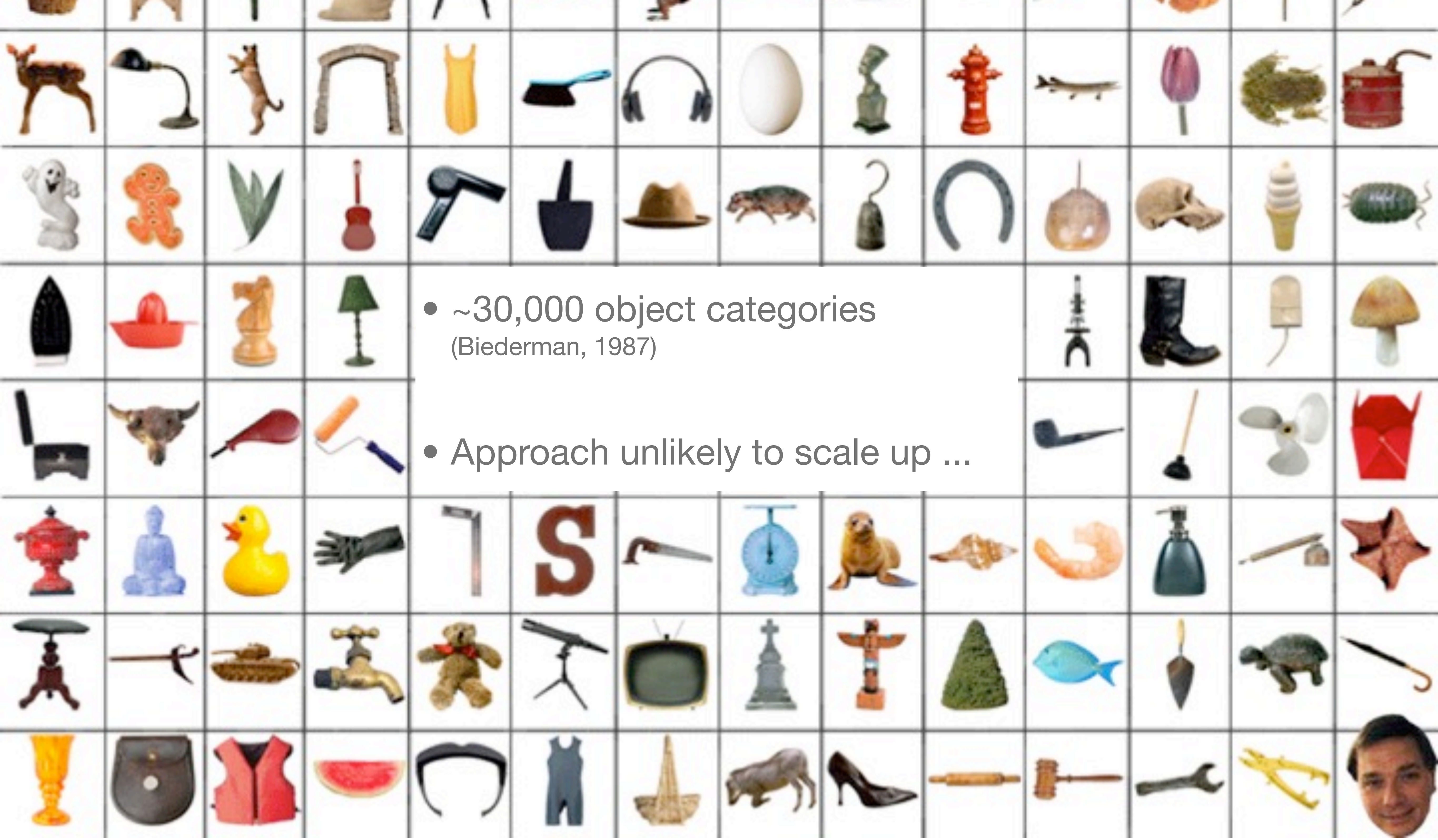
Already available on volvo S60 and soon on most car manufacturers



Machine: Millions of labeled examples for real-world applications

e.g., Mobileye pedestrian detection system

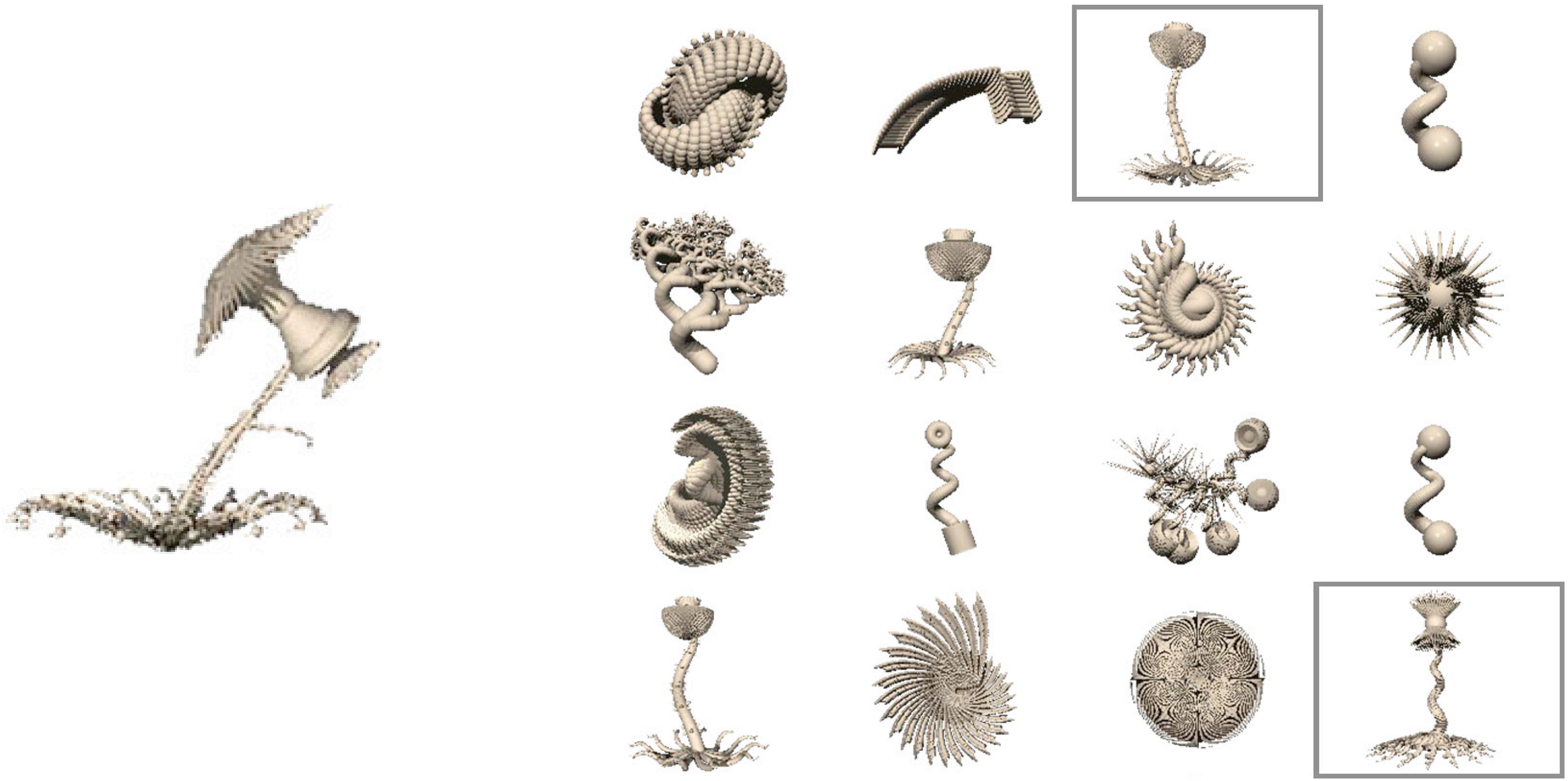
What's wrong with this picture?



- ~30,000 object categories
  (Biederman, 1987)

- Approach unlikely to scale up ...

What’s wrong with this picture?



Source: Tenenbaum

Humans: learning from very few examples

By age 6, a child knows 10-30K categories

Invariant recognition in natural images

- Subjects get the gist of the scene at 7 images/s from unpredictable random sequence of images
  - No time for eye movements
  - No top-down / expectations

movie courtesy of Jim DiCarlo

Potter 1971, 1975; see also Biederman 1972; Thorpe 1996

Invariant recognition in natural images

- Subjects get the gist of the scene at 7 images/s from unpredictable random sequence of images
  - No time for eye movements
  - No top-down / expectations

movie courtesy of Jim DiCarlo

Potter 1971, 1975; see also Biederman 1972; Thorpe 1996

Invariant recognition in natural images

- Subjects get the gist of the scene at 7 images/s from unpredictable random sequence of images
  - No time for eye movements
  - No top-down / expectations

Potter 1971, 1975; see also Biederman 1972; Thorpe 1996

Invariant recognition in natural images

- Subjects get the gist of the scene at 7 images/s from unpredictable random sequence of images
  - No time for eye movements
  - No top-down / expectations

Potter 1971, 1975; see also Biederman 1972; Thorpe 1996

Invariant recognition in natural images

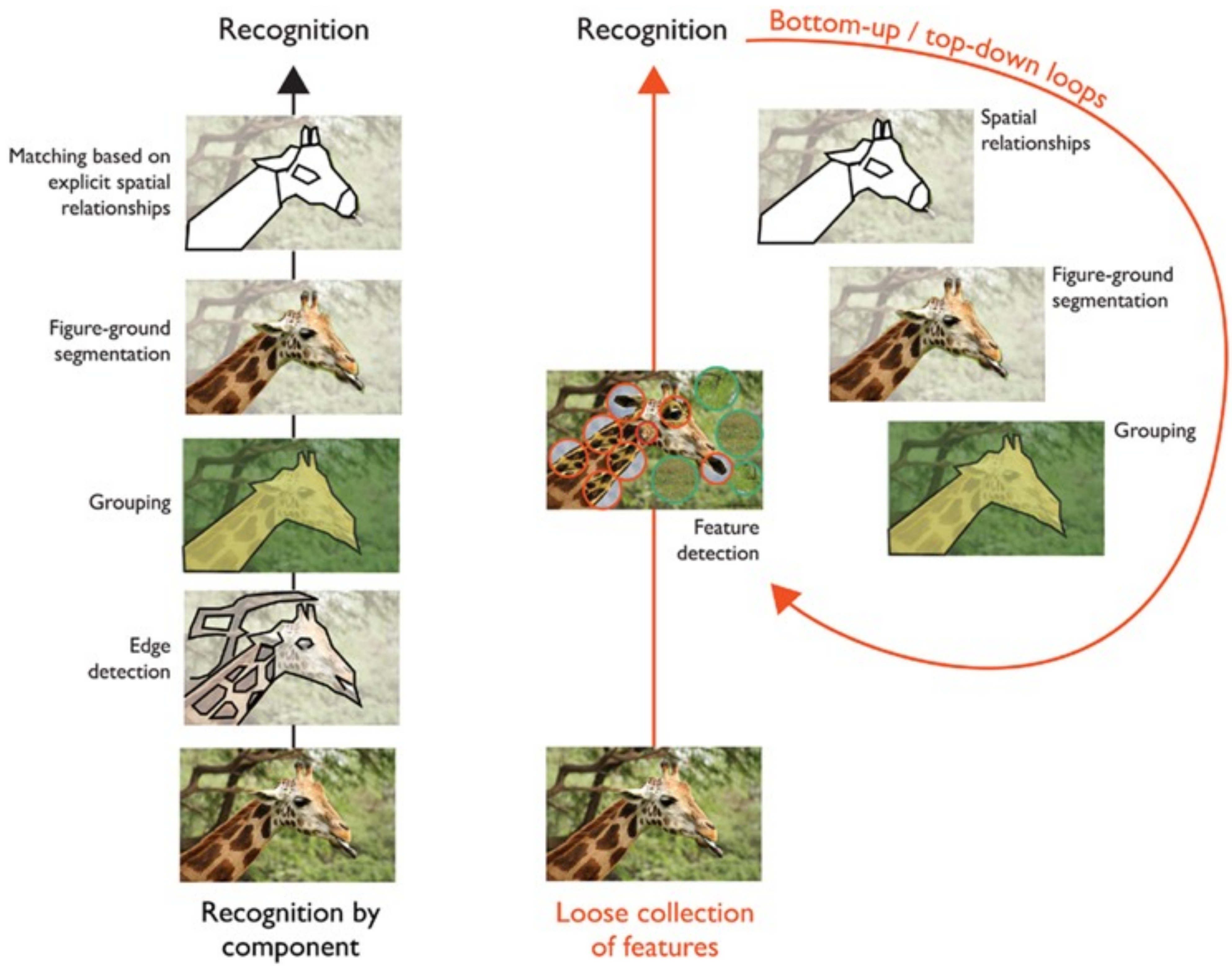
- Subjects get the gist of the scene at 7 images/s from unpredictable random sequence of images
  - No time for eye movements
  - No top-down / expectations

- Coarse initial base representation

- Enables rapid object detection/recognition ('what is there?')
- Insufficient for object localization
- Sensitive to presence of clutter

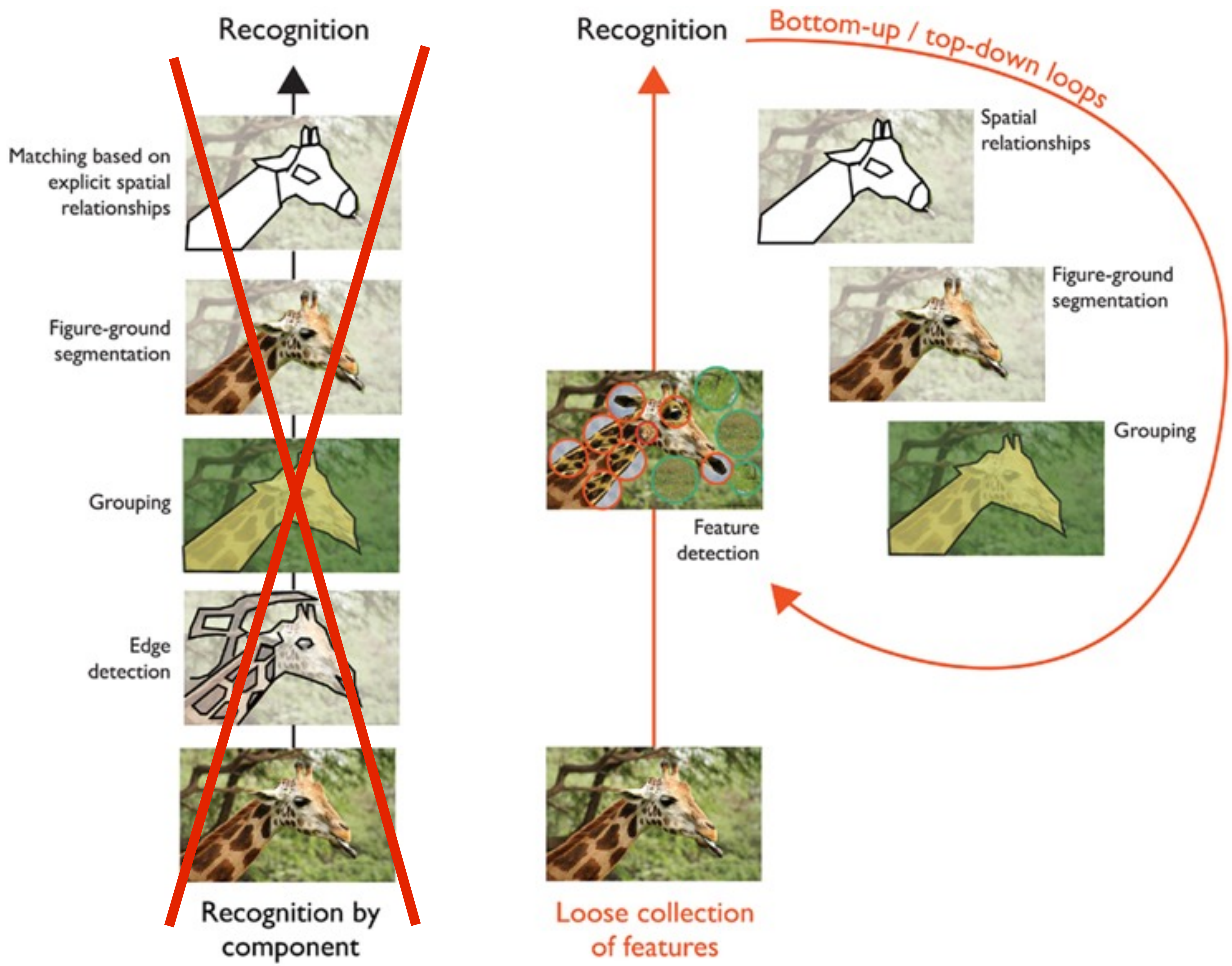
Potter 1971, 1975; see also Biederman 1972; Thorpe 1996

Two classes of models



Marr '82;
Biederman '87

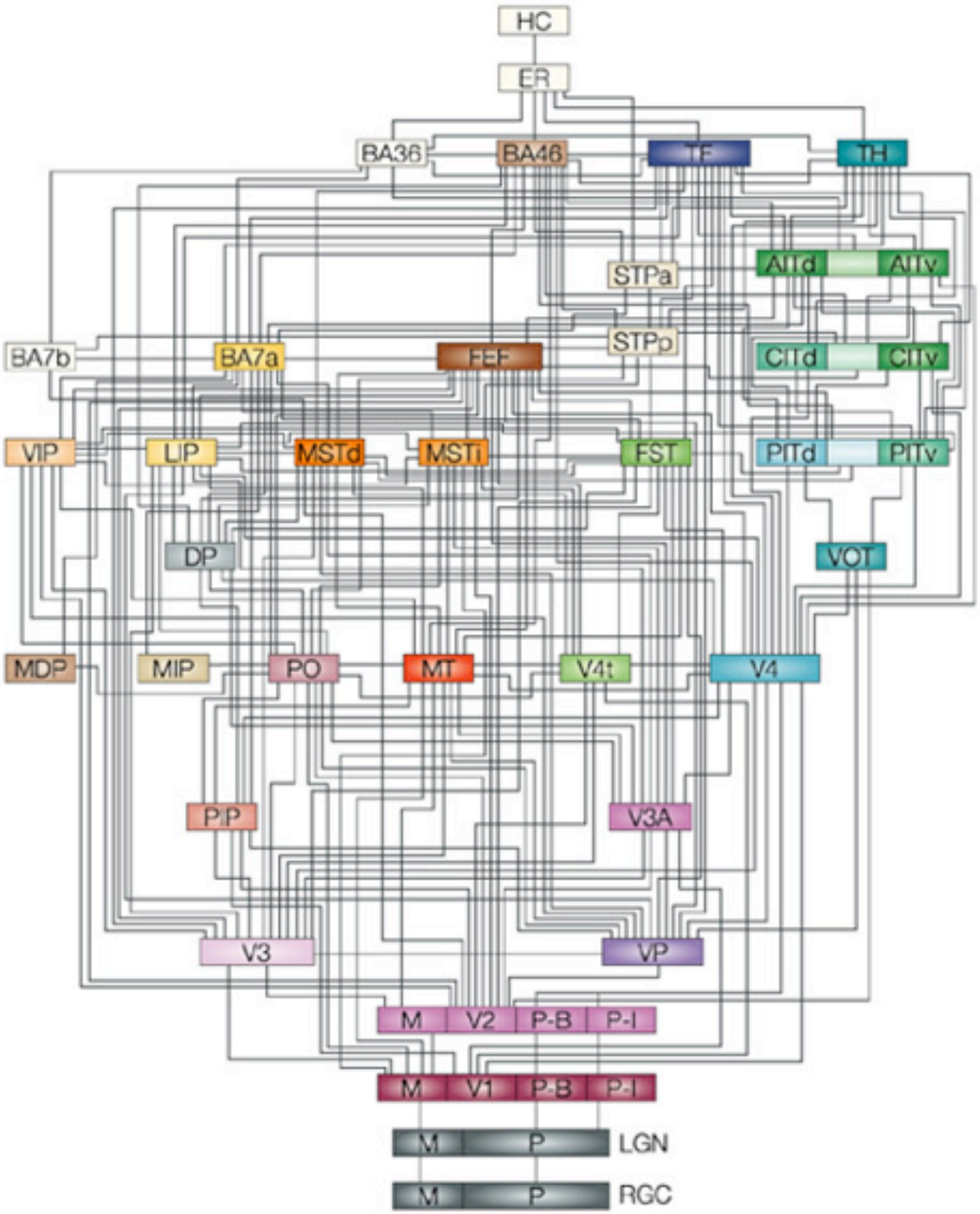
Two classes of models

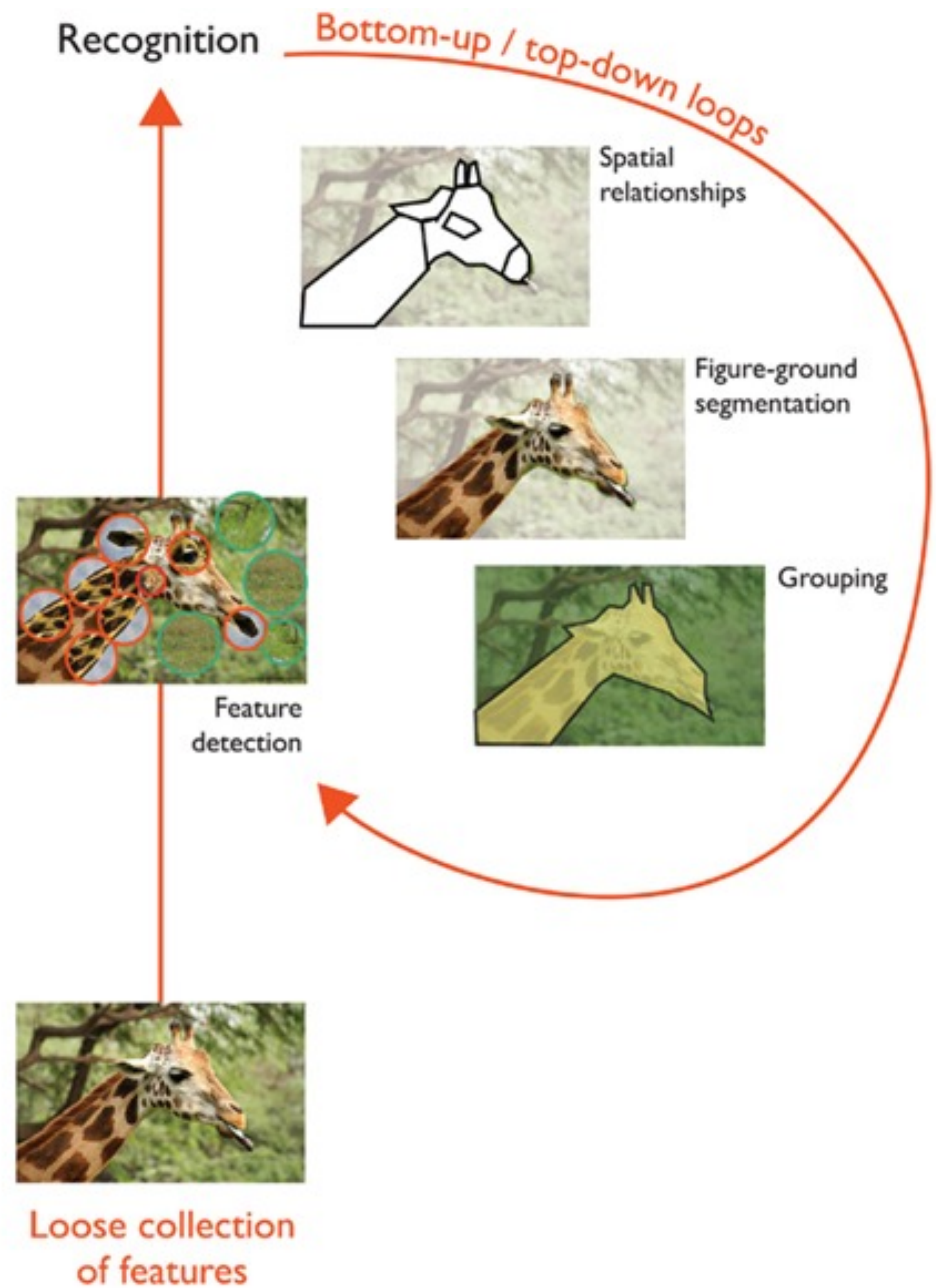


Marr '82;
Biederman '87

Two classes of models

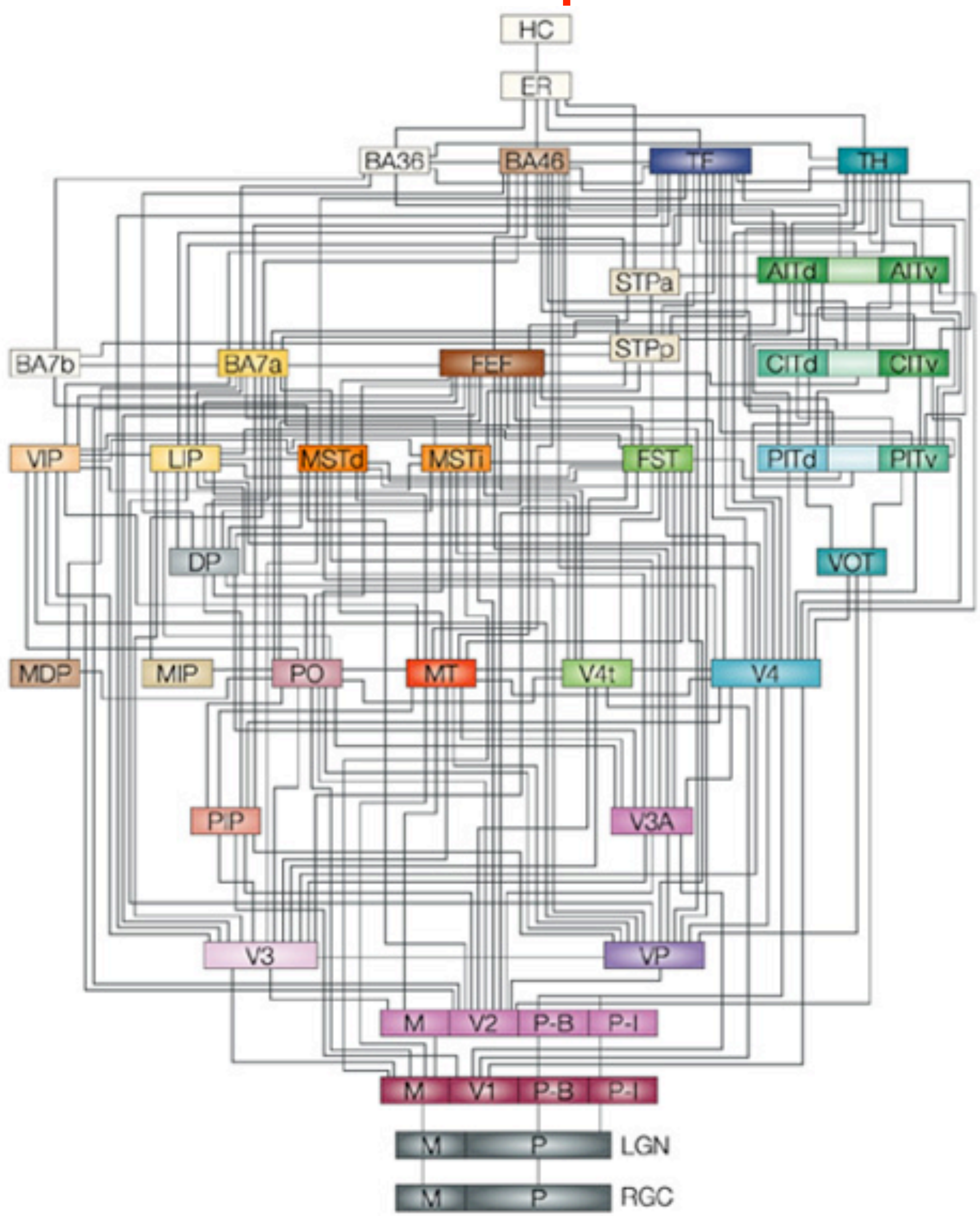
I. Fundamentals of primate vision



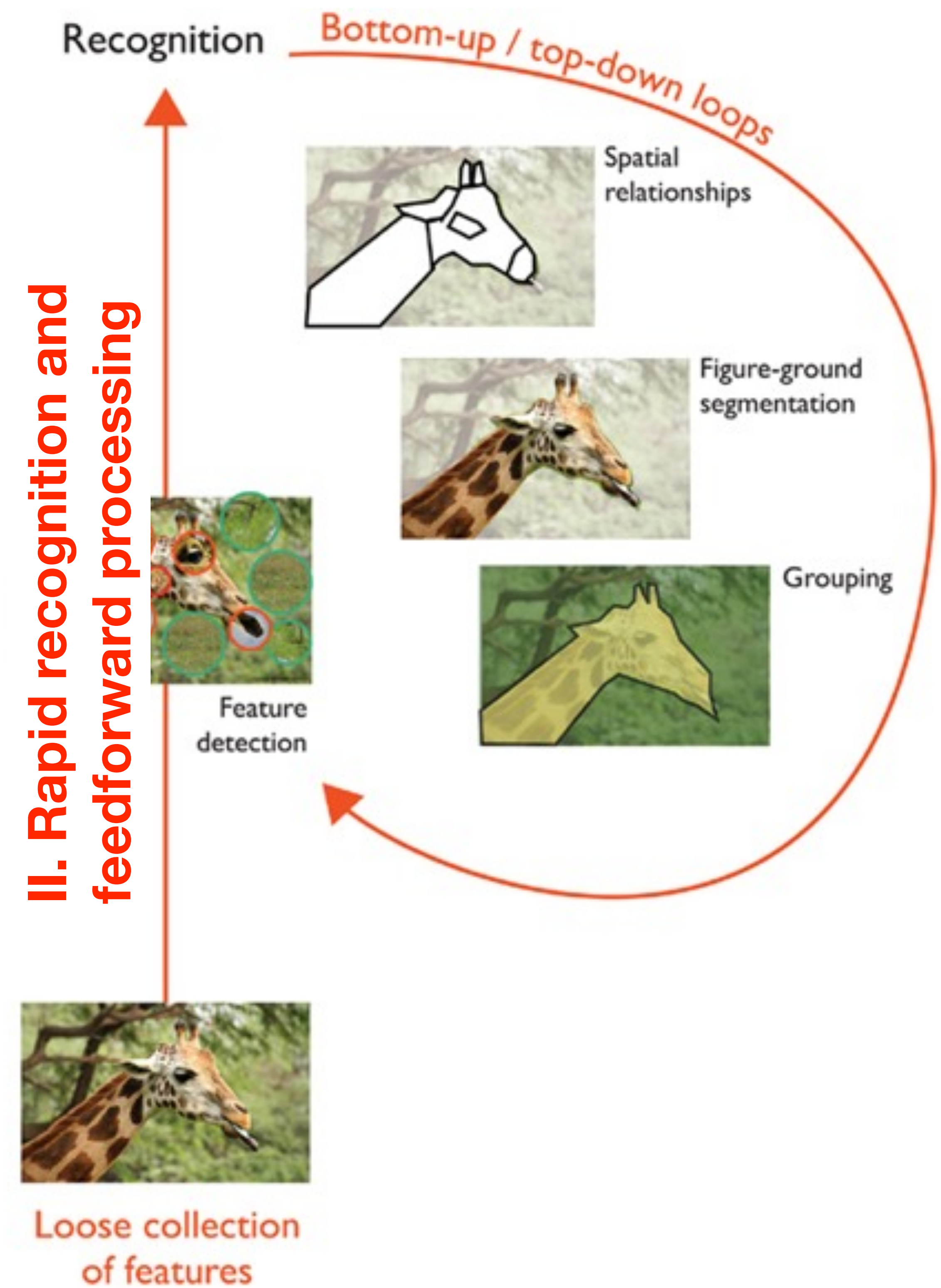


Two classes of models

I. Fundamentals of primate vision

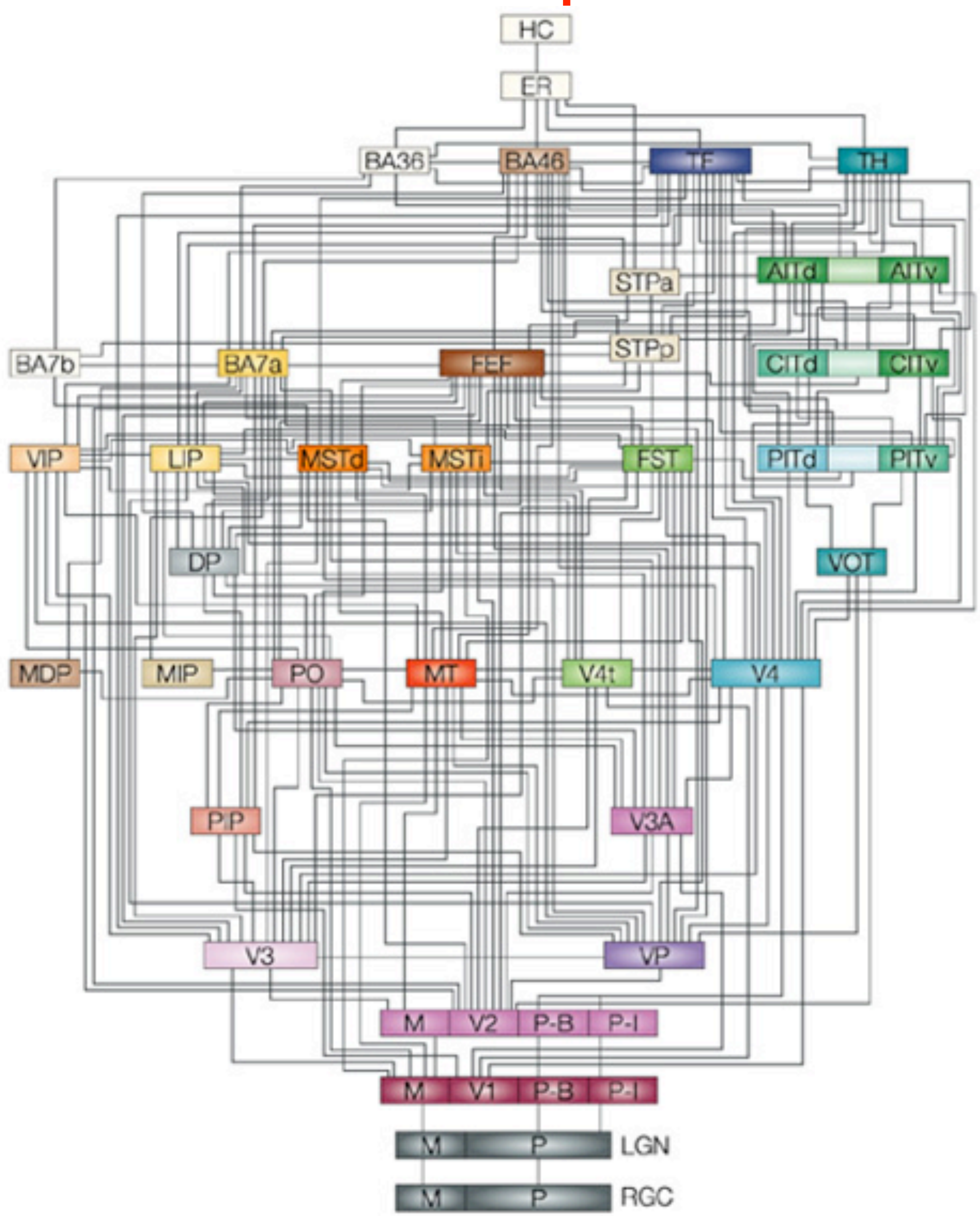


II. Rapid recognition and feedforward processing

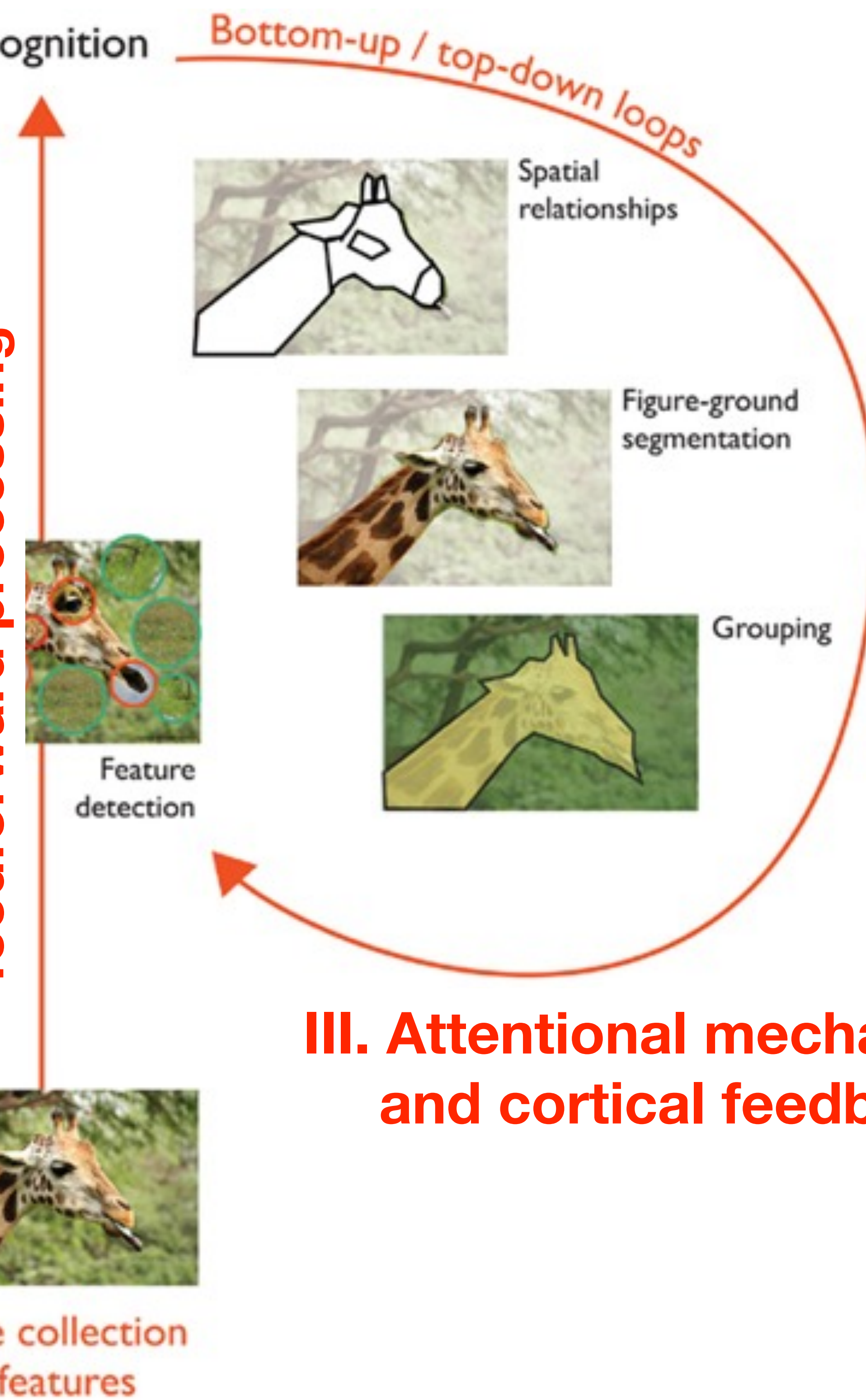


Two classes of models

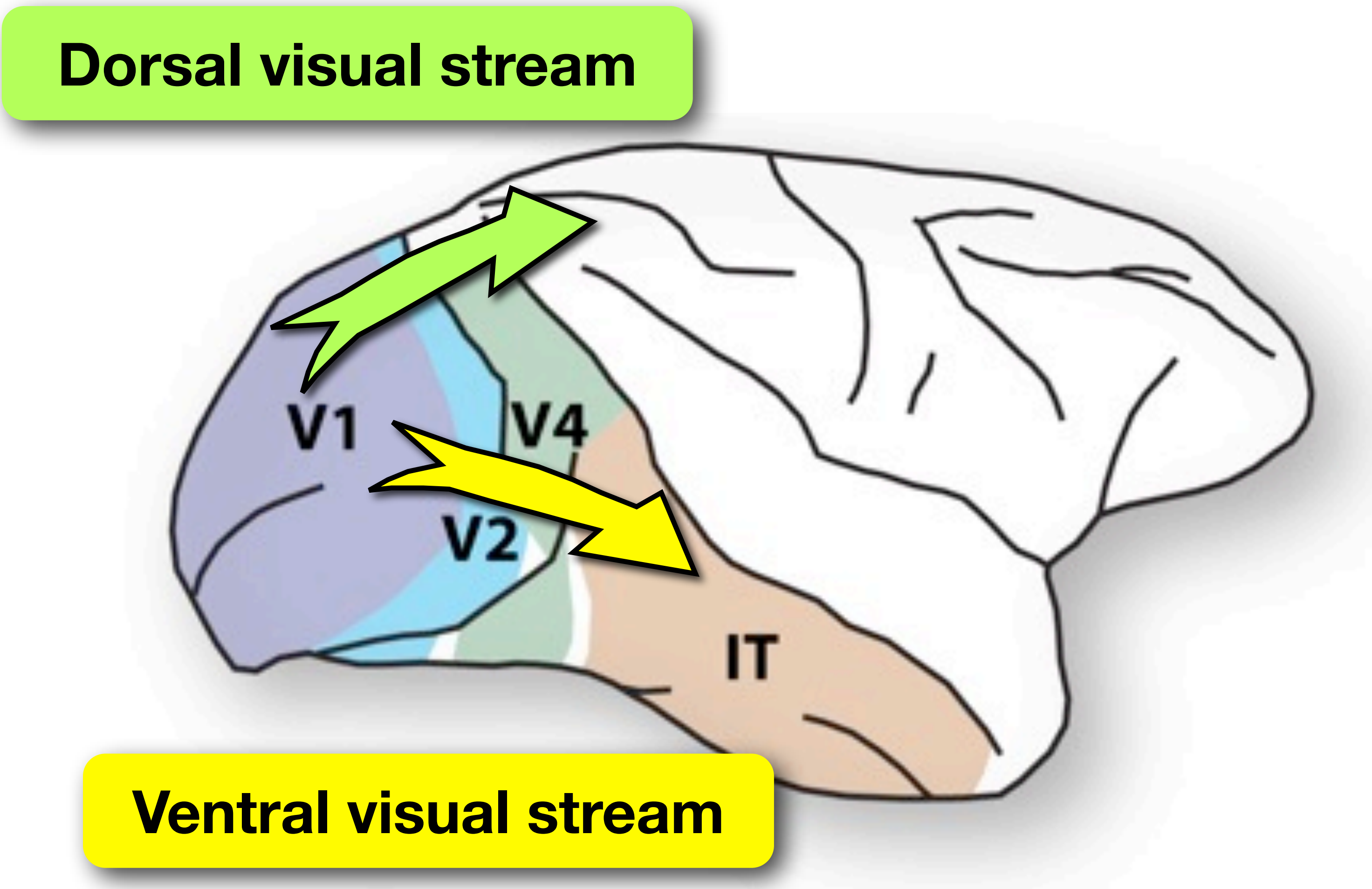
I. Fundamentals of primate vision



II. Rapid recognition and feedforward processing

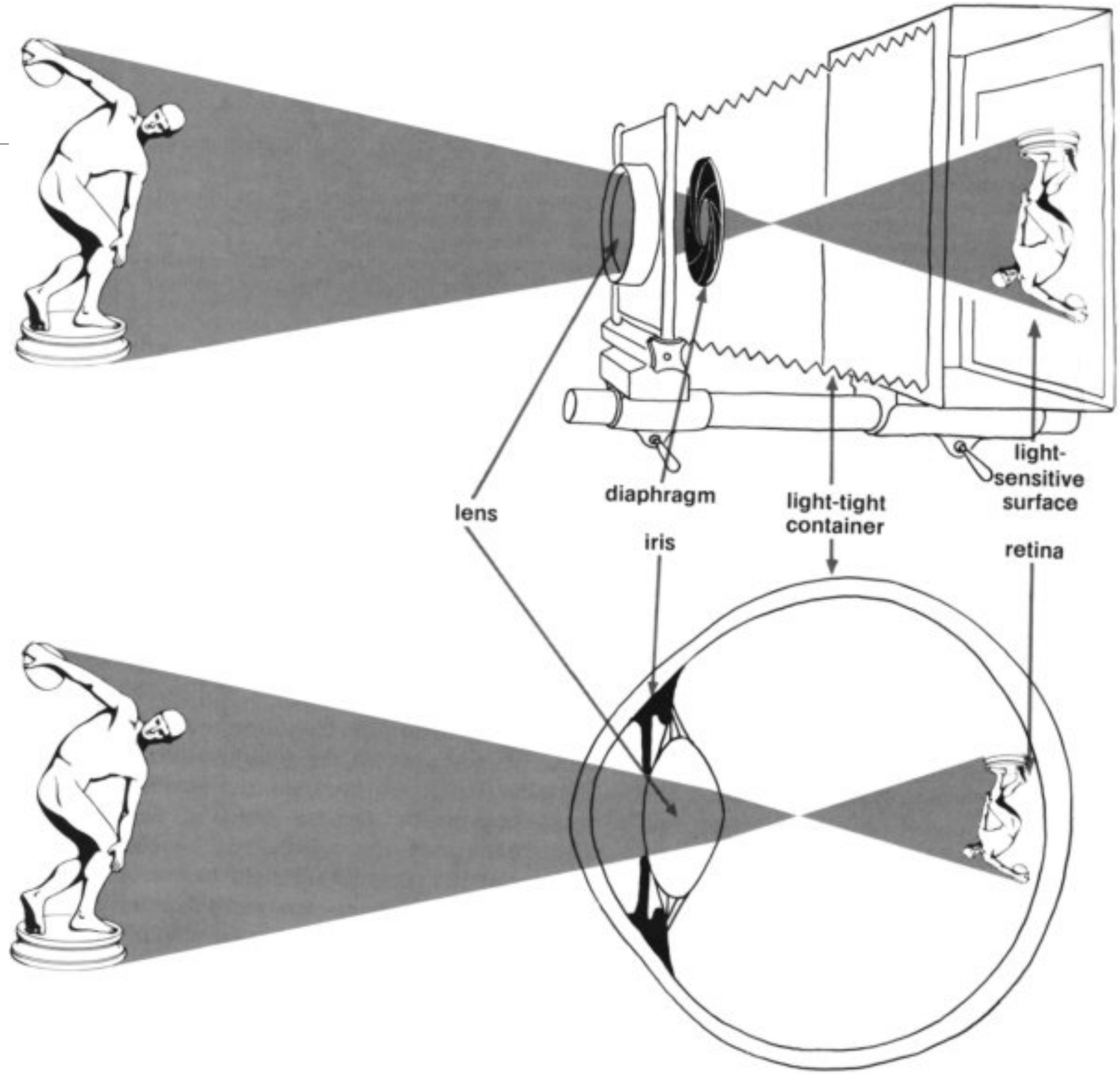


III. Attentional mechanisms and cortical feedback



Streams of processing

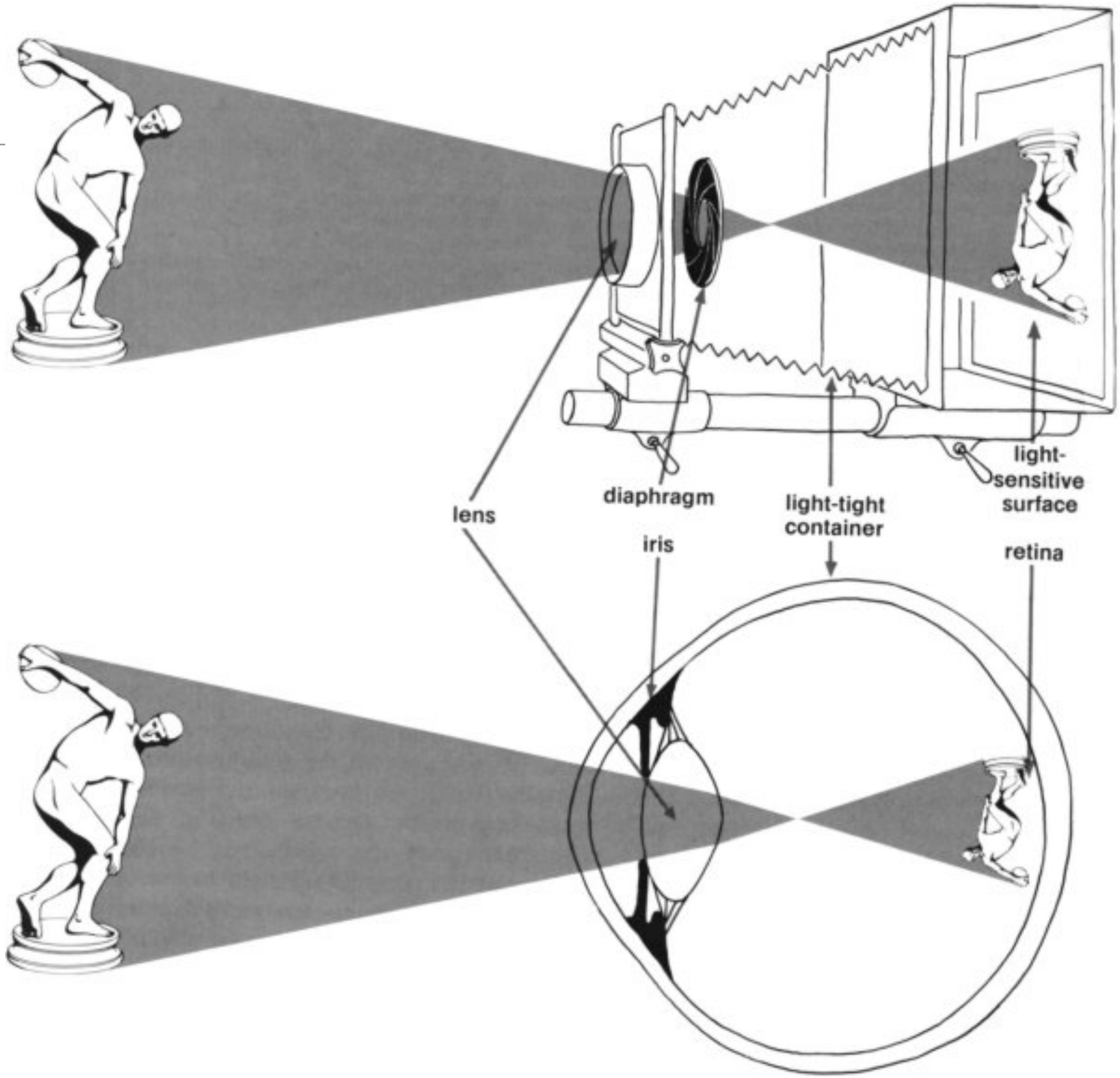
Simple eye vs. camera



Source: unknown

Simple eye vs. camera

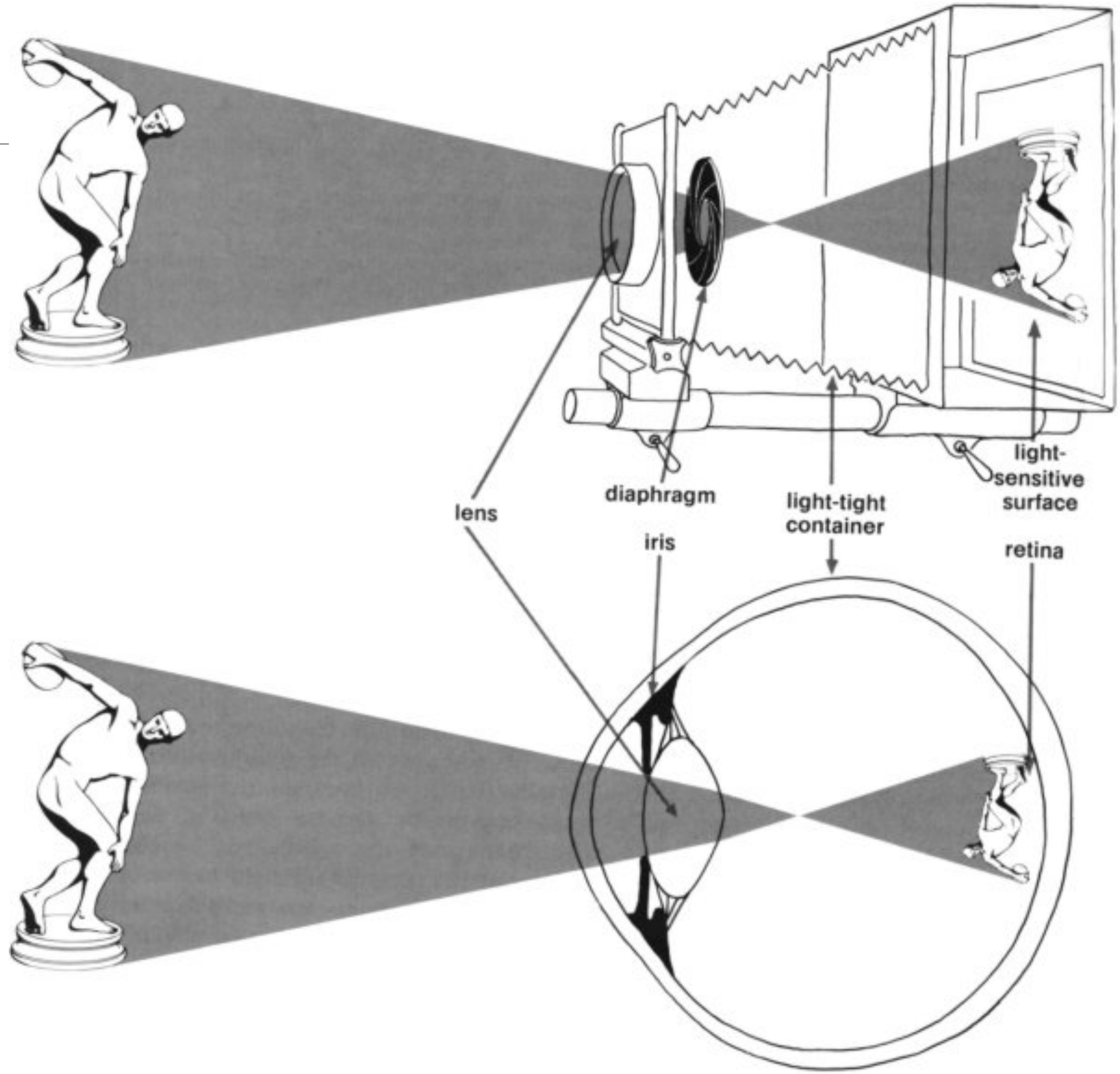
- ~10M pixels



Source: unknown

Simple eye vs. camera

- ~10M pixels



- ~100M photoreceptors
  - ~5M cones
  - ~90M rods

Source: unknown

Eye stimulated by stuff in the world

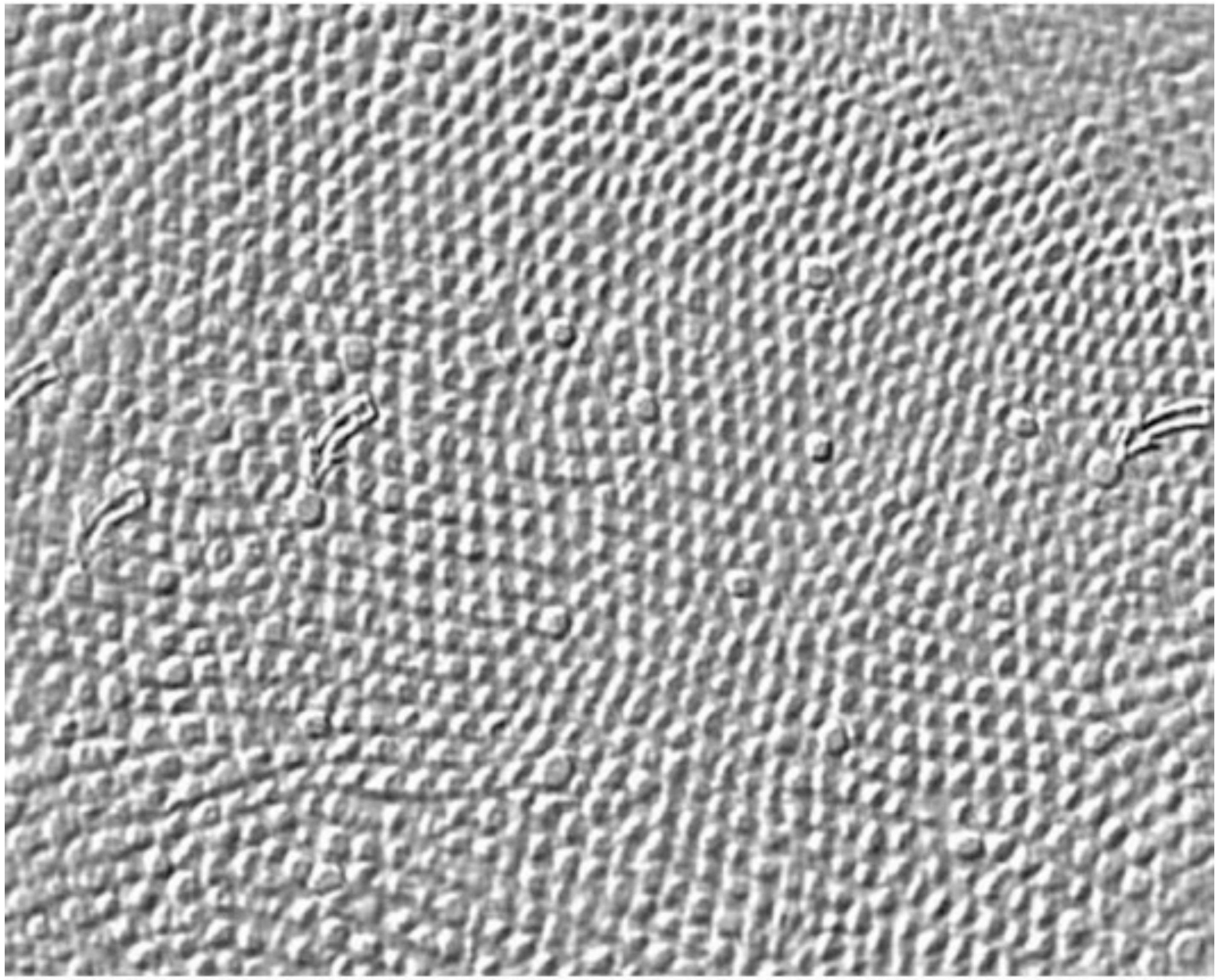


Fig. 13. Tangential section through the human fovea.
Larger cones (arrows) are blue cones. *From Ahnelt et al. 1987.*

Source: webvision

Eye stimulated by stuff in the world

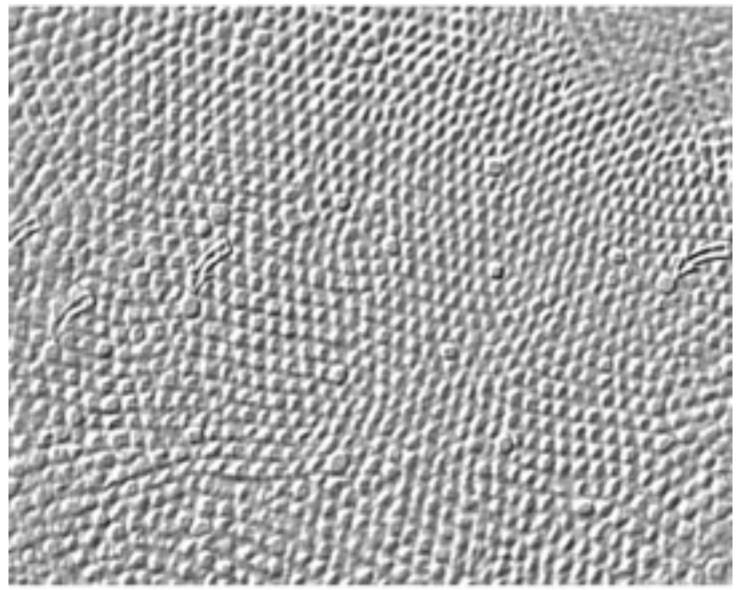
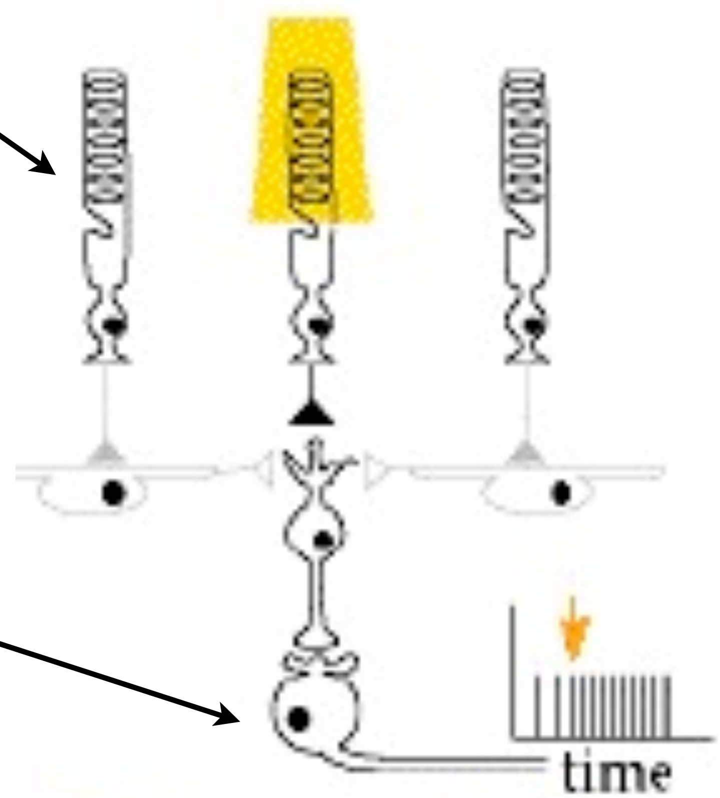


Fig. 13. Tangential section through the human fovea. Larger cones (arrows) are blue cones. From Ahnelt et al. 1987.

photoreceptors



Retinal output
(ganglion cells)

Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

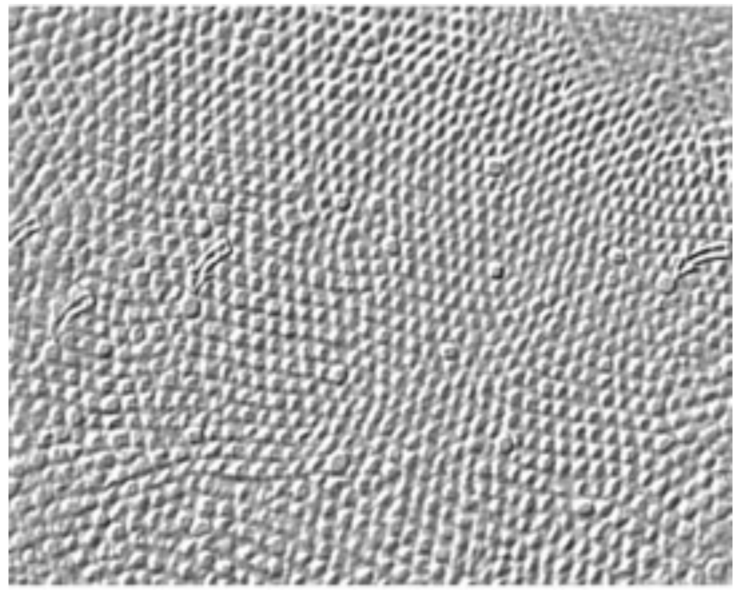
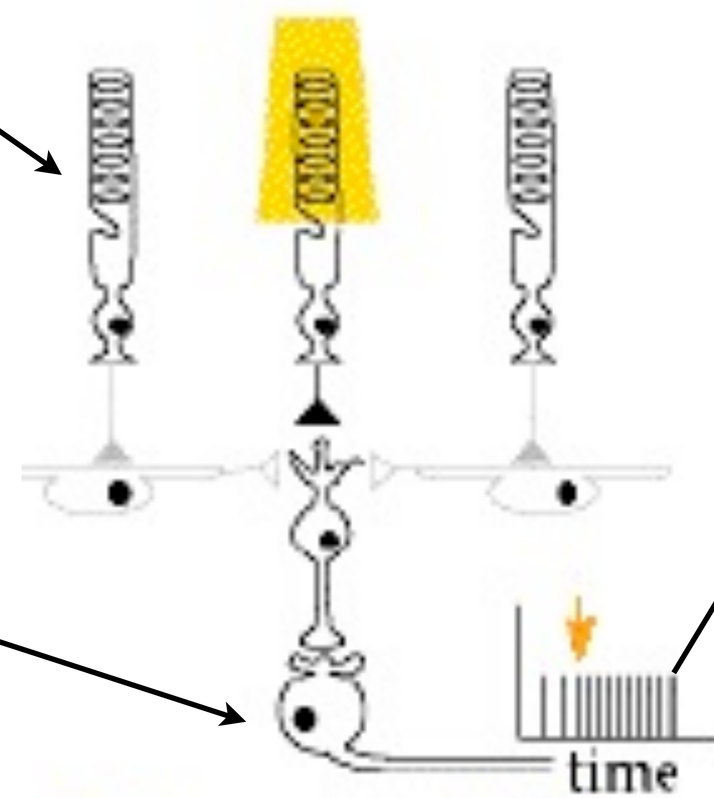
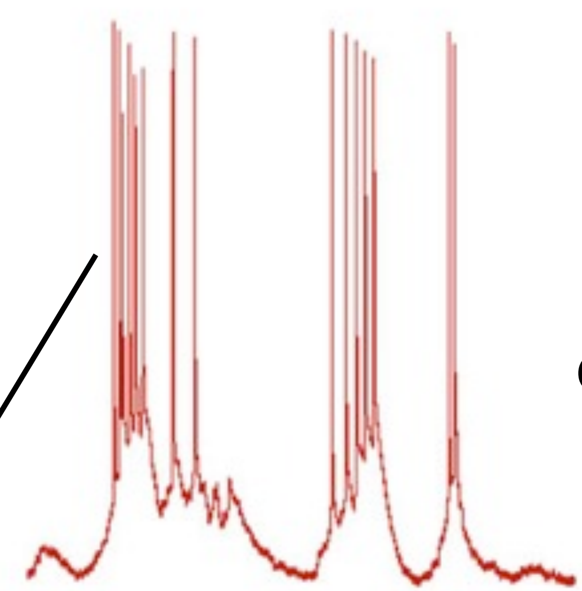


Fig. 13. Tangential section through the human fovea. Larger cones (arrows) are blue cones. From Ahnelt et al. 1987.

photoreceptors



Retinal output (ganglion cells)



Increase in firing rate driven by external stimulus

Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

source: webvision

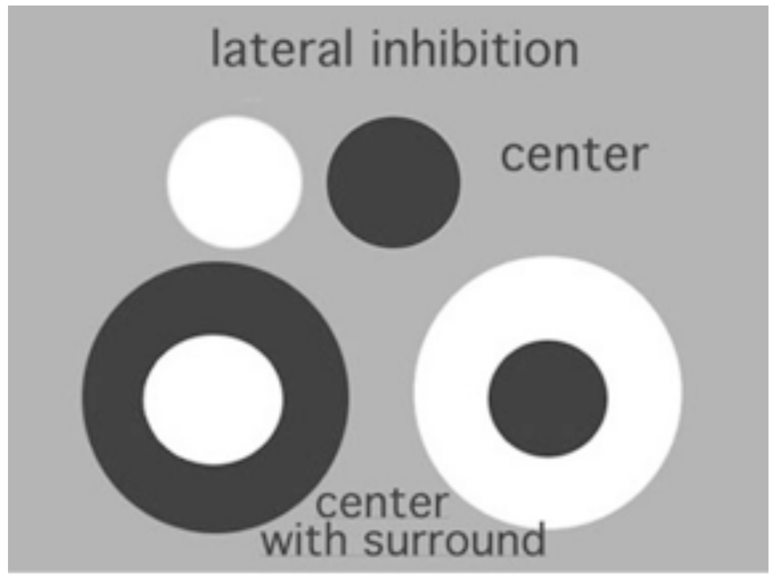
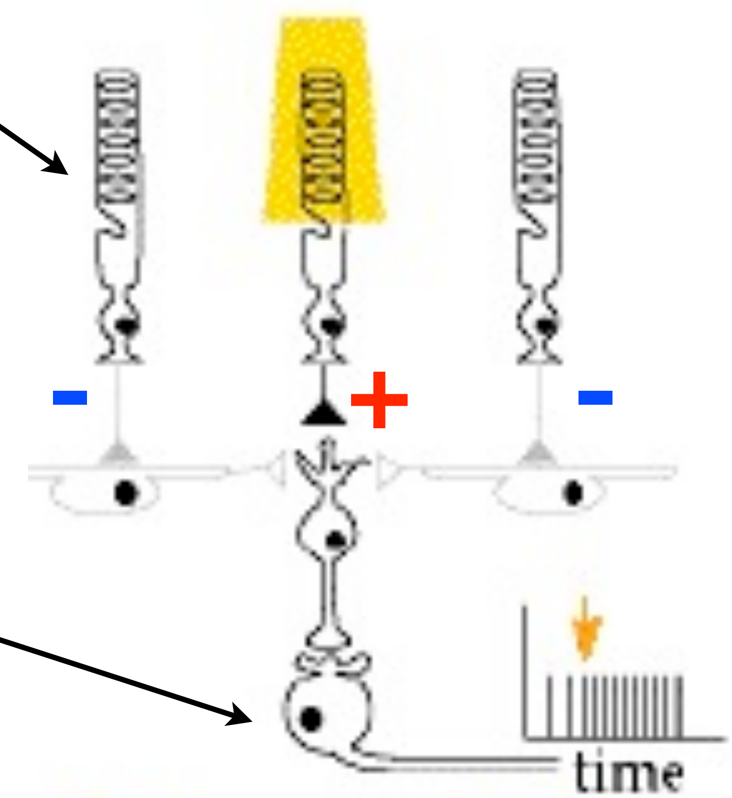


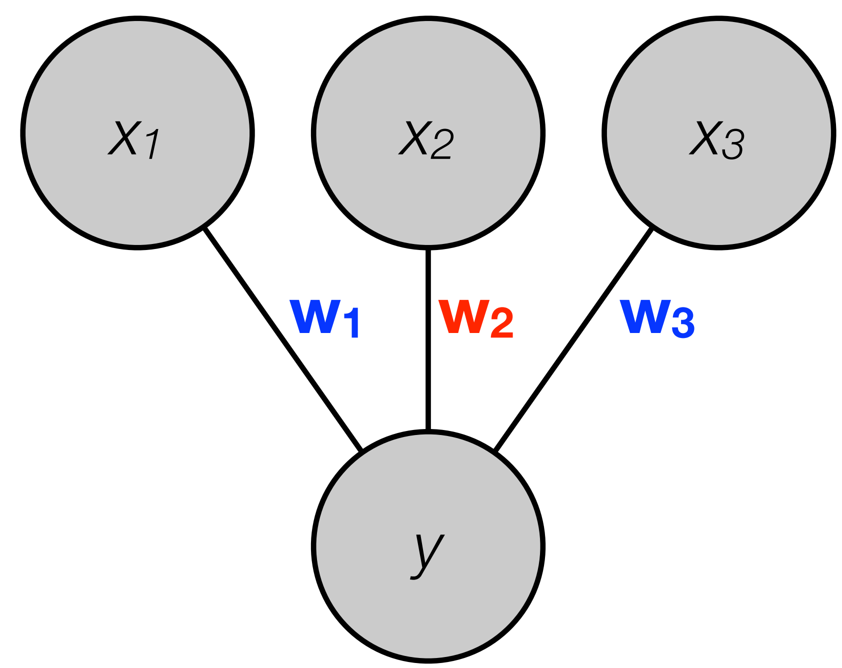
Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

$$y = f\left(\sum w_i x_i\right)$$

photoreceptors



Retinal output
(ganglion cells)



Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

source: webvision

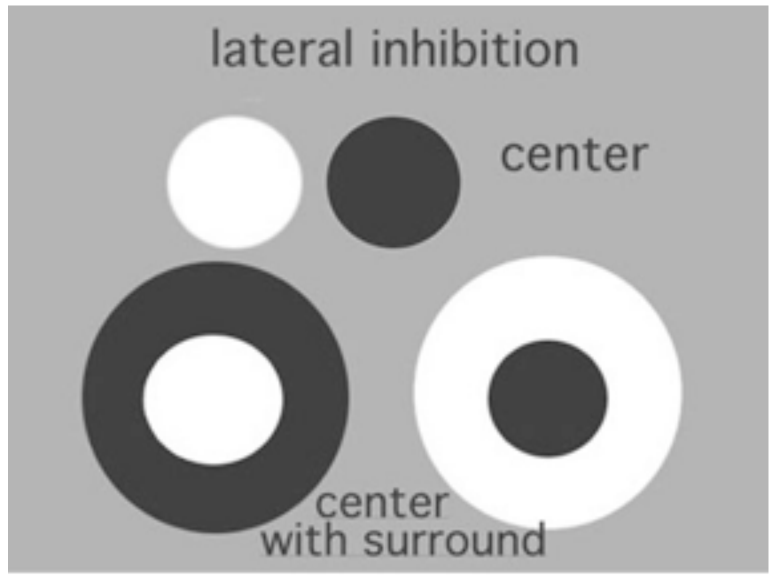
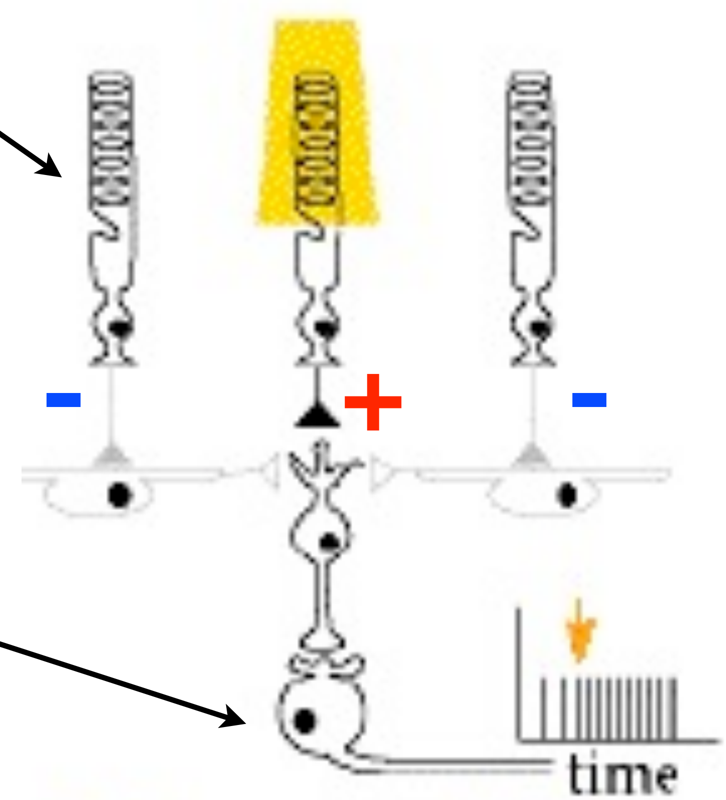


Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

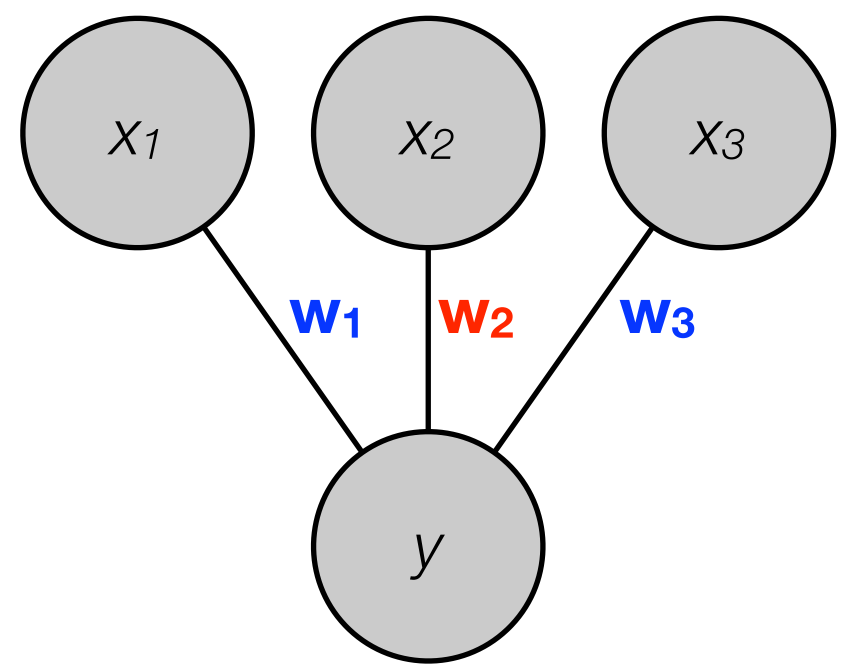
photoreceptors



Retinal output
(ganglion cells)

$$y = f\left(\sum w_i x_i\right)$$

Afferent activity
from neuron i



Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

source: webvision

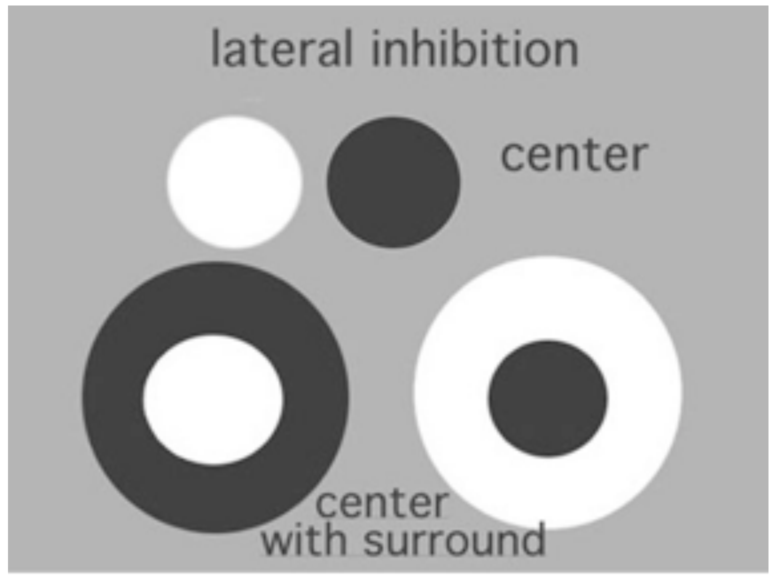
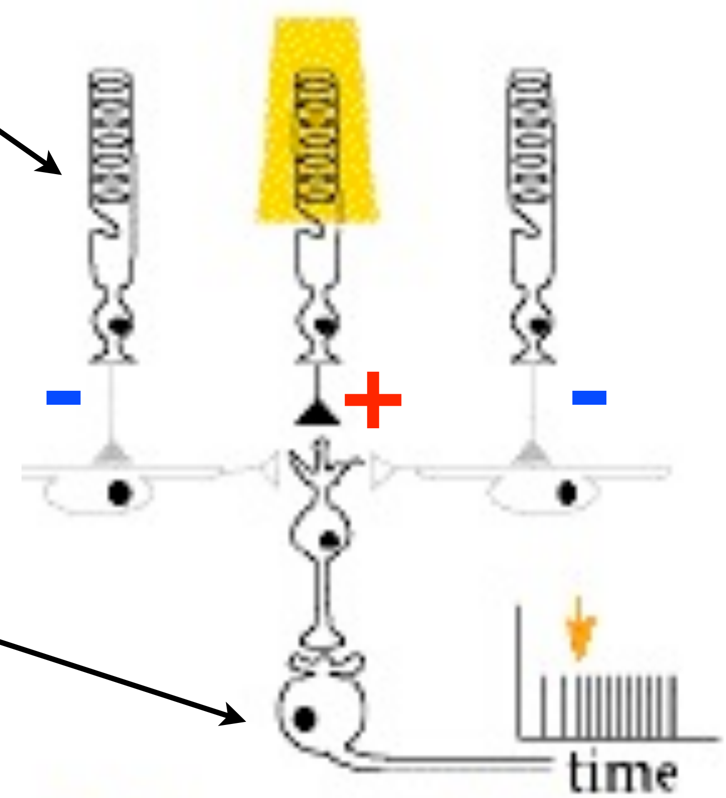


Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

photoreceptors

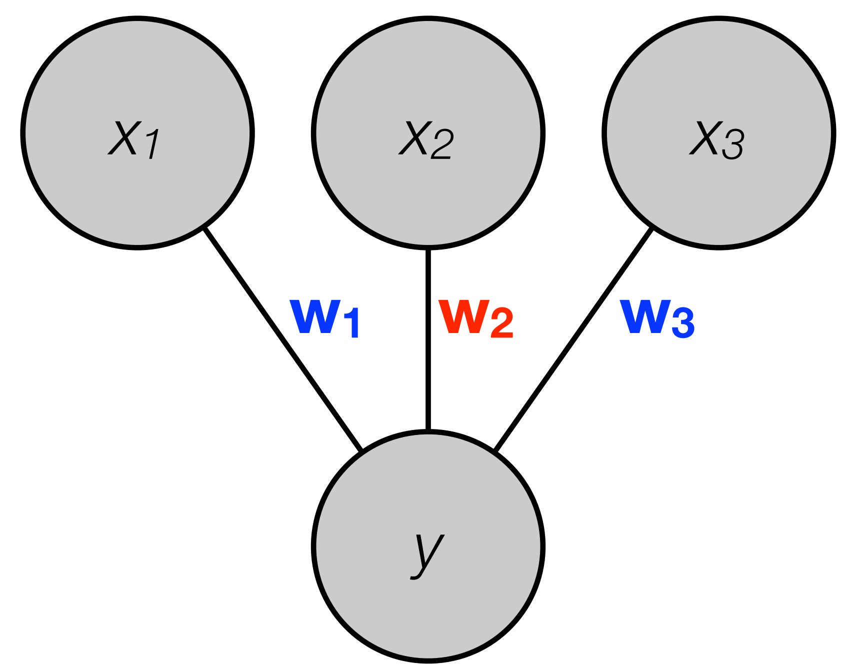


Retinal output
(ganglion cells)

$$y = f\left(\sum w_i x_i\right)$$

Synaptic efficacies

Afferent activity from neuron i



Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

source: webvision

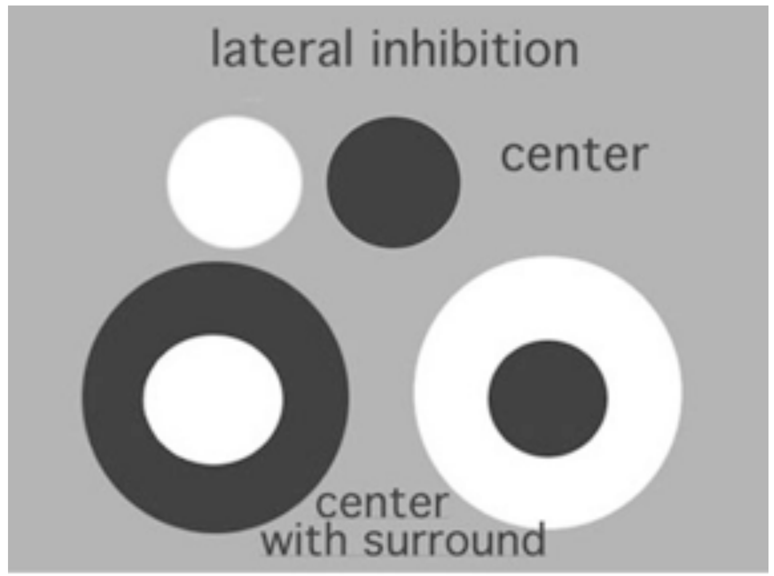
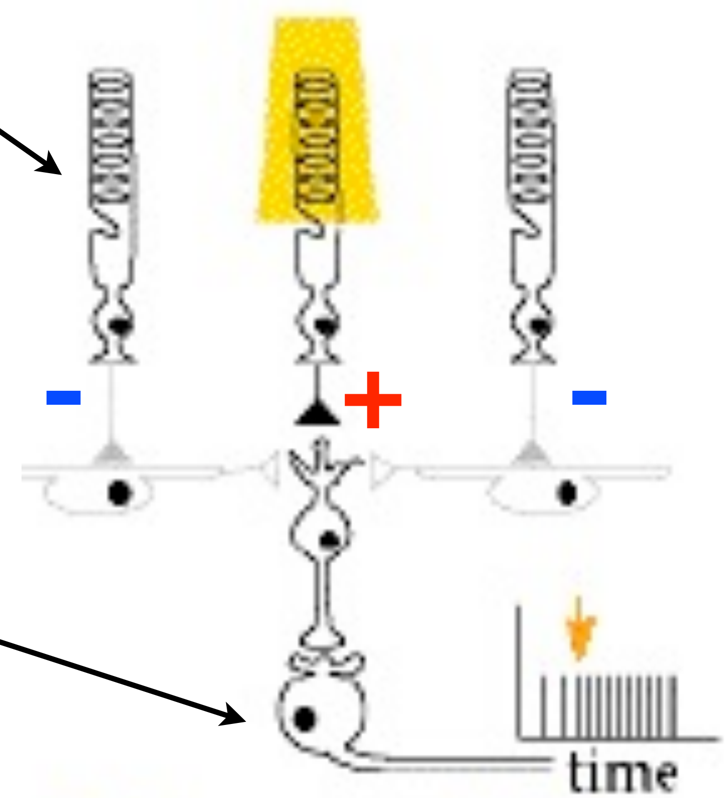


Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

photoreceptors



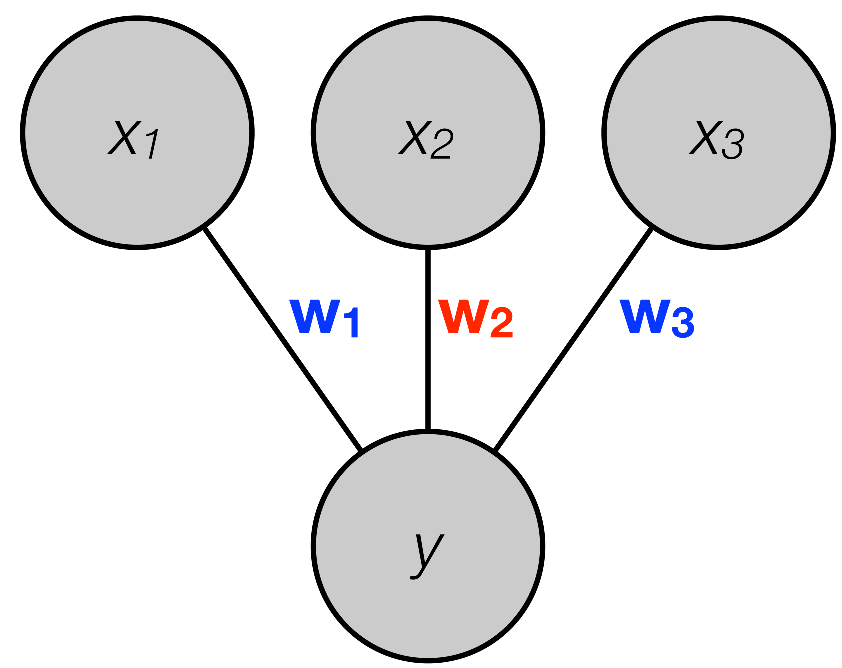
Retinal output
(ganglion cells)

Summation at the soma

$$y = f\left(\sum w_i x_i\right)$$

Synaptic efficacies

Afferent activity from neuron i



Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

source: webvision

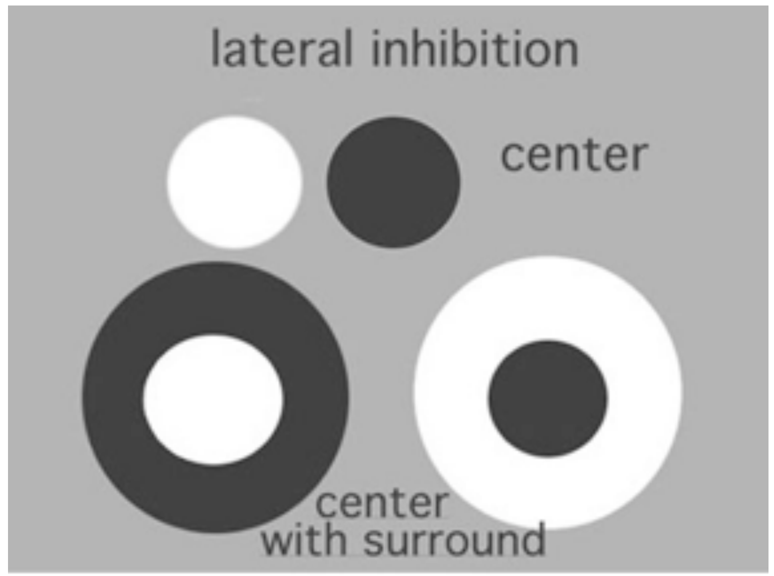
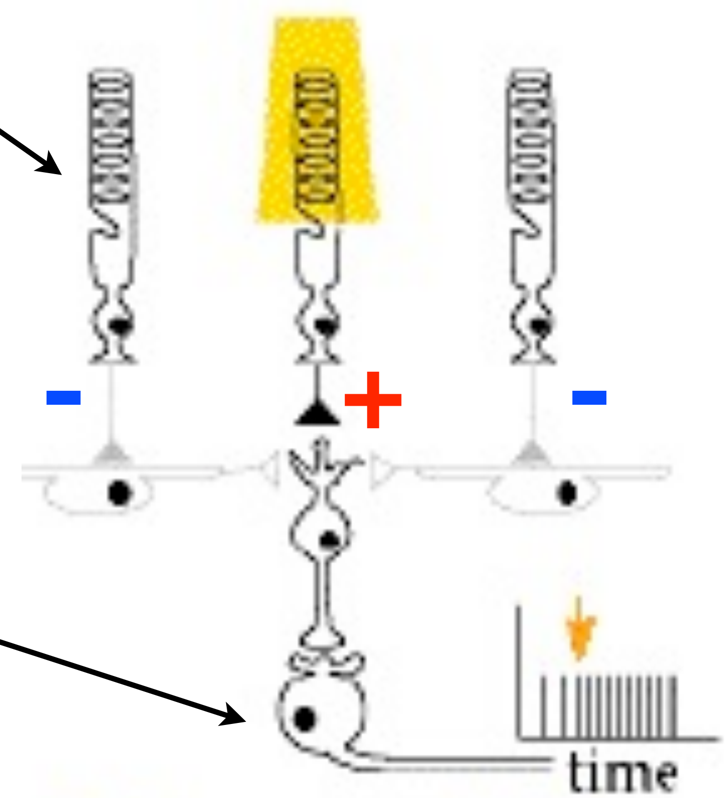


Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

photoreceptors



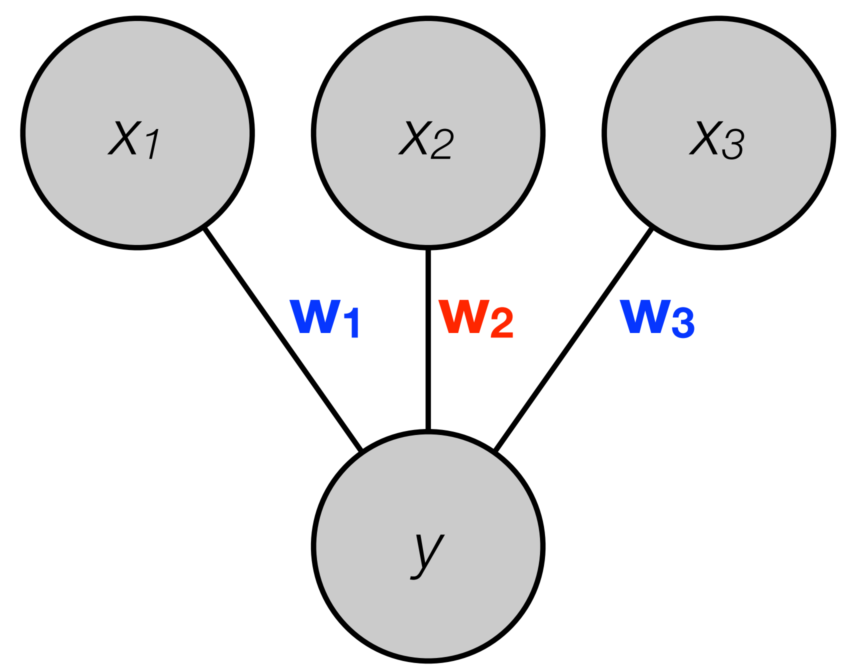
Retinal output
(ganglion cells)

Summation at
the soma

$$y = f\left(\sum w_i x_i\right)$$

Synaptic
efficacies

Afferent activity
from neuron i



Modified from http://thalamus.wustl.edu/course/eyeret.html

Eye stimulated by stuff in the world

source: webvision

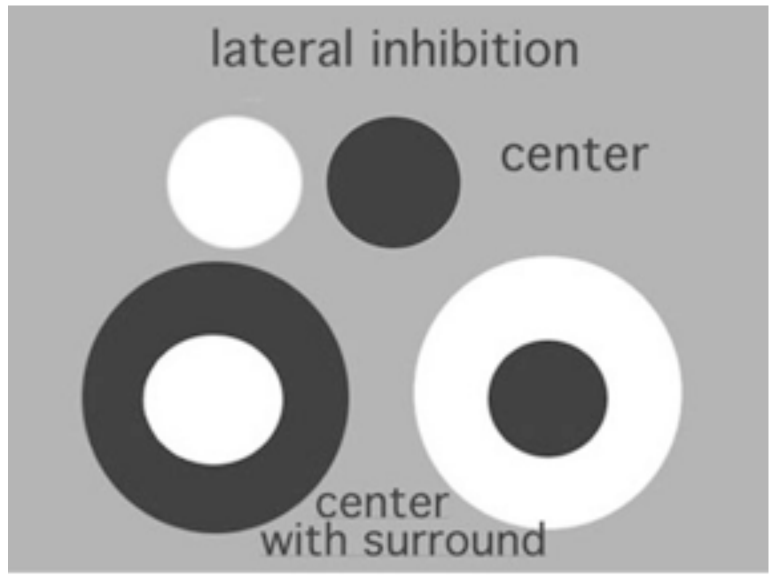
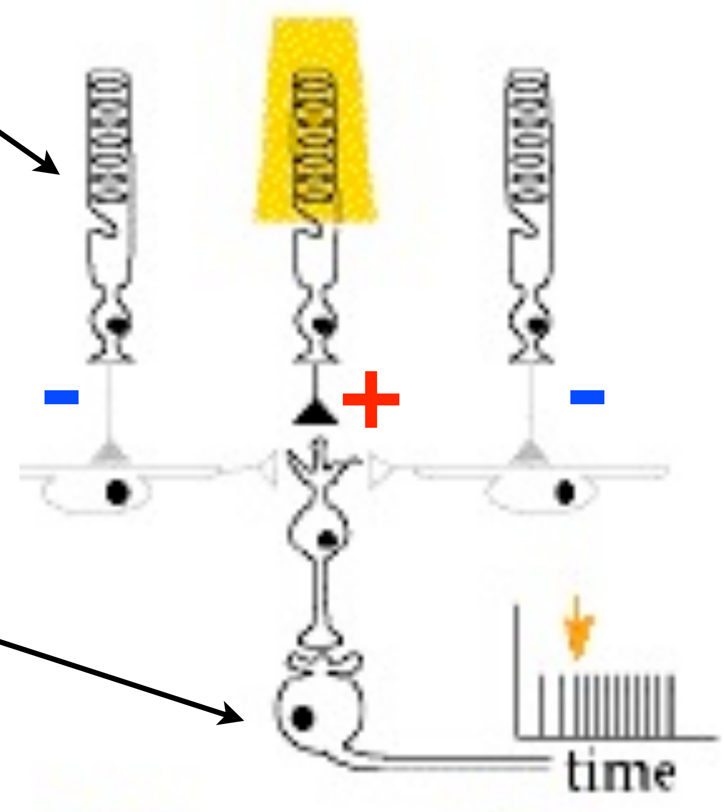


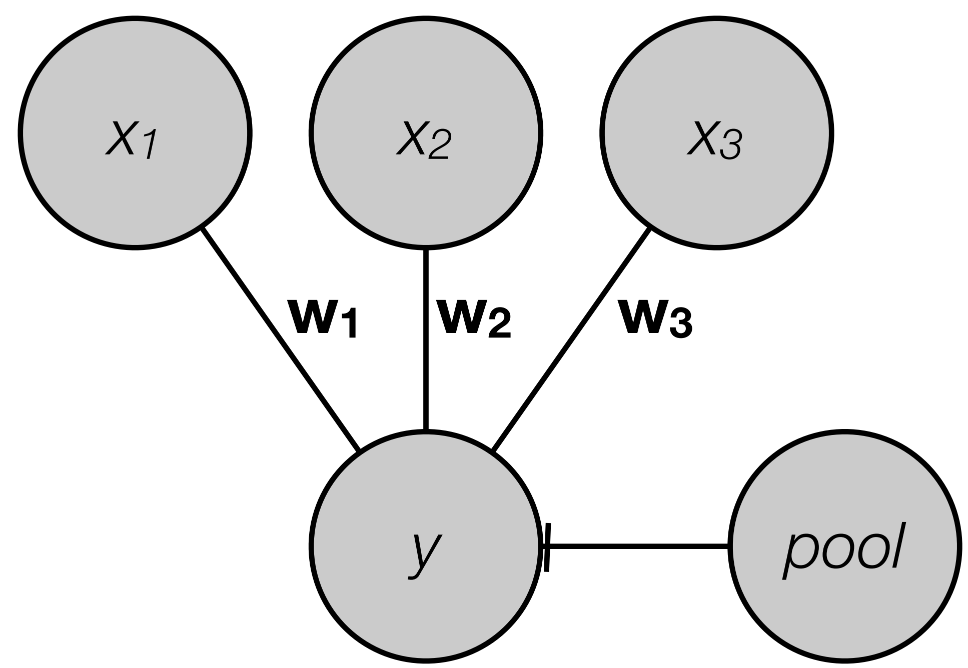
Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

$$y = f\left(\sum w_i x_i\right) / (\text{pool})$$

photoreceptors



Retinal output
(ganglion cells)



Eye stimulated by stuff in the world

Normalization via pool of neighboring units (e.g., tuned to different orientations)

source: webvision

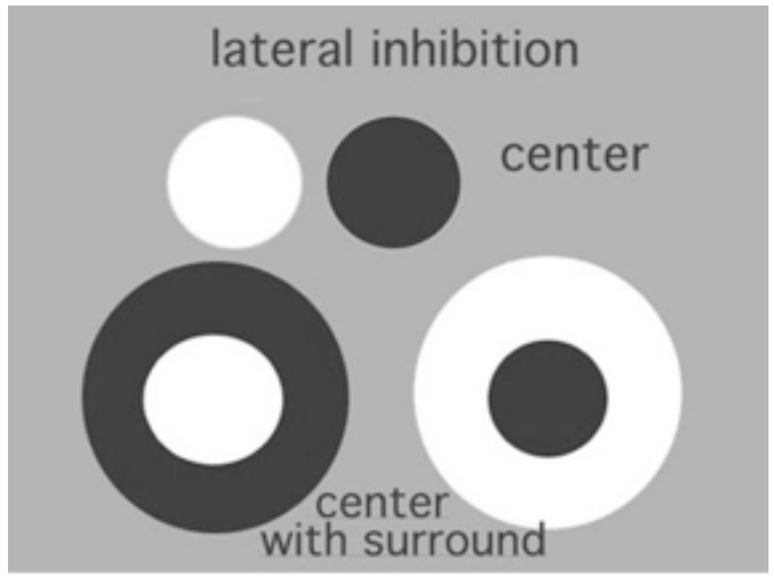
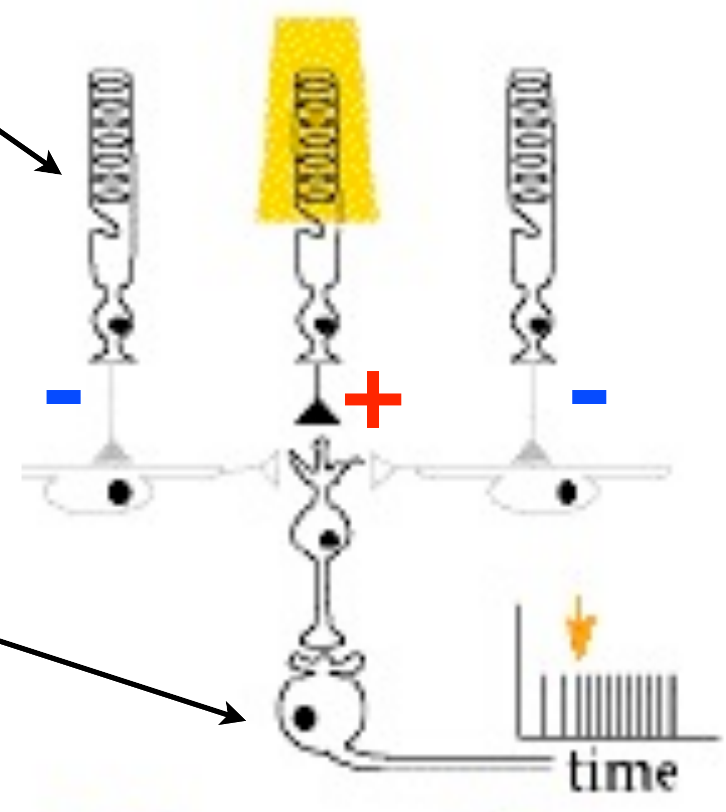


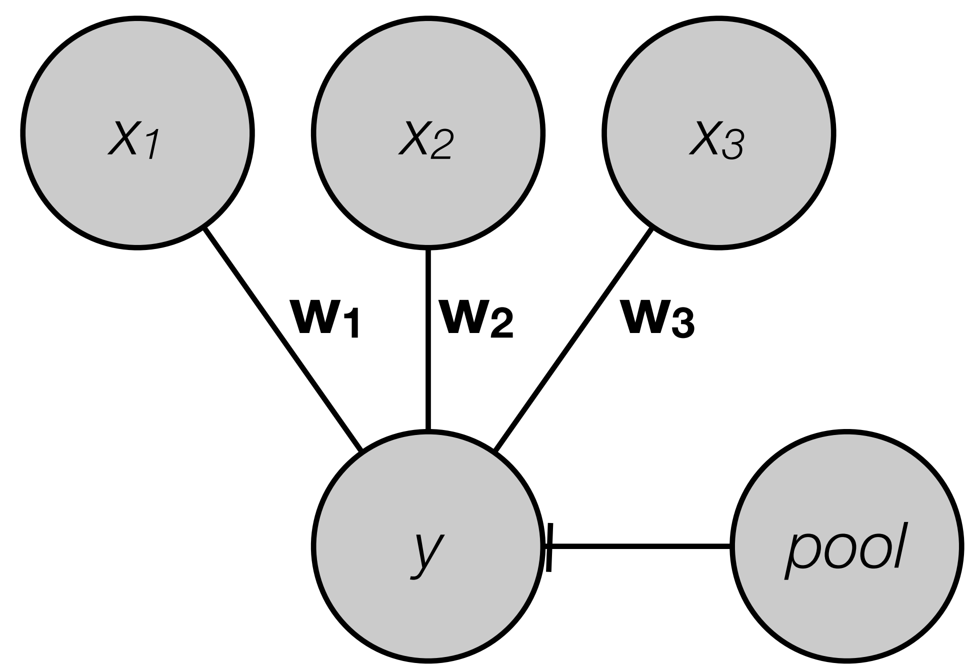
Fig. 10. Center-surround receptive fields can be ON center or OFF center with the oposite sign annular surround.

$$y = f\left(\sum w_i x_i\right) / \,(\mathrm{pool})$$

photoreceptors



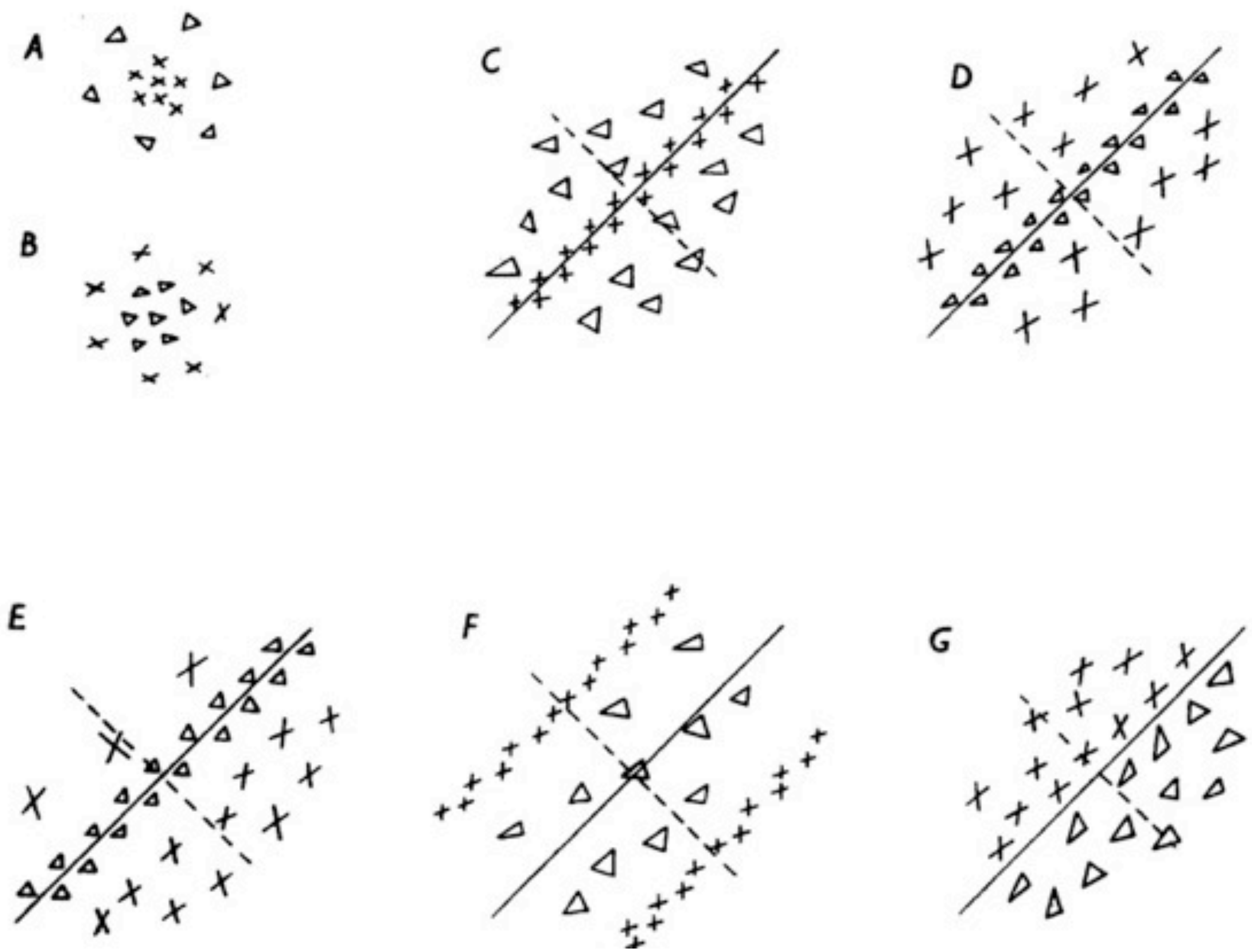
Retinal output (ganglion cells)

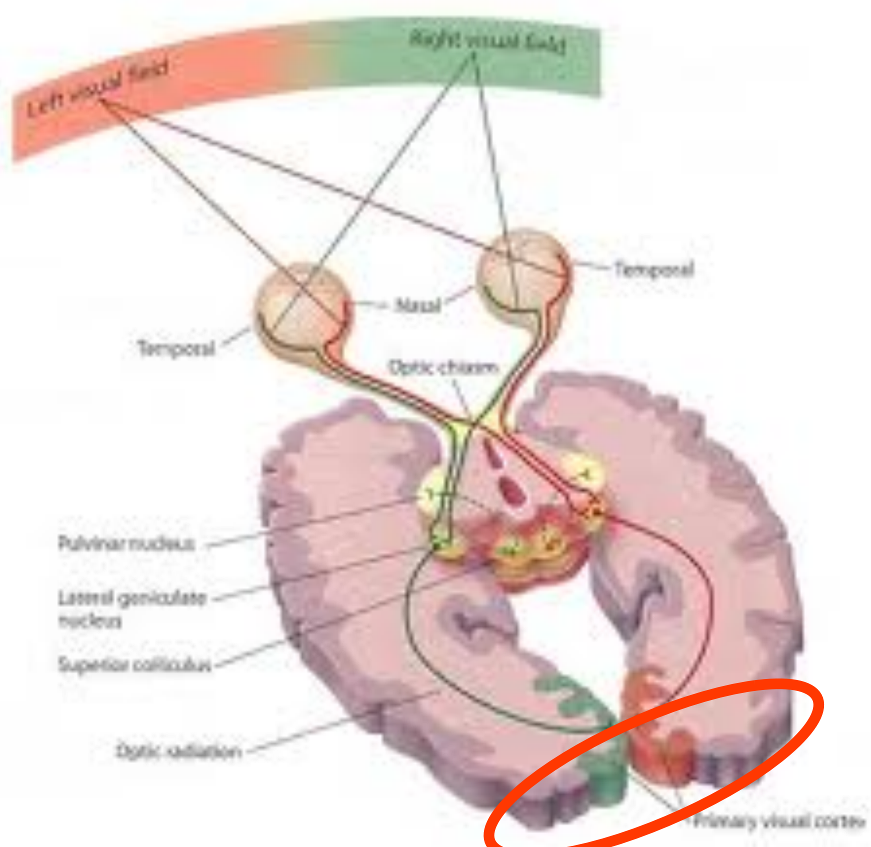


RF organization in V1

Simple cell

Hubel & Wiesel

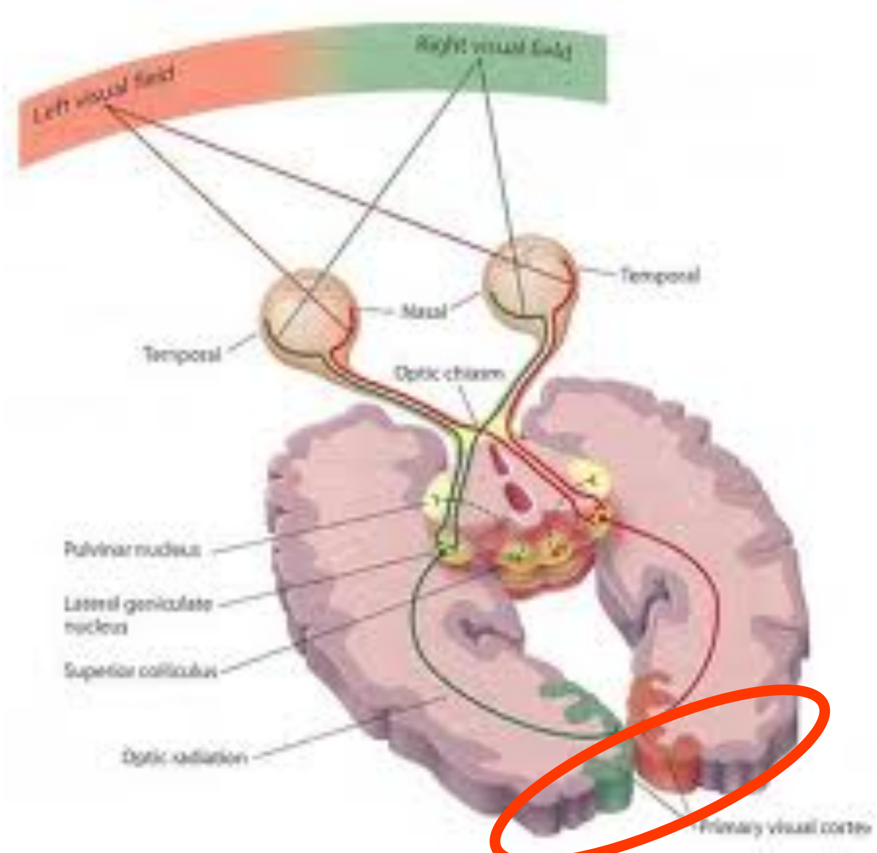




Hubel & Wiesel ’59 ’62 ’68

RF organization in V1

Hubel & Wiesel

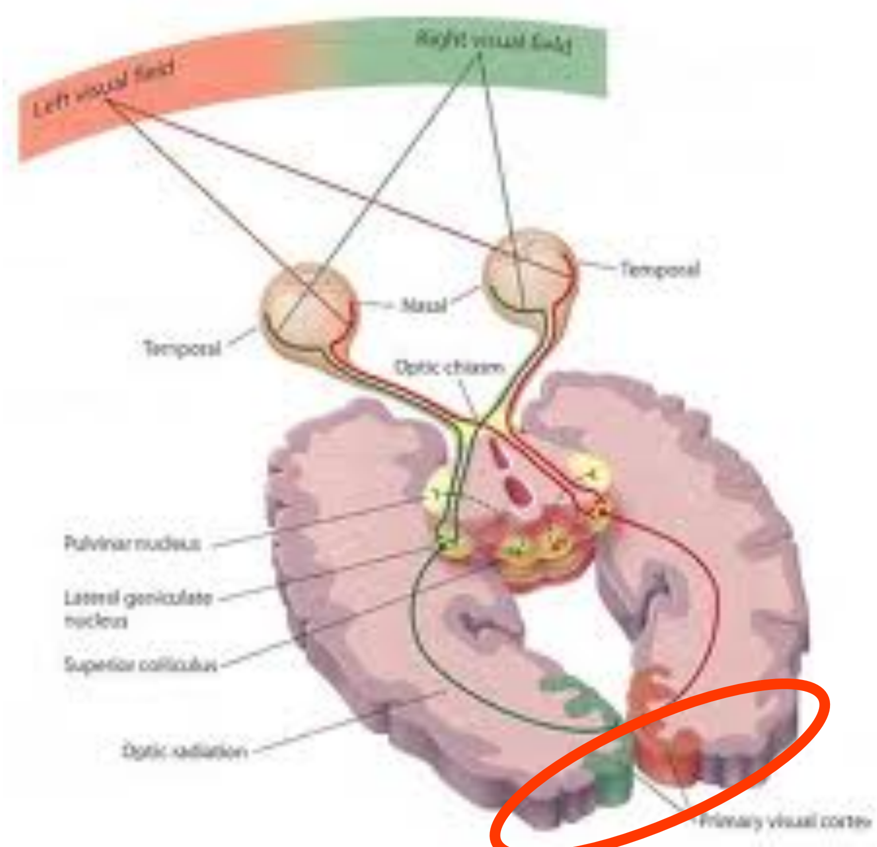


Simple cell

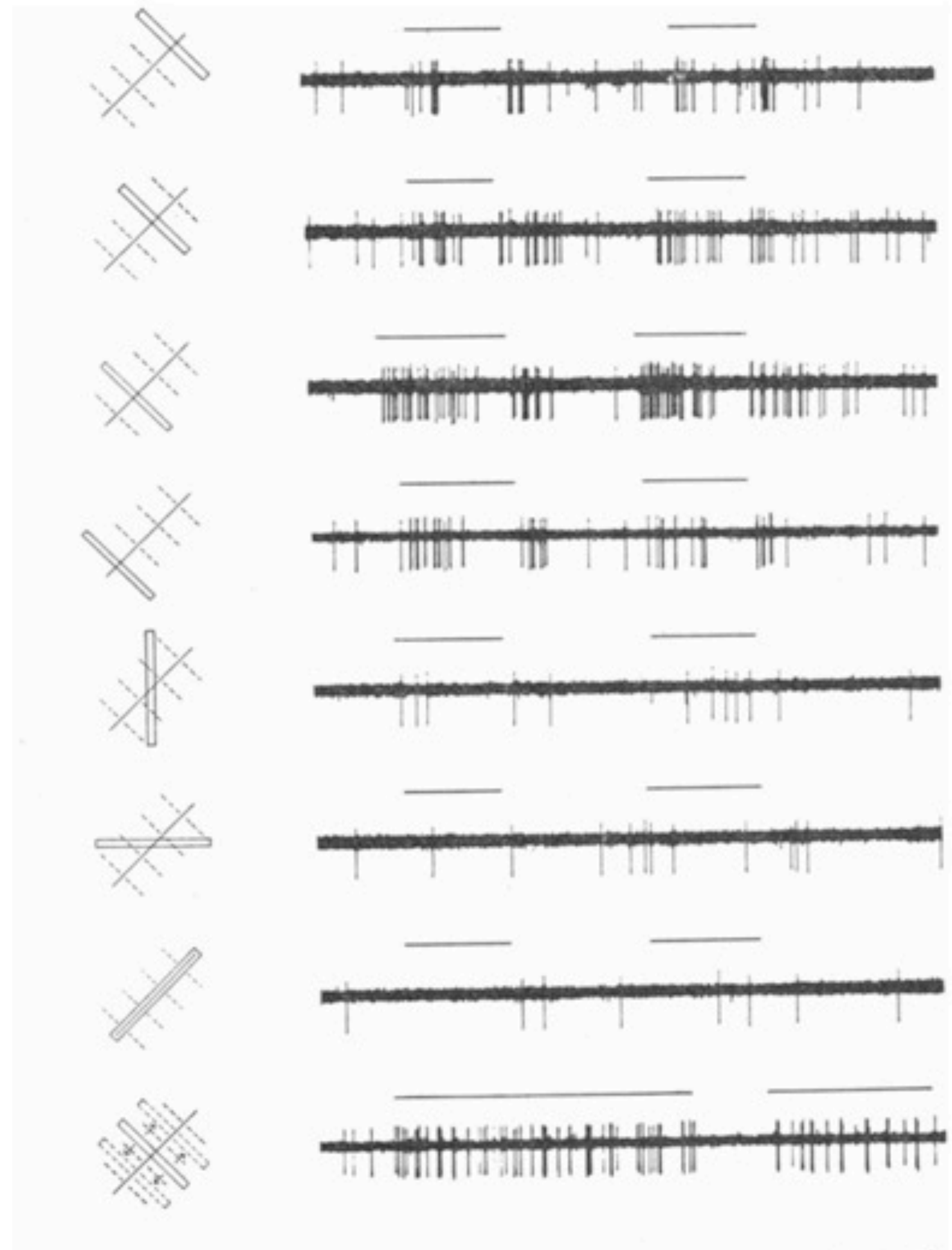
Hubel & Wiesel '59 '62 '68

RF organization in V1

Hubel & Wiesel



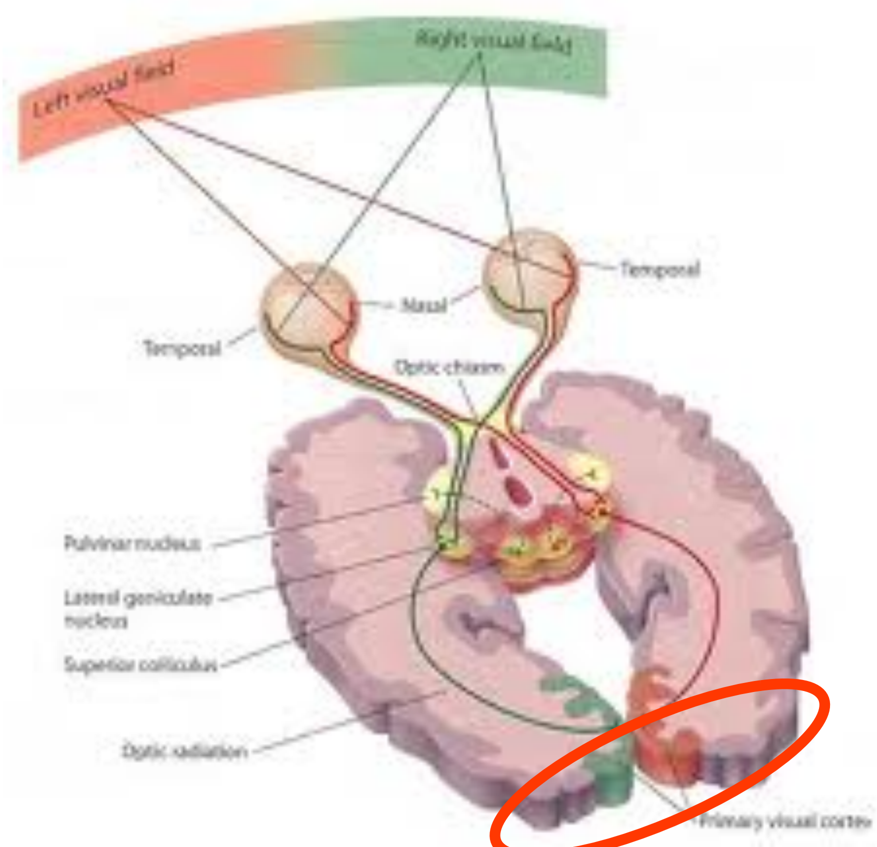
Complex cell



Hubel & Wiesel ’59 ’62 ’68

RF organization in V1

Hubel & Wiesel

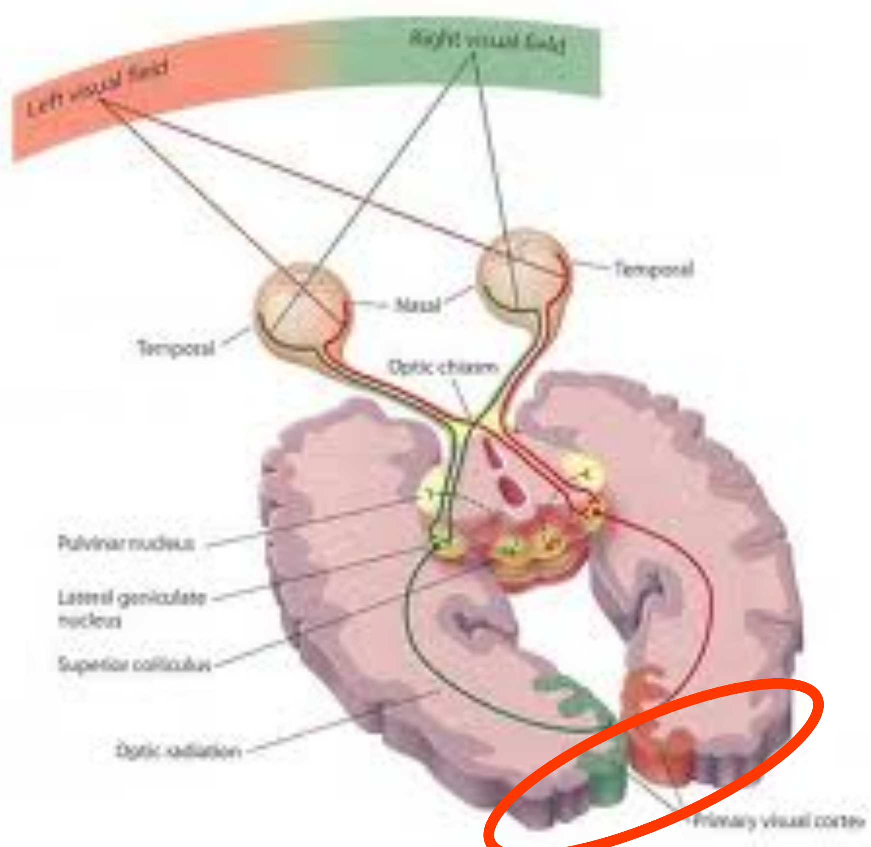


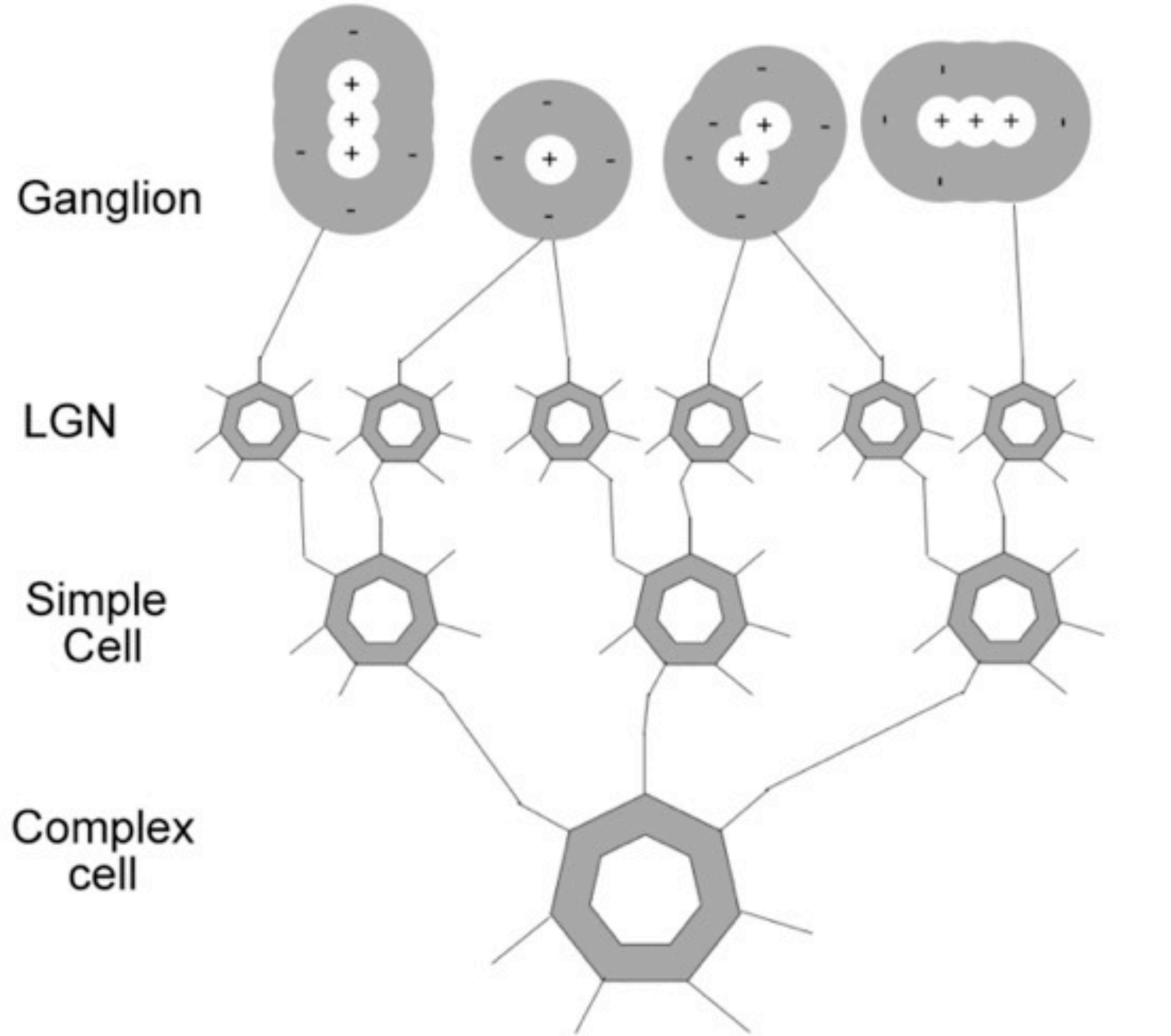
Complex cell

Hubel & Wiesel ’59 ’62 ’68

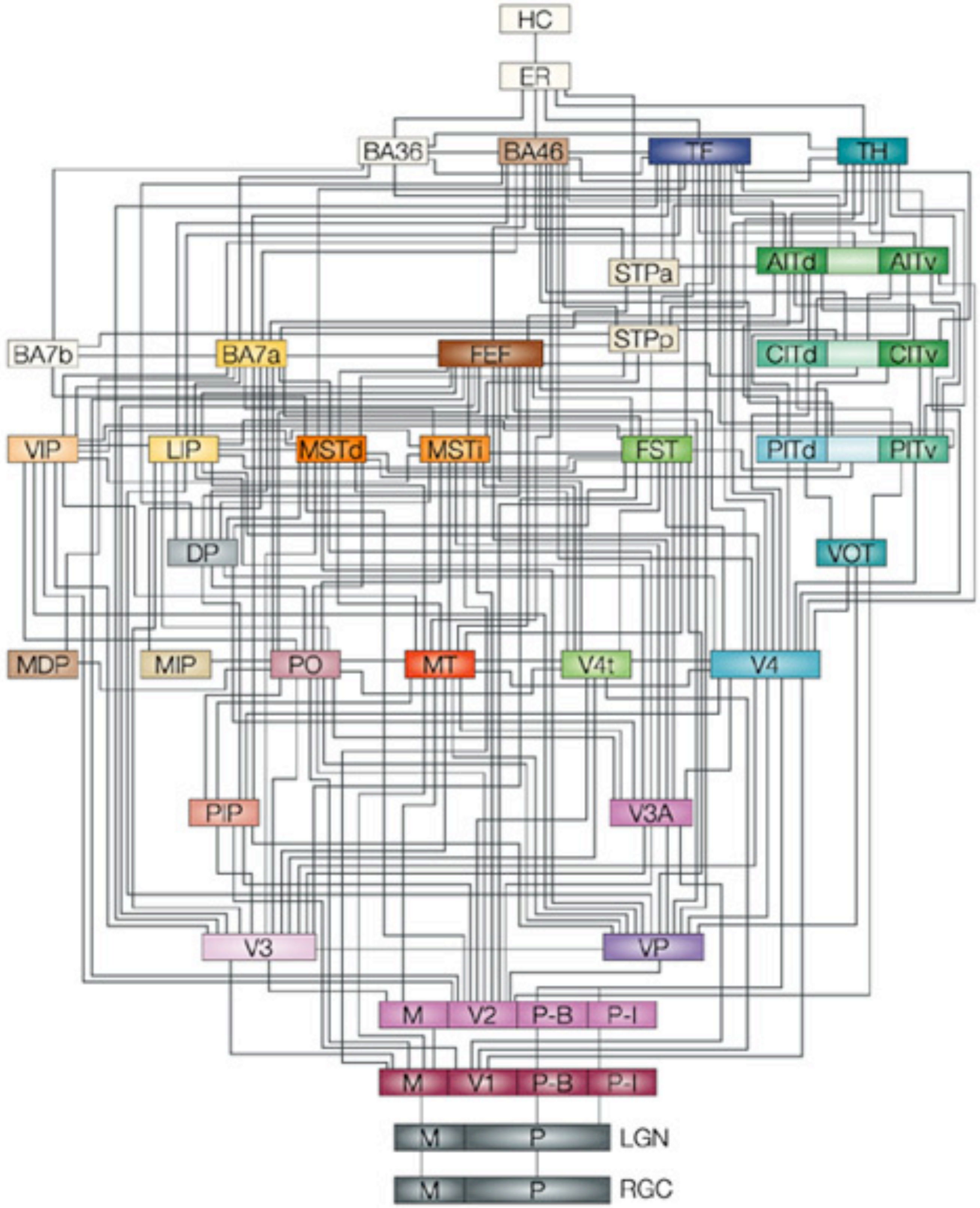
RF organization in V1

Hubel & Wiesel



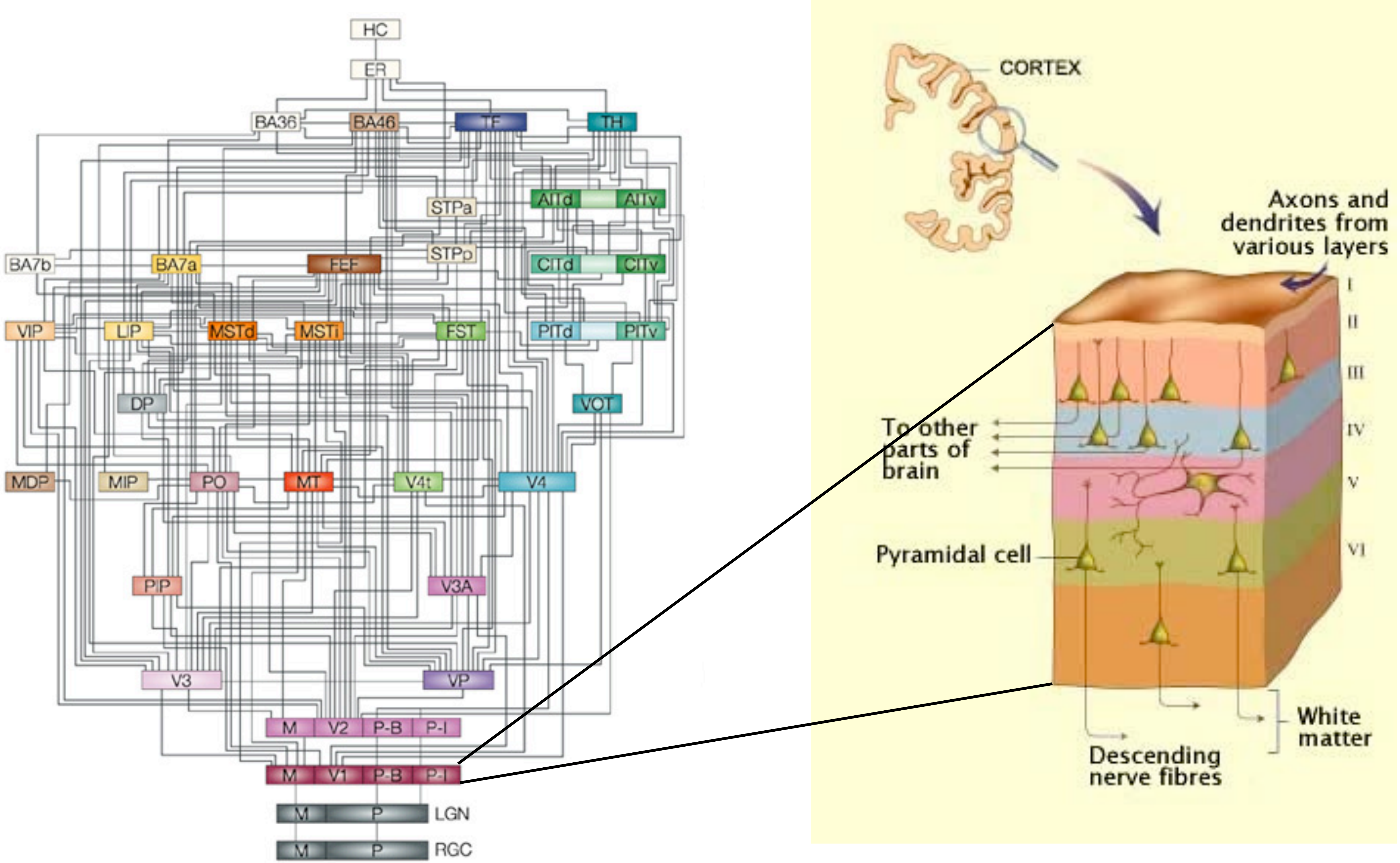


Hubel & Wiesel ’59 ’62 ’68



Hierarchical architecture: Anatomy

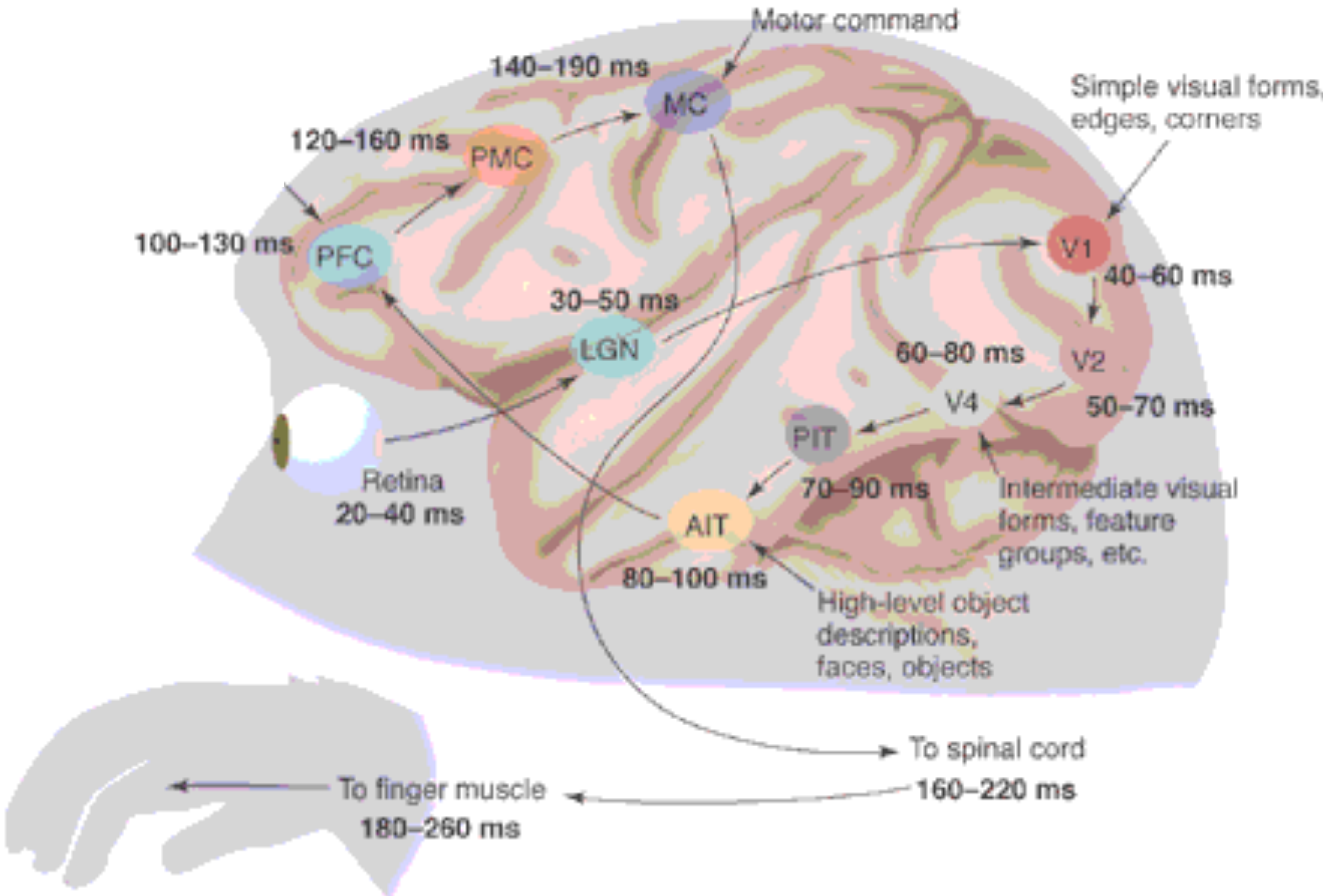
Rockland & Pandya '79;
Maunsell & Van Essen '83;
Felleman & Van Essen '91



Hierarchical architecture: Anatomy

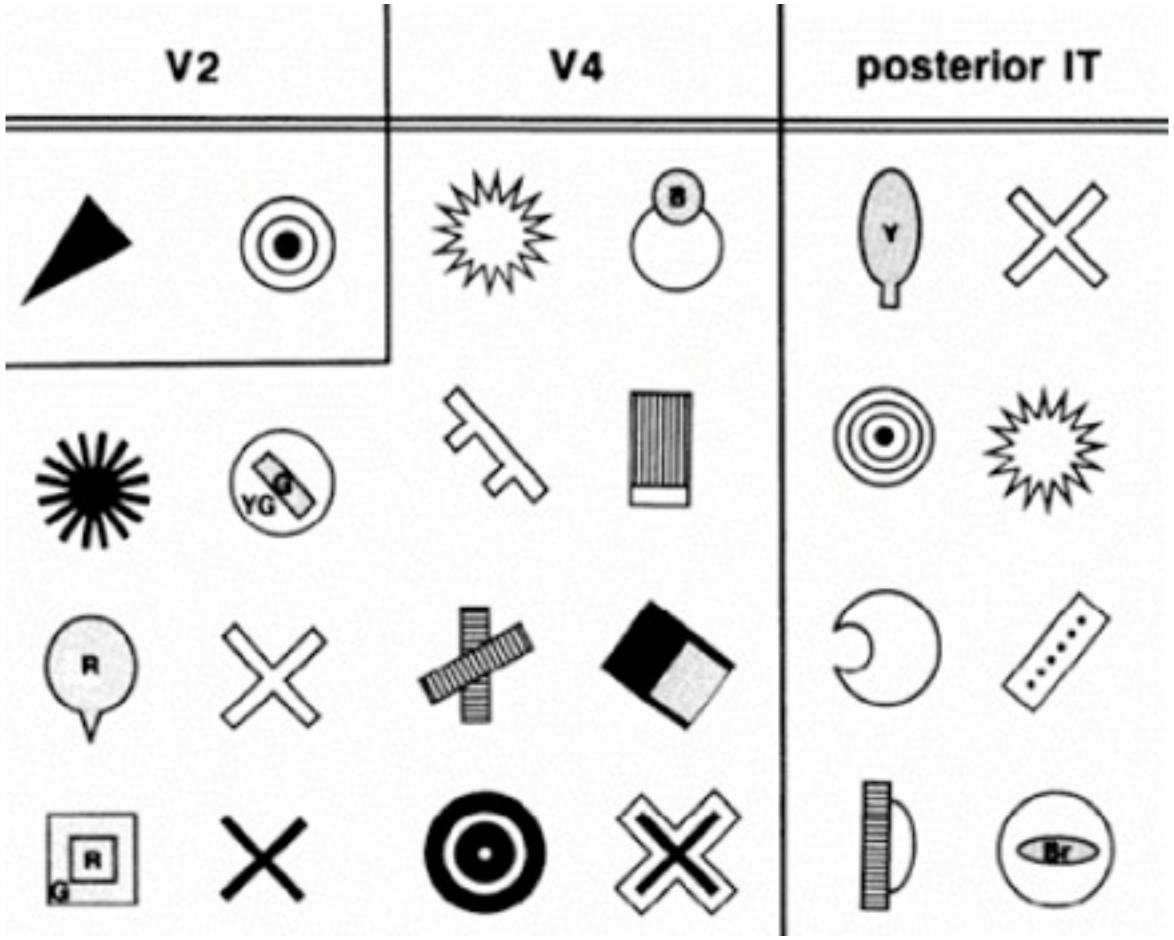
Rockland & Pandya ’79;
Maunsell & Van Essen ‘83;
Felleman & Van Essen ’91

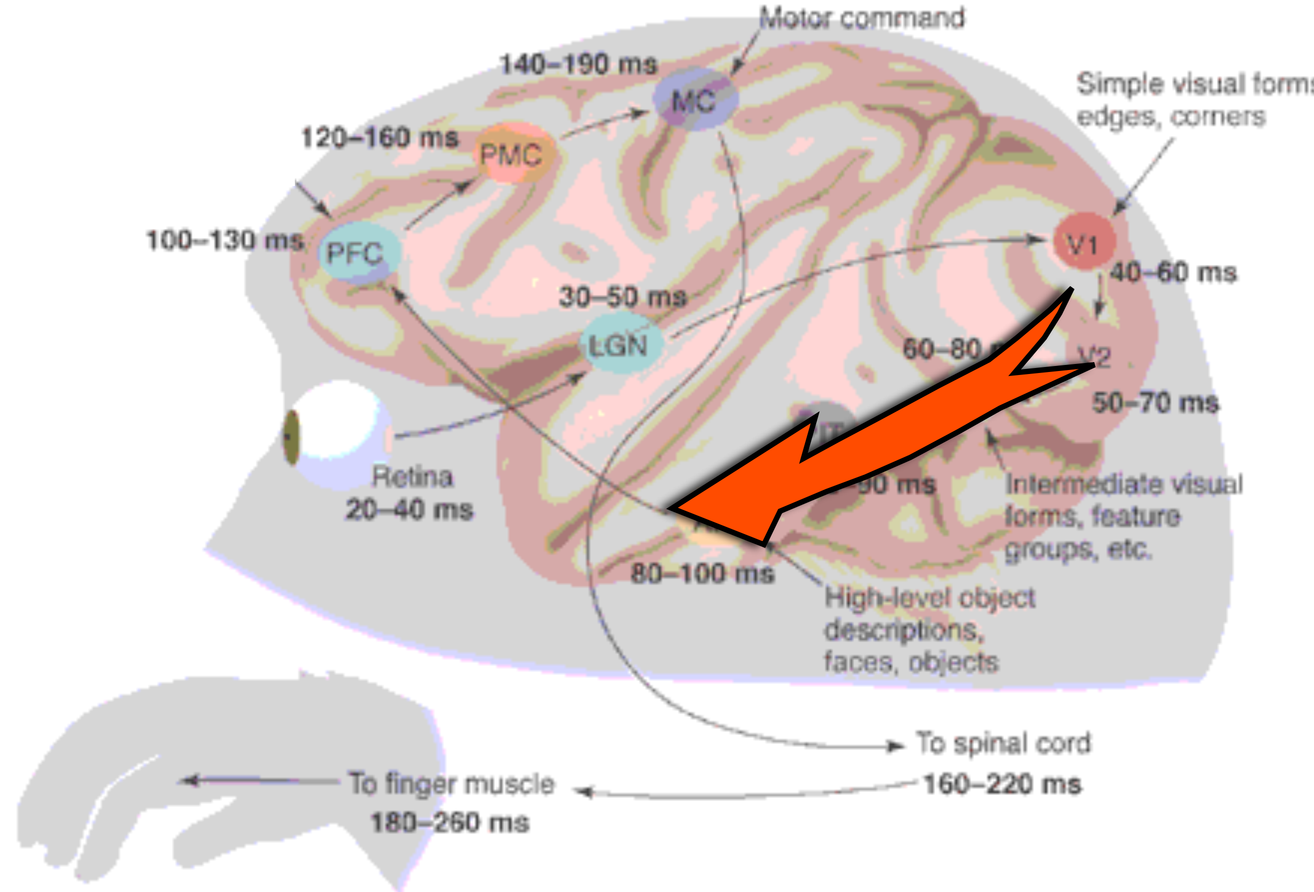
source: Thorpe & Fabre-Thorpe ‘01



Hierarchical architecture: Latencies

Nowak & Bullier ’97
Schmolesky et al ’98

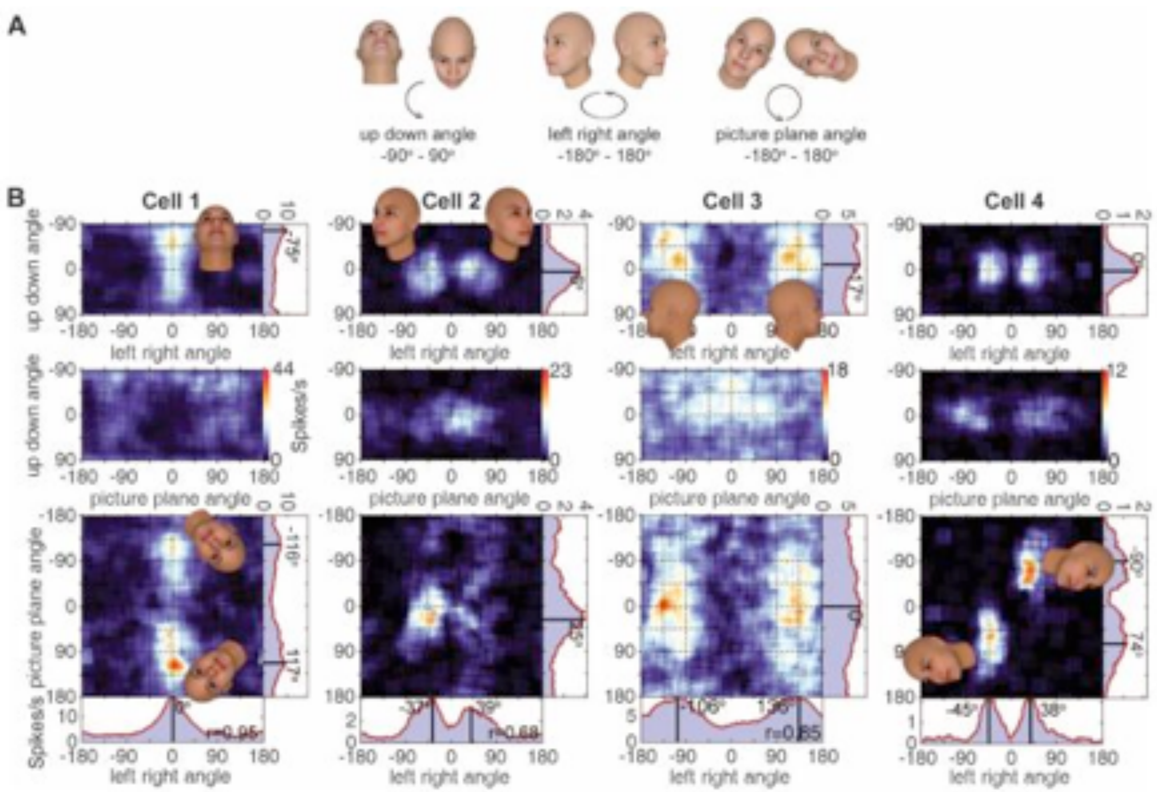


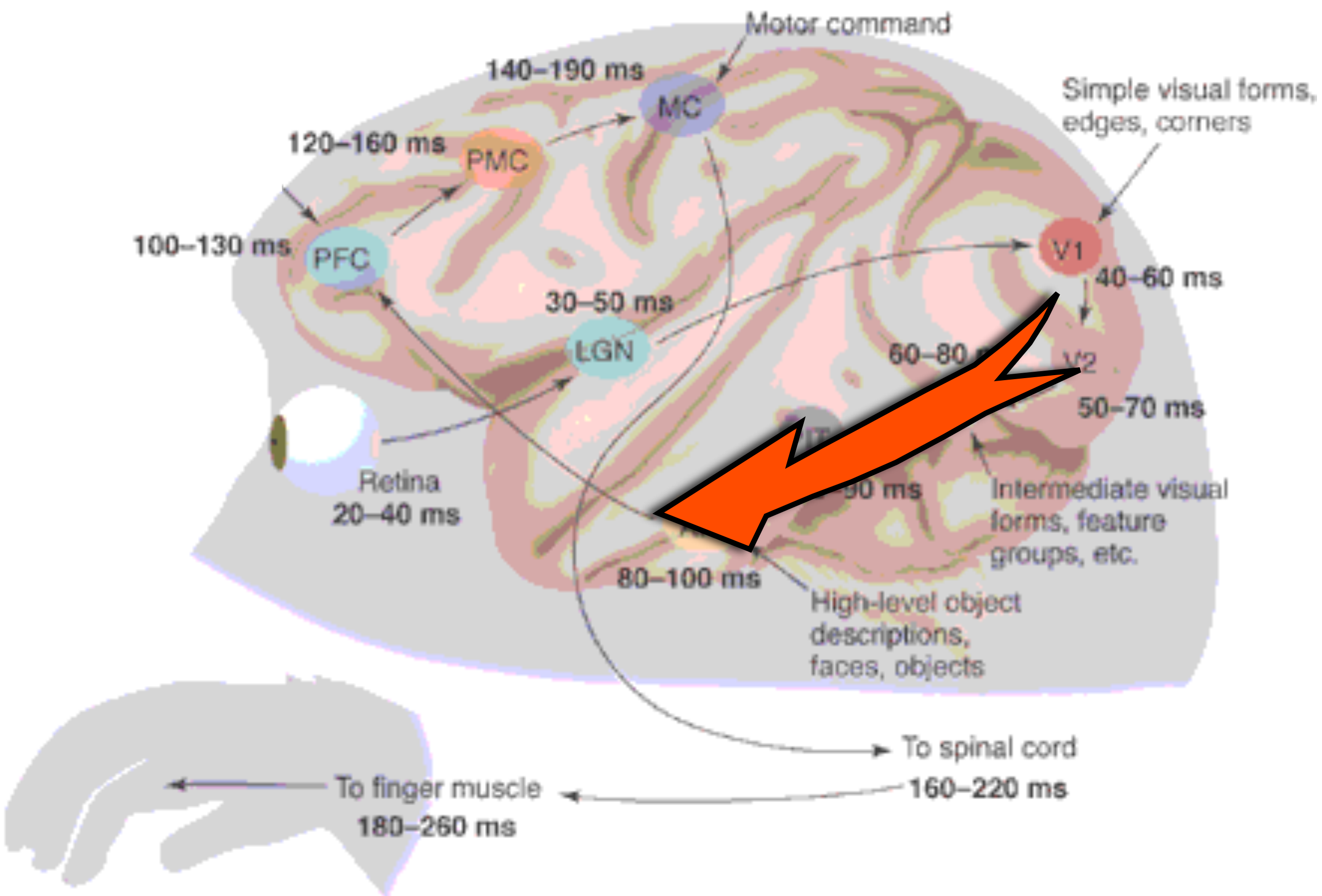


gradual increase in complexity of preferred stimulus

Hierarchical architecture: Function

source: Kobatake & Tanaka ’94; Freiwald & Tsao ’10
see also Oram & Perrett 1993; Sheinberg & Logothetis ‘96; Gallant et al ‘96; Riesenhuber & Poggio ’99

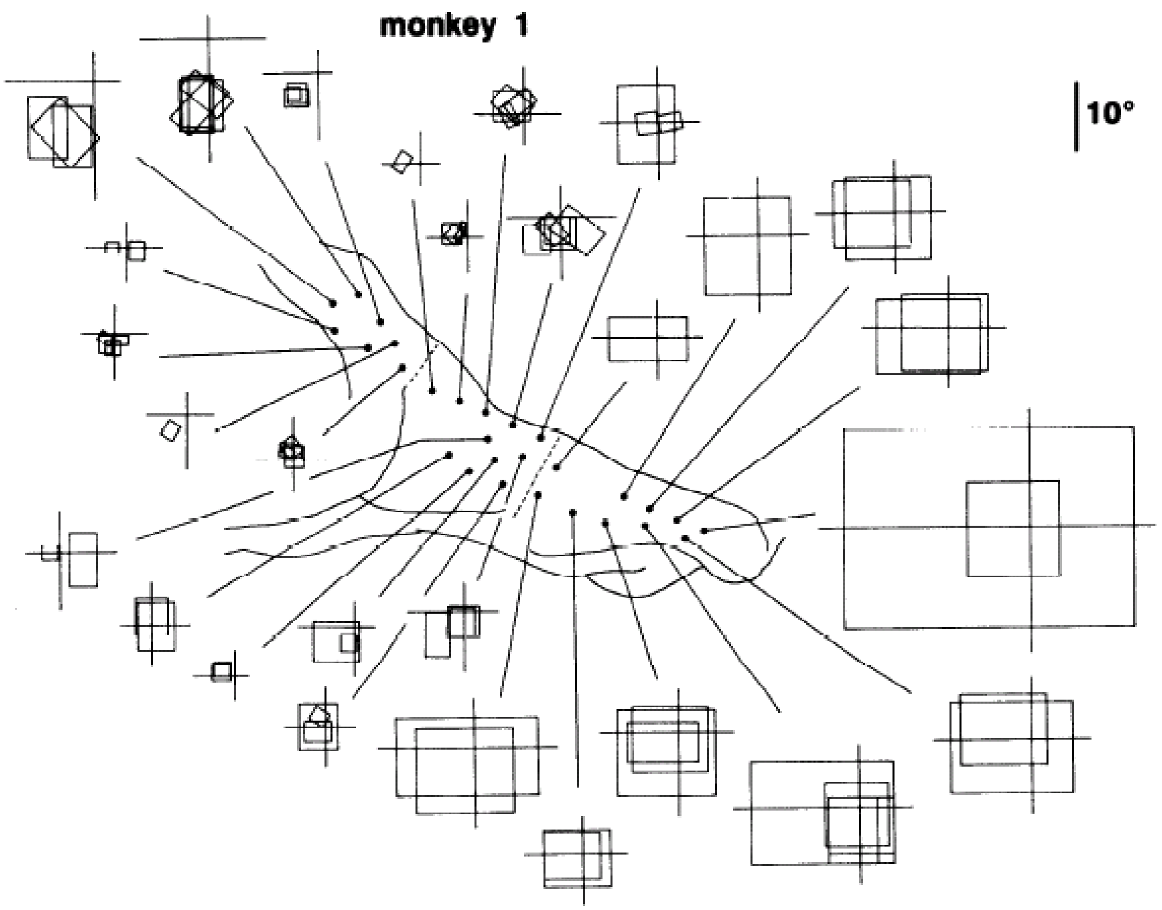


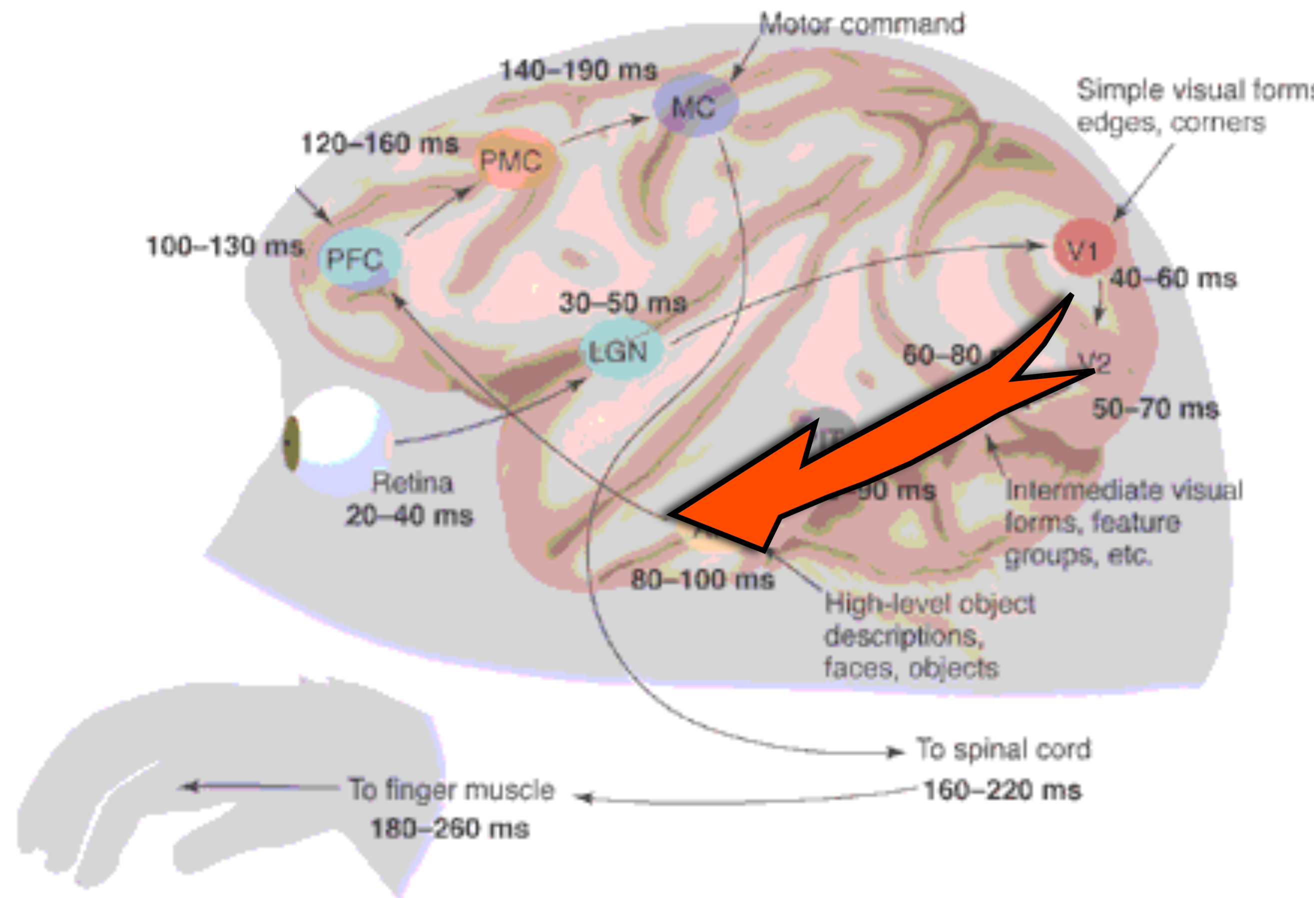


gradual increase in complexity of preferred stimulus

Hierarchical architecture: Function

source: Kobatake & Tanaka ’94; Freiwald & Tsao ’10 see also Oram & Perrett 1993; Sheinberg & Logothetis ‘96; Gallant et al ‘96; Riesenhuber & Poggio ’99





Parallel increase in invariance properties (position and scale) of neurons

Hierarchical architecture: Function

source: Kobatake & Tanaka 1994
see also Oram & Perrett 1993; Sheinberg & Logothetis 1996; Gallant et al 1996; Riesenhuber & Poggio 1999

Columnar organization in V1

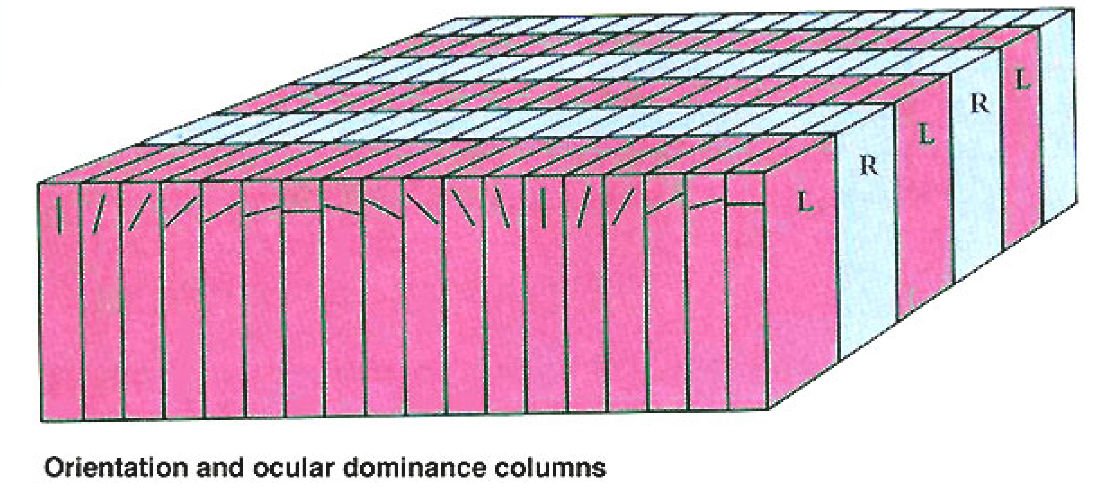
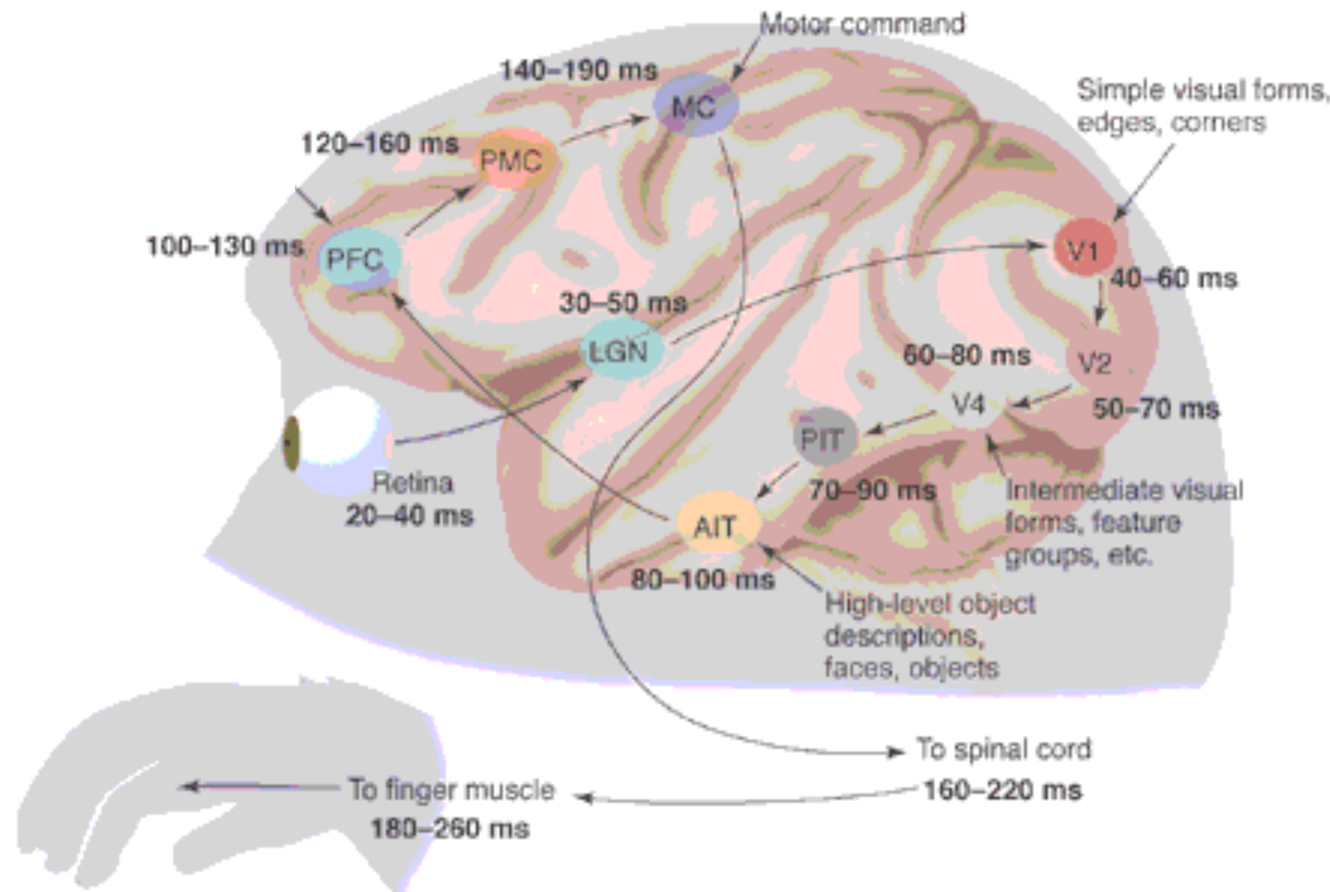
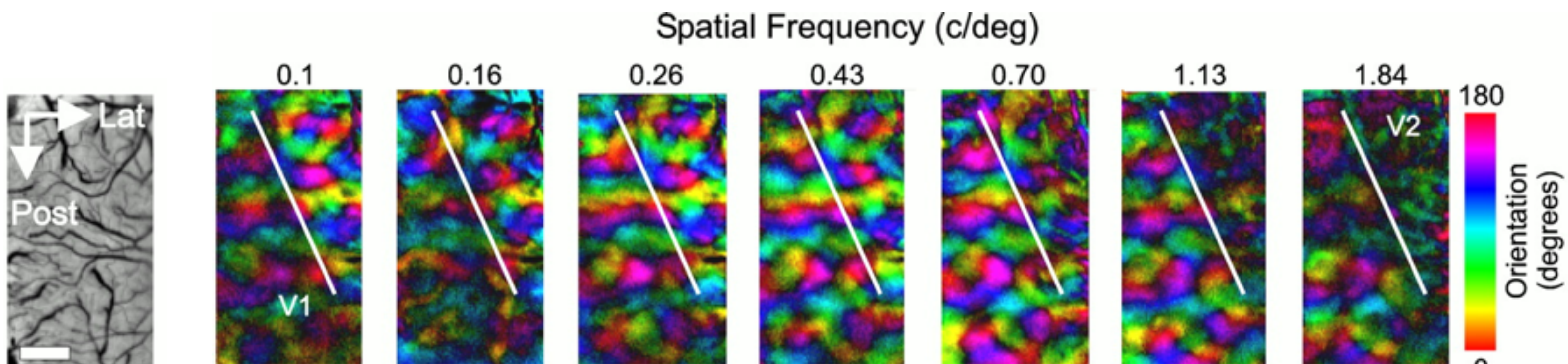


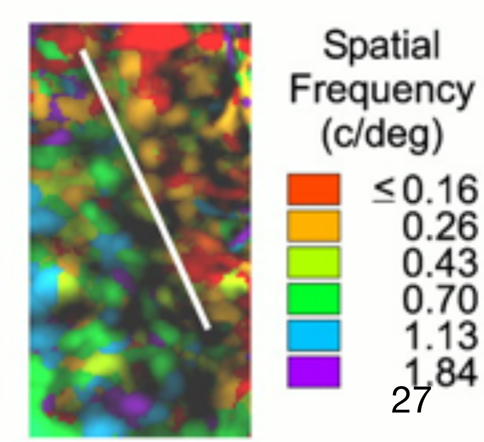
Figure 23. The ice-cube model of the cortex. It illustrates how the cortex is divided, at the same time, into two kinds of slabs, one set of ocular dominance (left and right) and one set for orientation. The model should not be taken literally: Neither set is as regular as this, and the orientation slabs especially are far from parallel or straight.





Source: unknown

Columnar organization



Columnar organization in V1

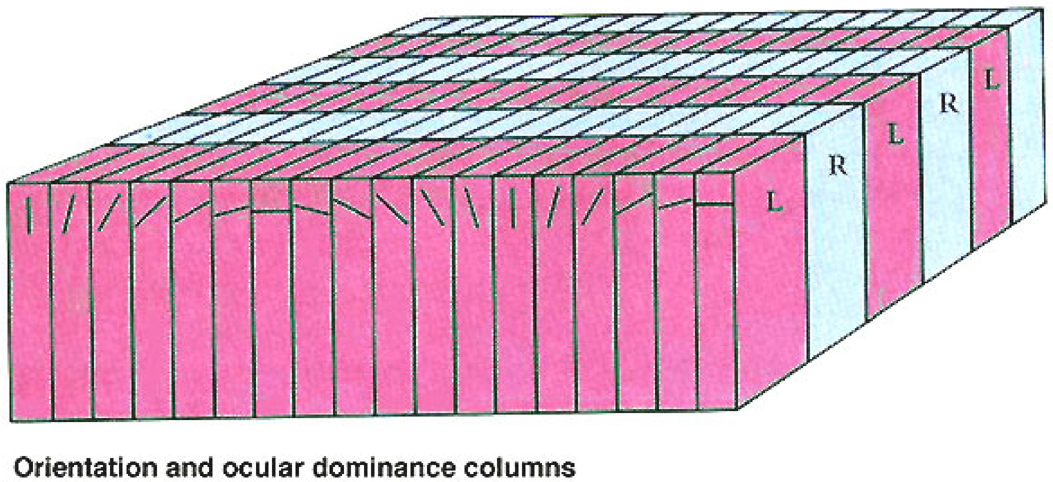
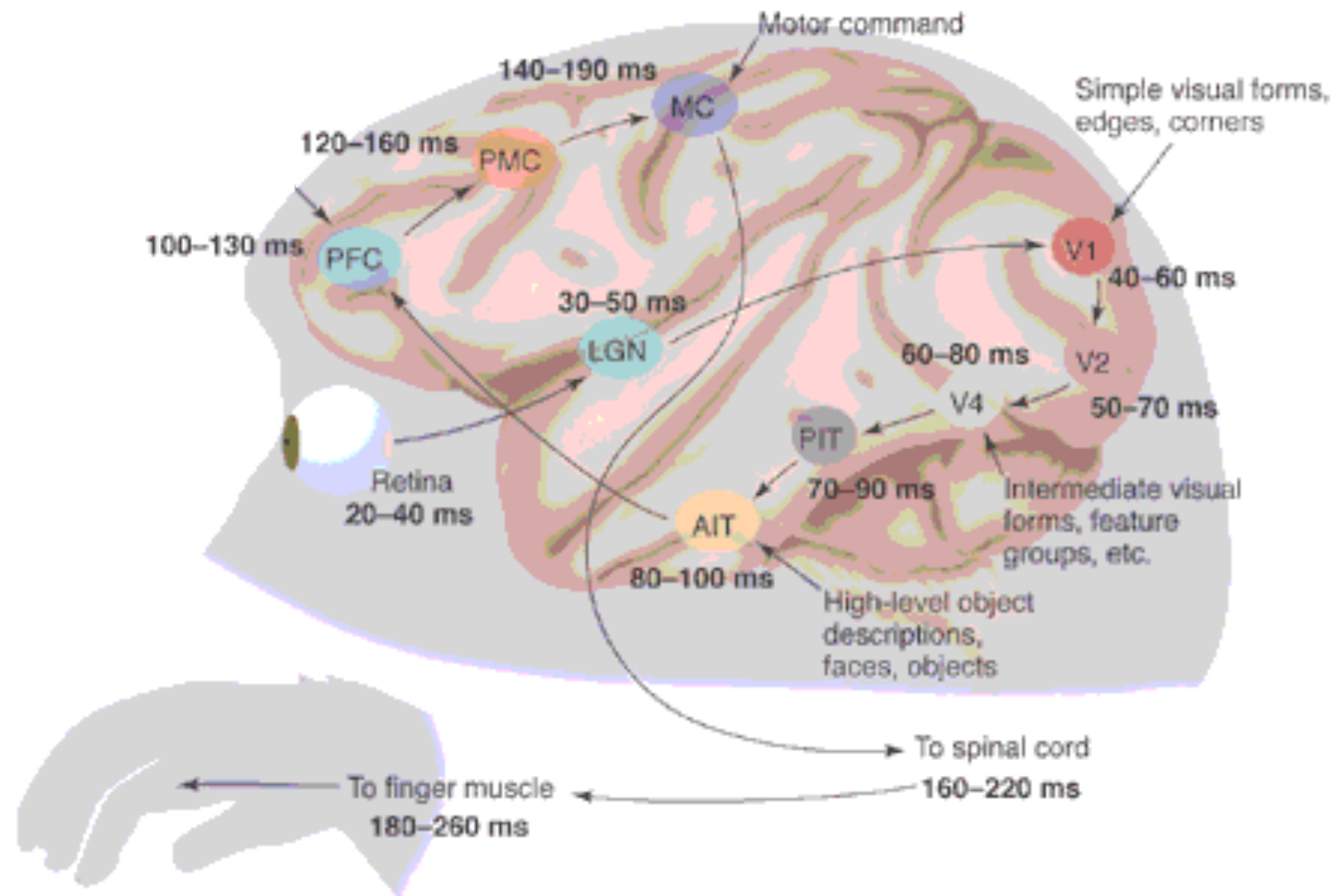
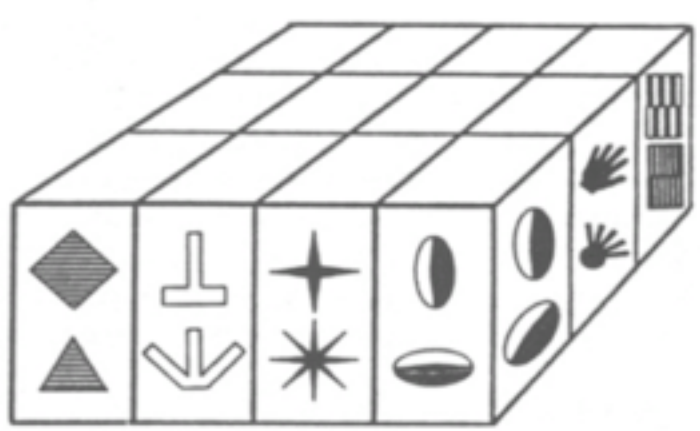
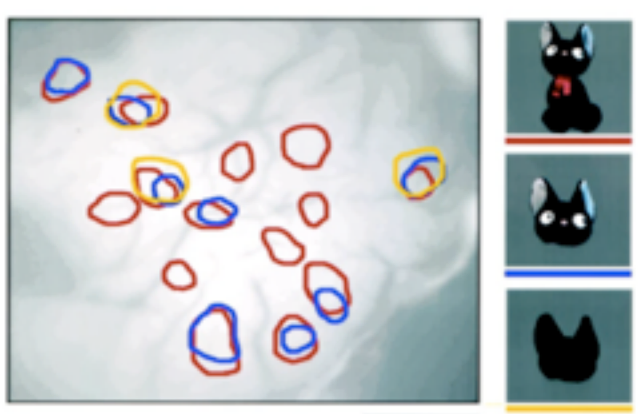


Figure 23. The ice-cube model of the cortex. It illustrates how the cortex is divided, at the same time, into two kinds of slabs, one set of ocular dominance (left and right) and one set for orientation. The model should not be taken literally: Neither set is as regular as this, and the orientation slabs especially are far from parallel or straight.





Tanaka et al.

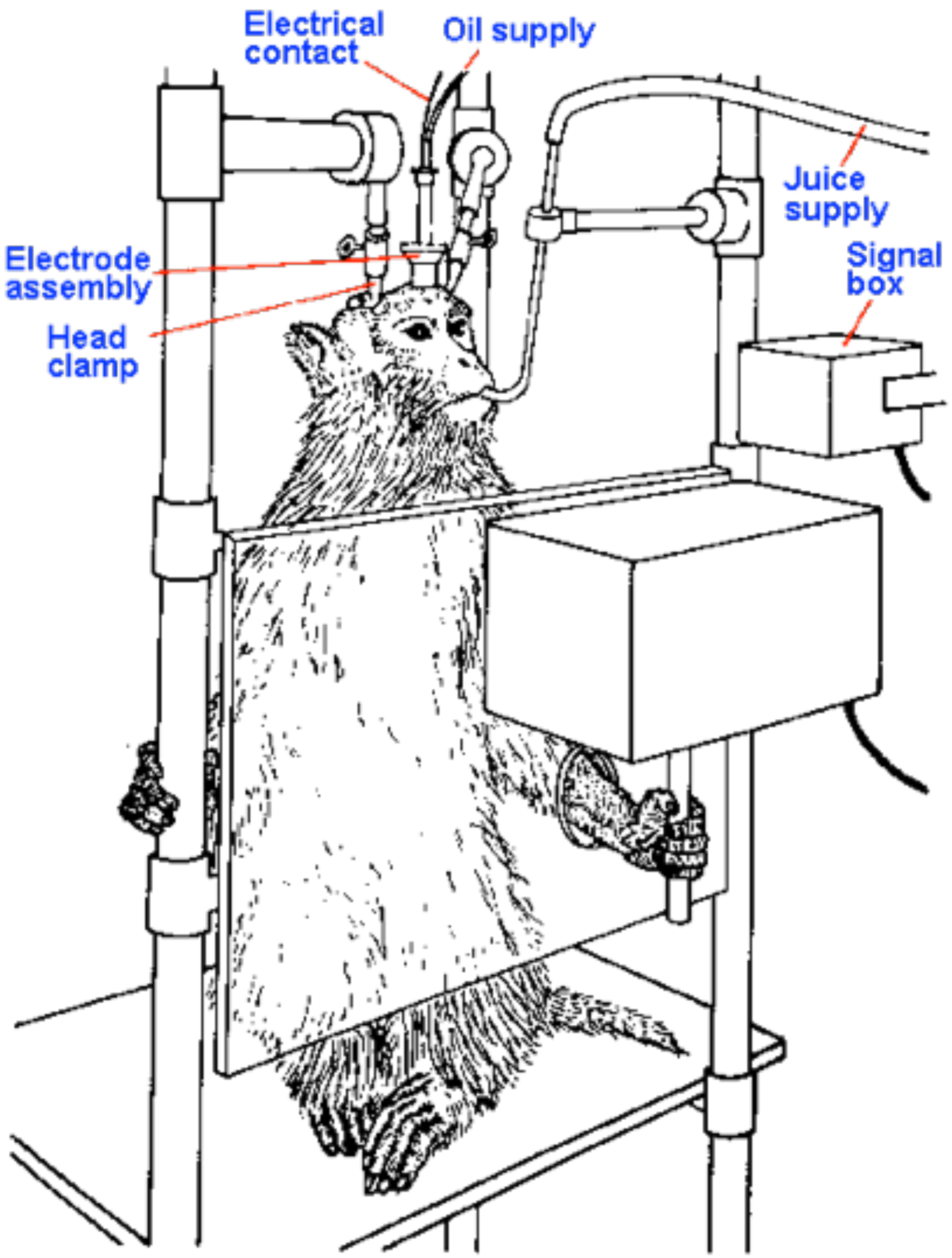


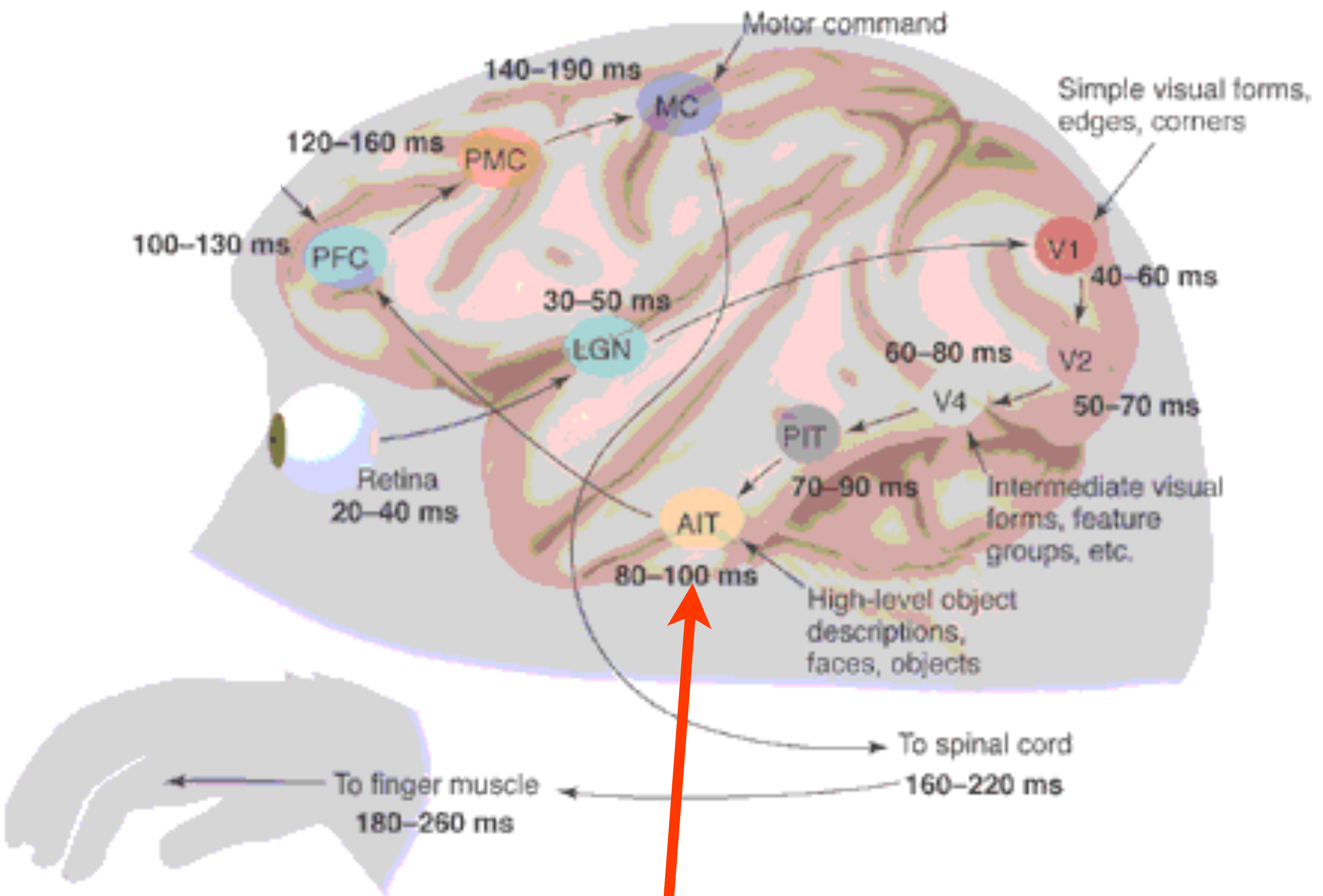
Tsunoda et al.

Dictionary of visual features of intermediate visual areas (V4/PIT)

Columnar organization

Kobatake & Tanaka 1994
see also Oram & Perrett 1993; Sheinberg & Logothetis 1996; Gallant et al 1996; Riesenhuber & Poggio 1999



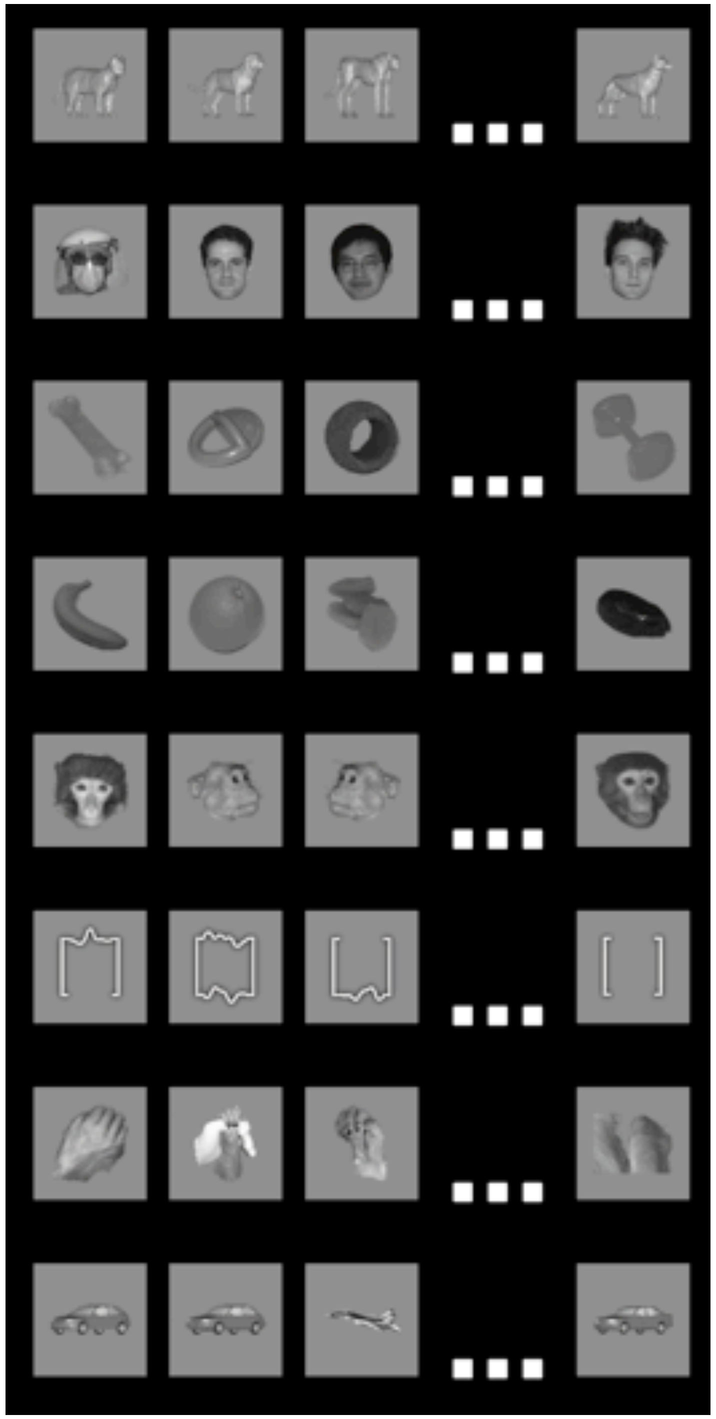


Invariant object category information can be decoded from small populations of cells in IT

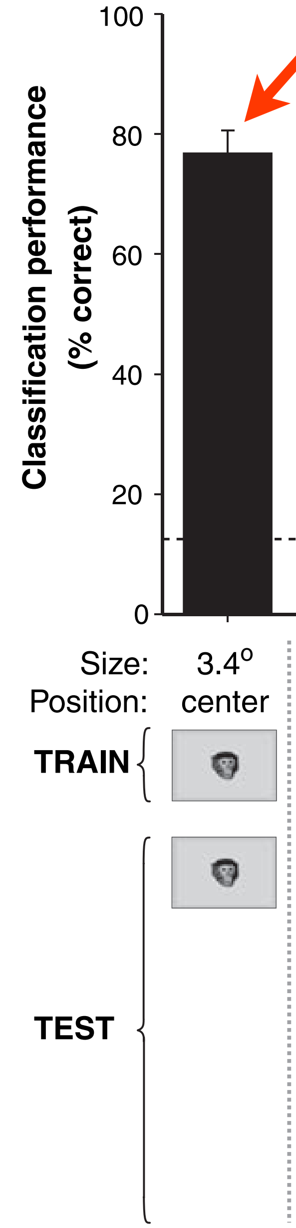
Invariant image representation in the IT

77 unique stimuli

8 object category



Decoding possible from around 100 ms

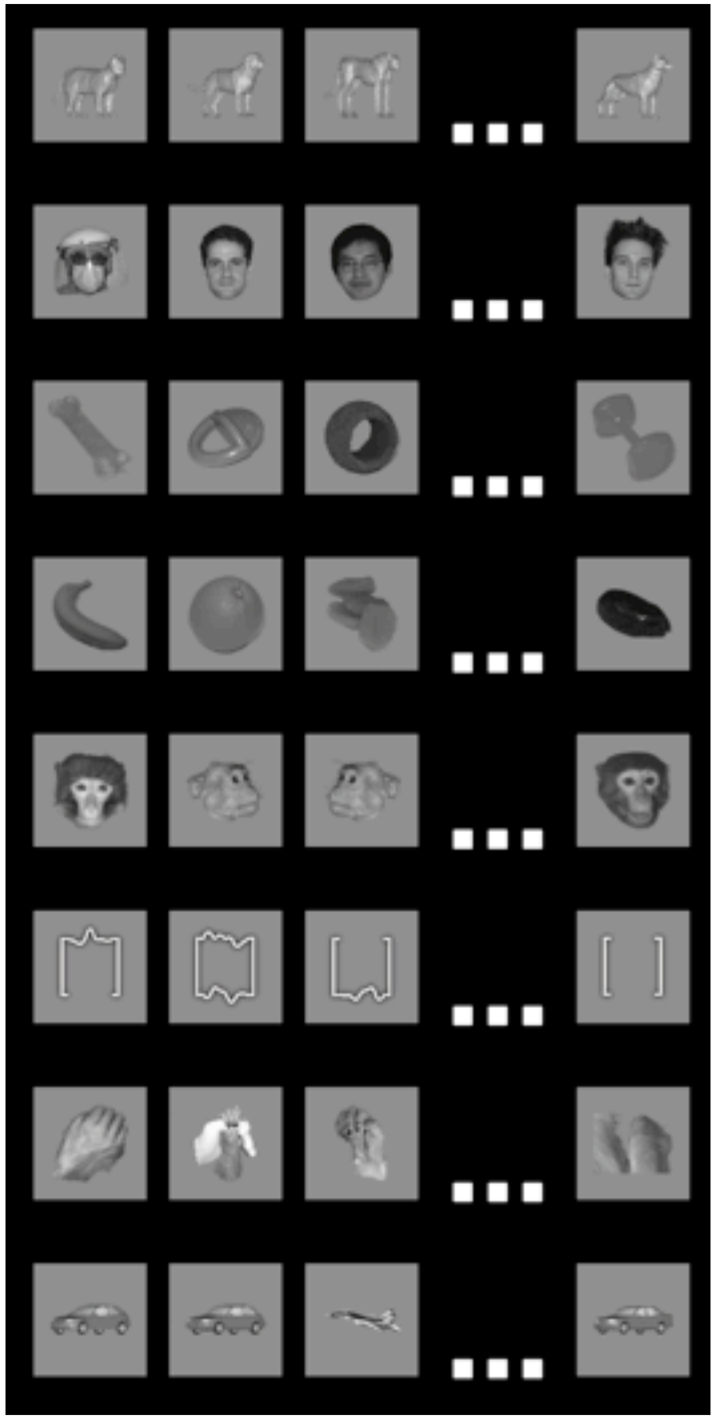


Invariant representation in IT

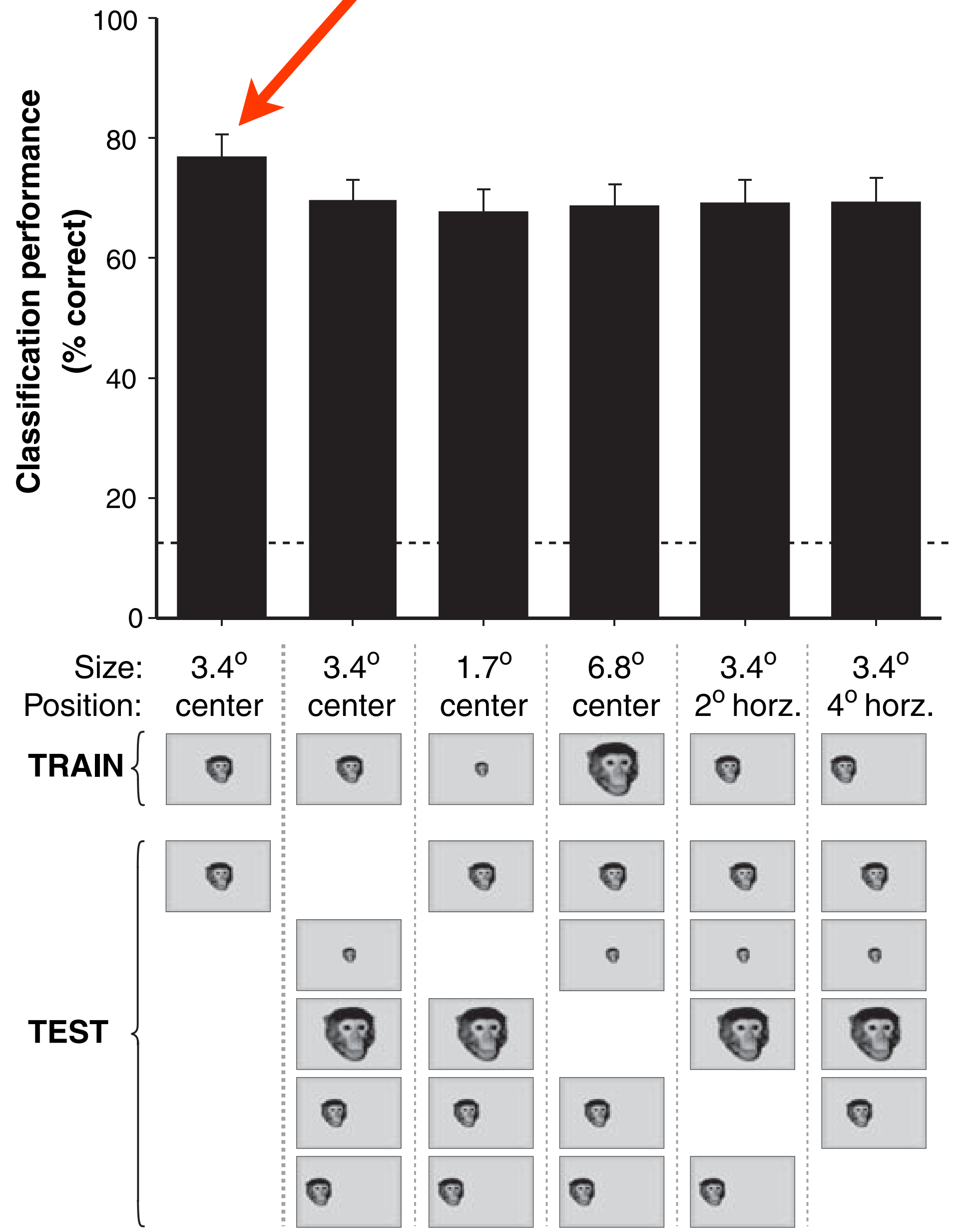
Hung et al ’05

77 unique stimuli

8 object category



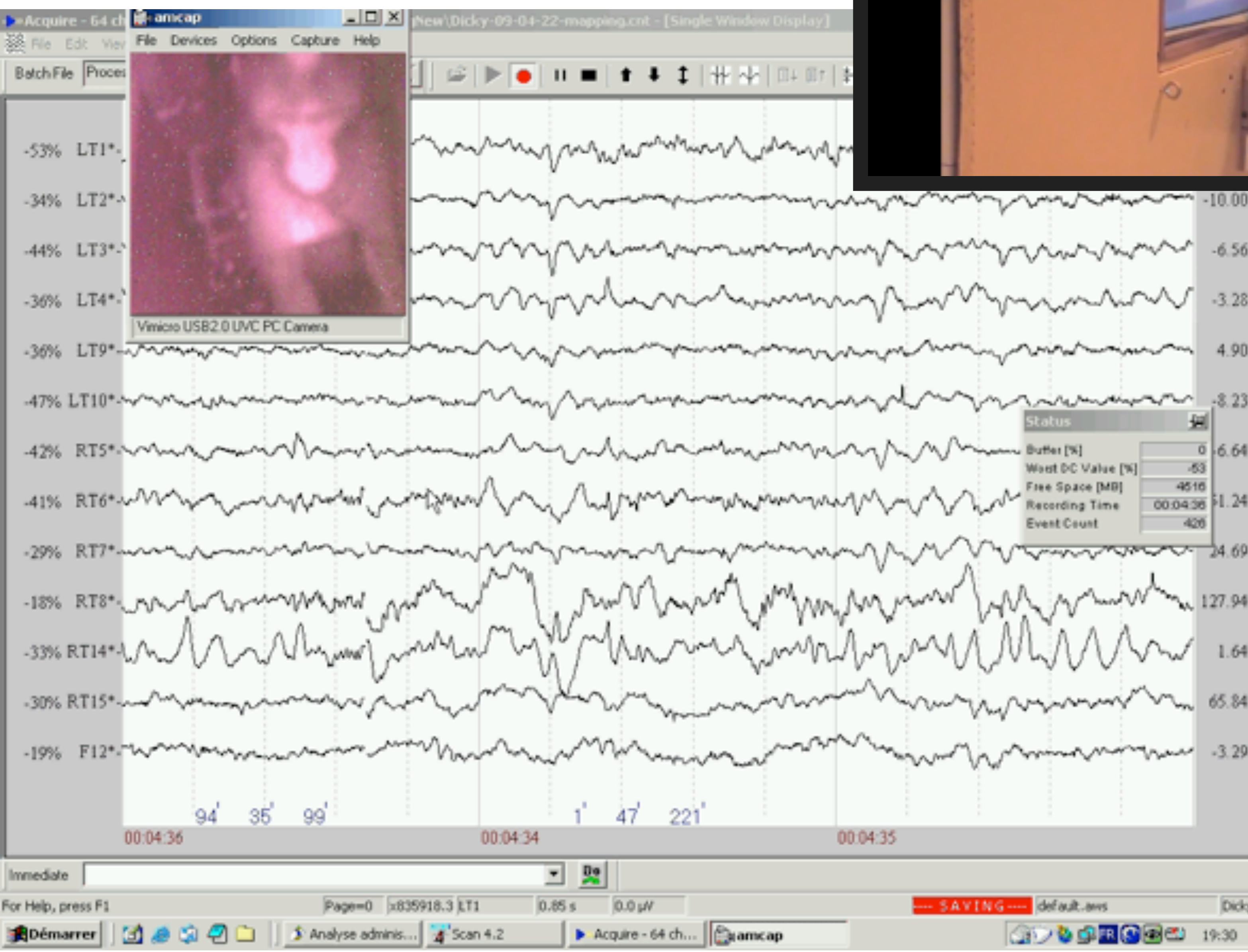
Decoding possible from around 100 ms



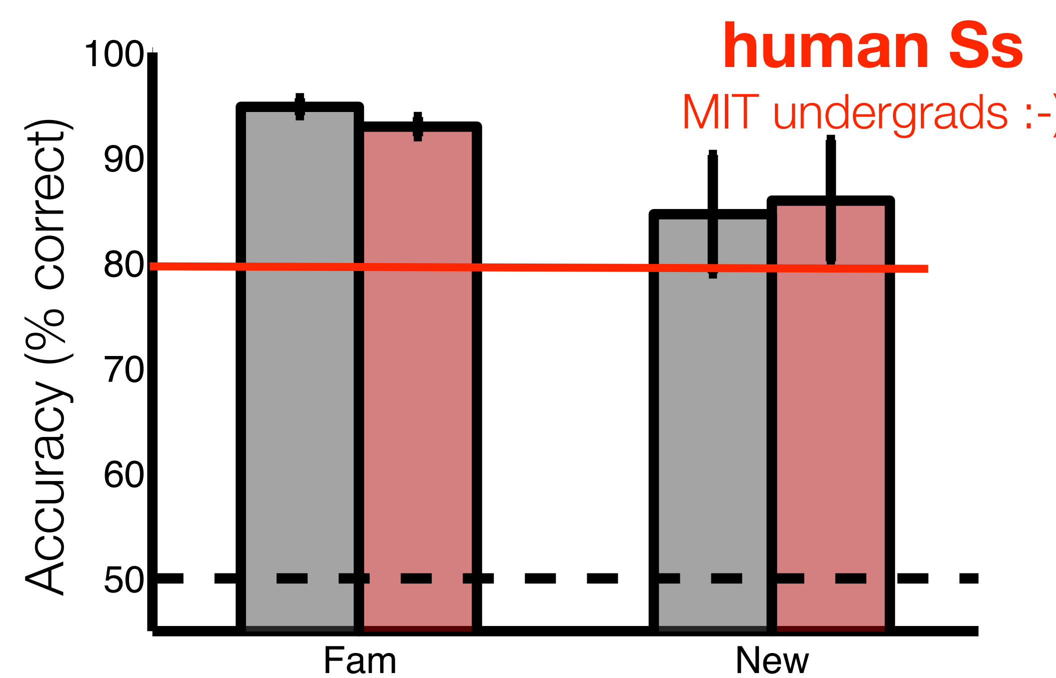
Invariant representation in IT

Hung et al ’05

Button release and touch screen on targets

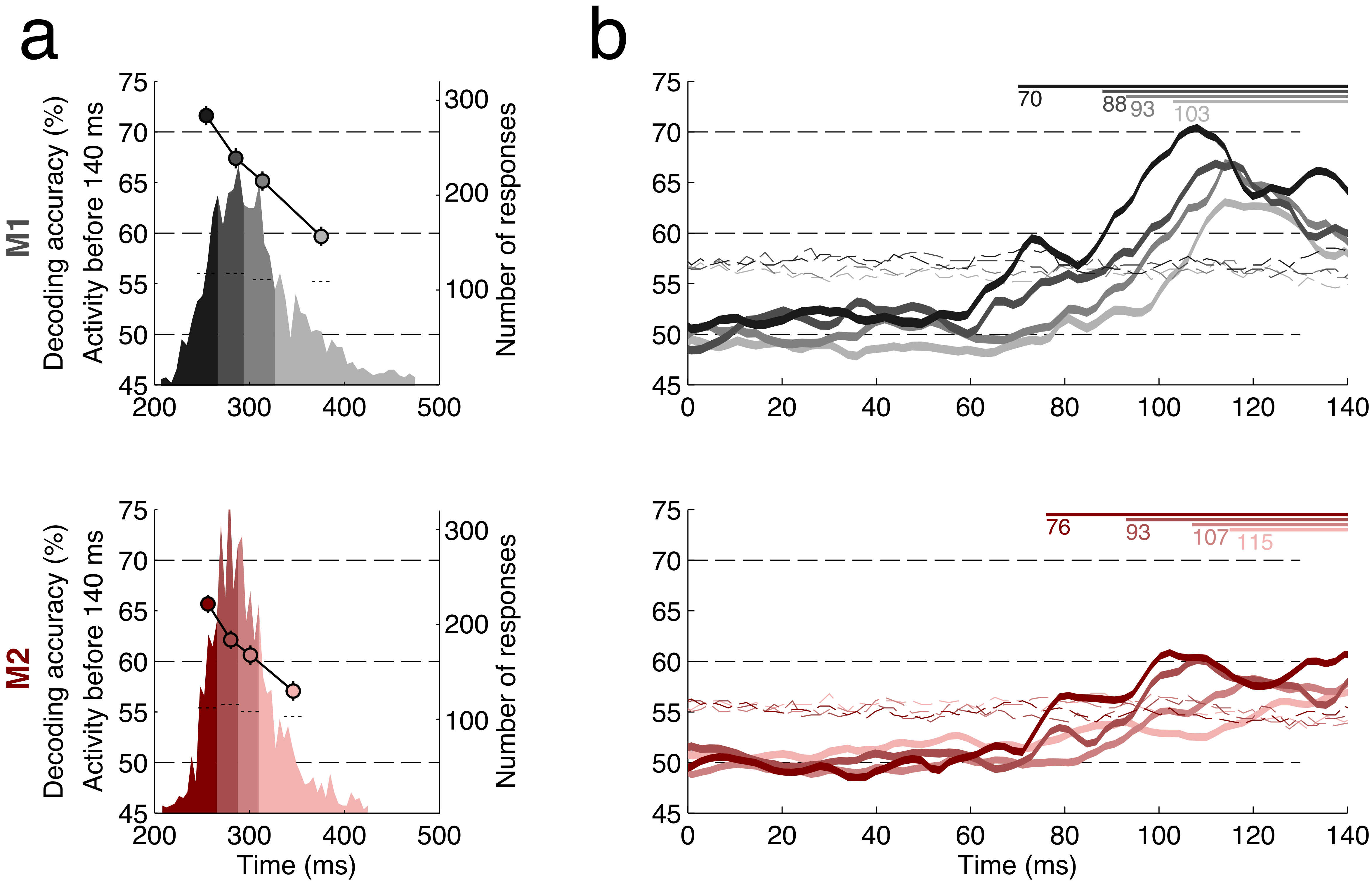


- Head-free monkeys
- Multiple electrodes implanted along ventral stream (V4+PIT)



Feedforward processing

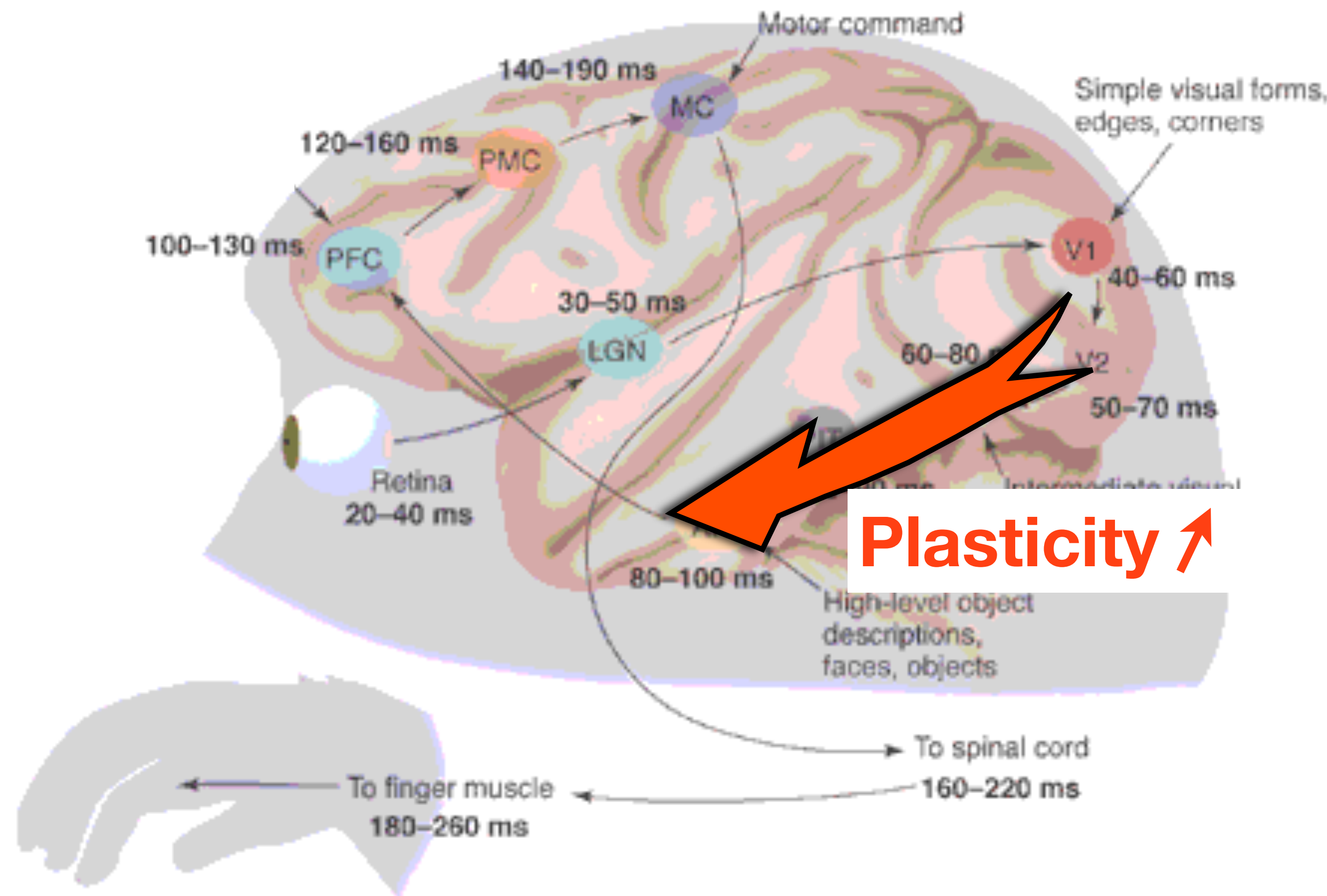
Data collected by Maxime Cauchoix and Denis Fize (CNRS, France)

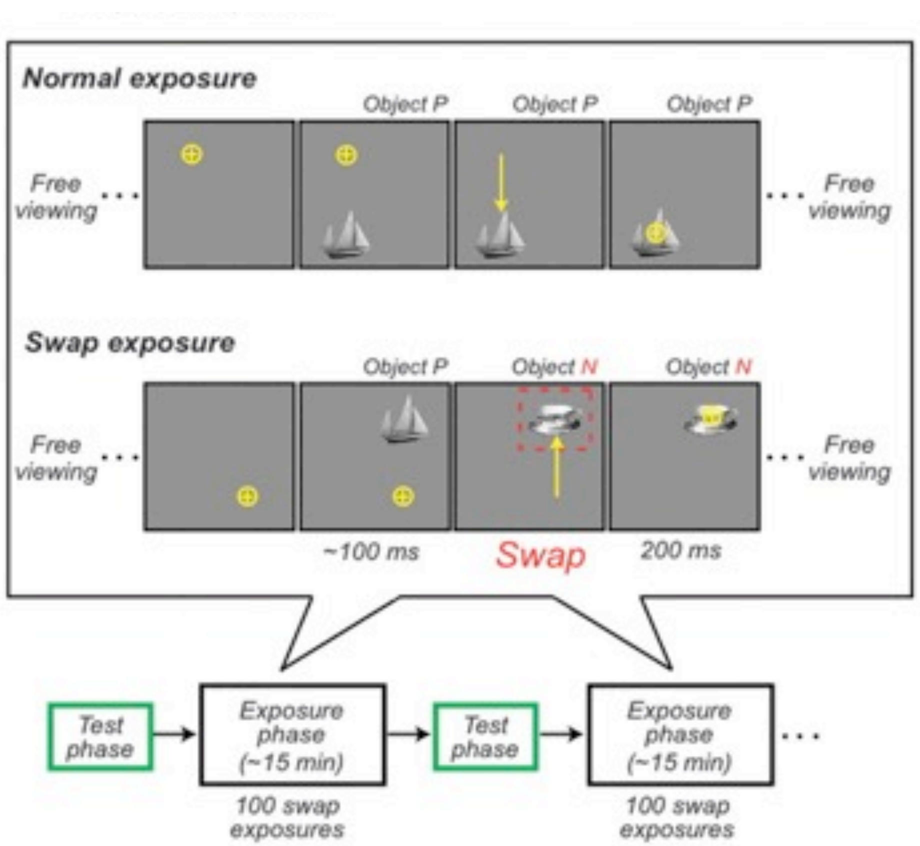


Feedforward processing

Data collected by Maxime Cauchoix and Denis Fize (CNRS, France)

- We know very little (i.e., STDP) about the learning rules at work in the visual cortex



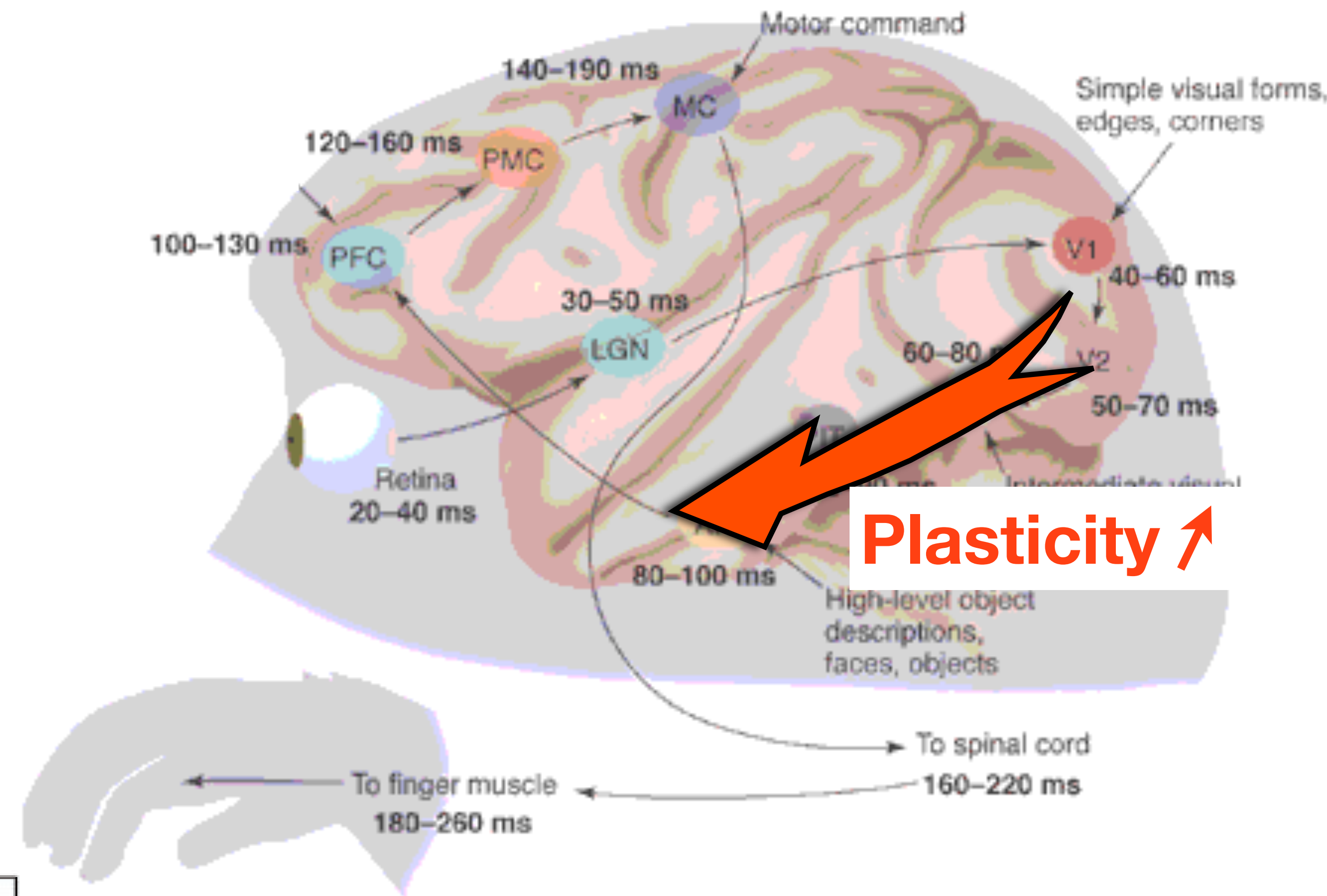


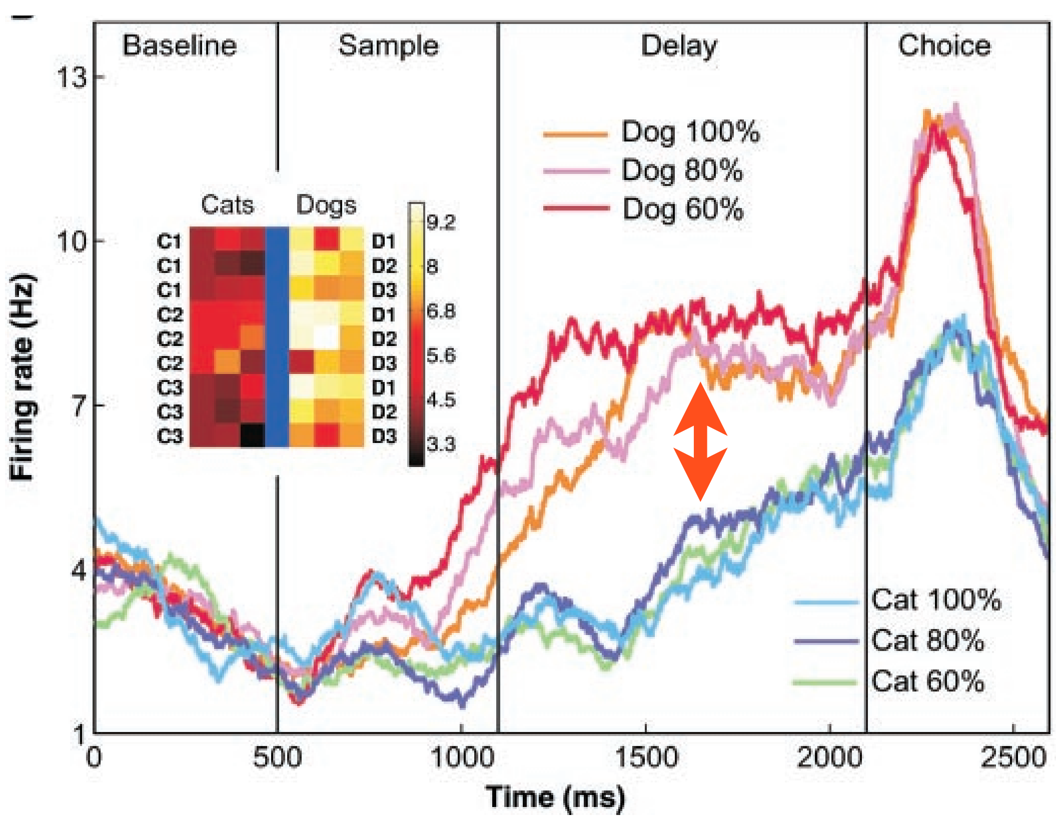
Very fast learning in IT

Learning and plasticity

Li & DiCarlo '08

- We know very little (i.e., STDP) about the learning rules at work in the visual cortex



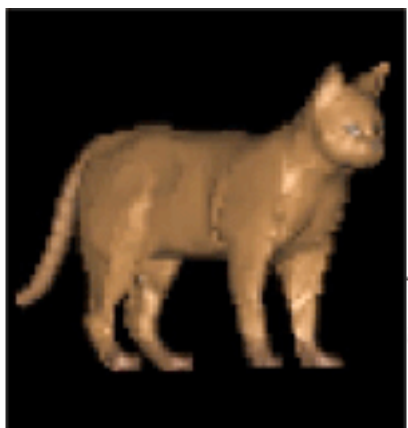


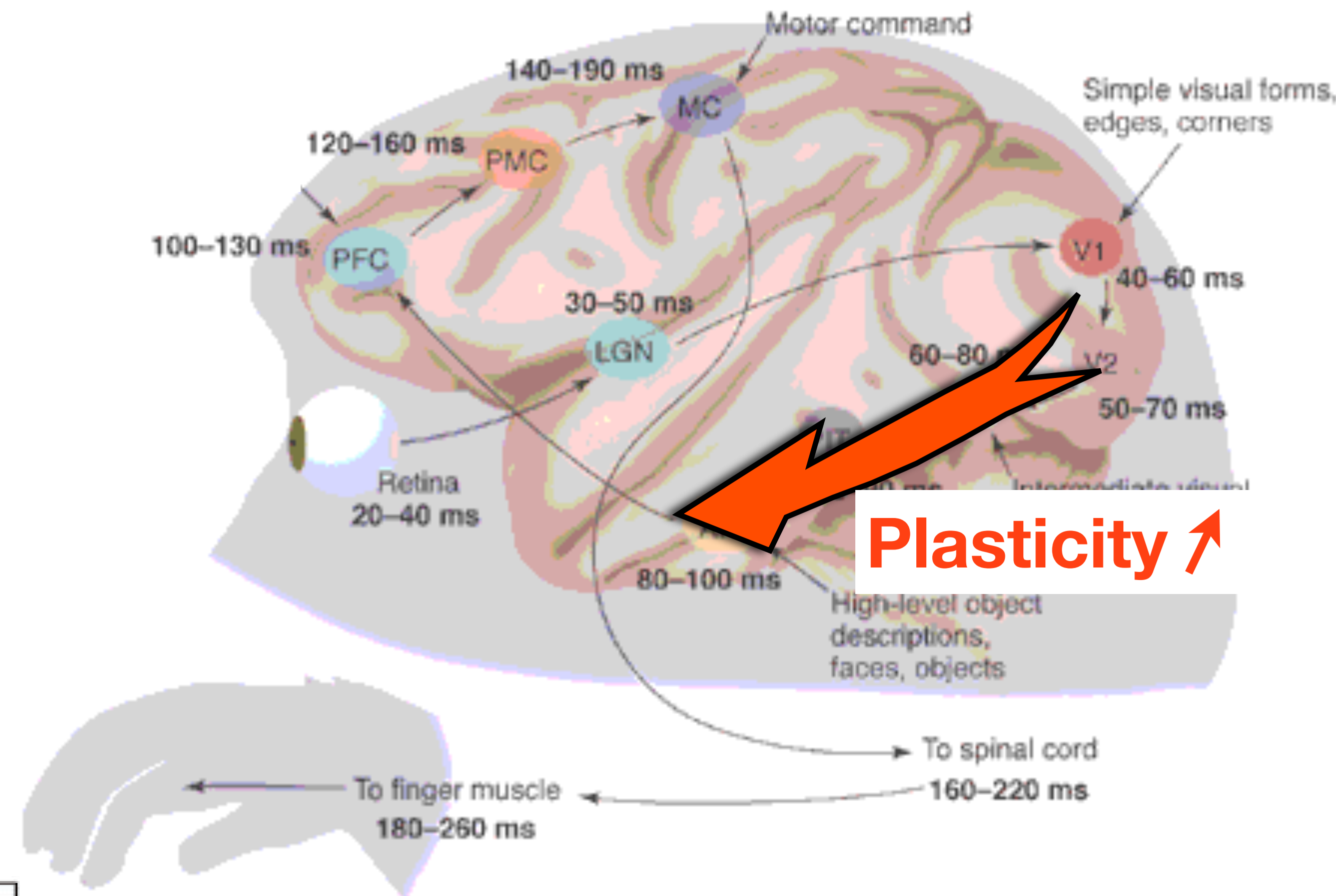
Supervised category learning in PFC

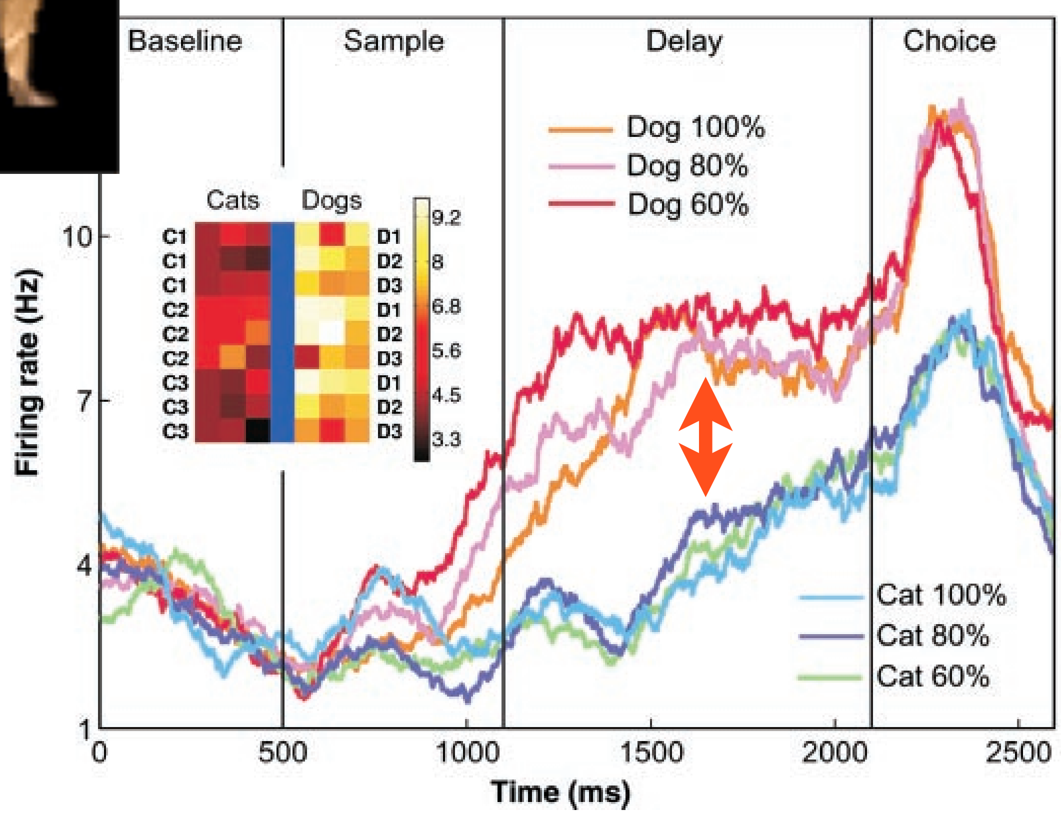
Learning and plasticity

Li & DiCarlo '08

- We know very little (i.e., STDP) about the learning rules at work in the visual cortex



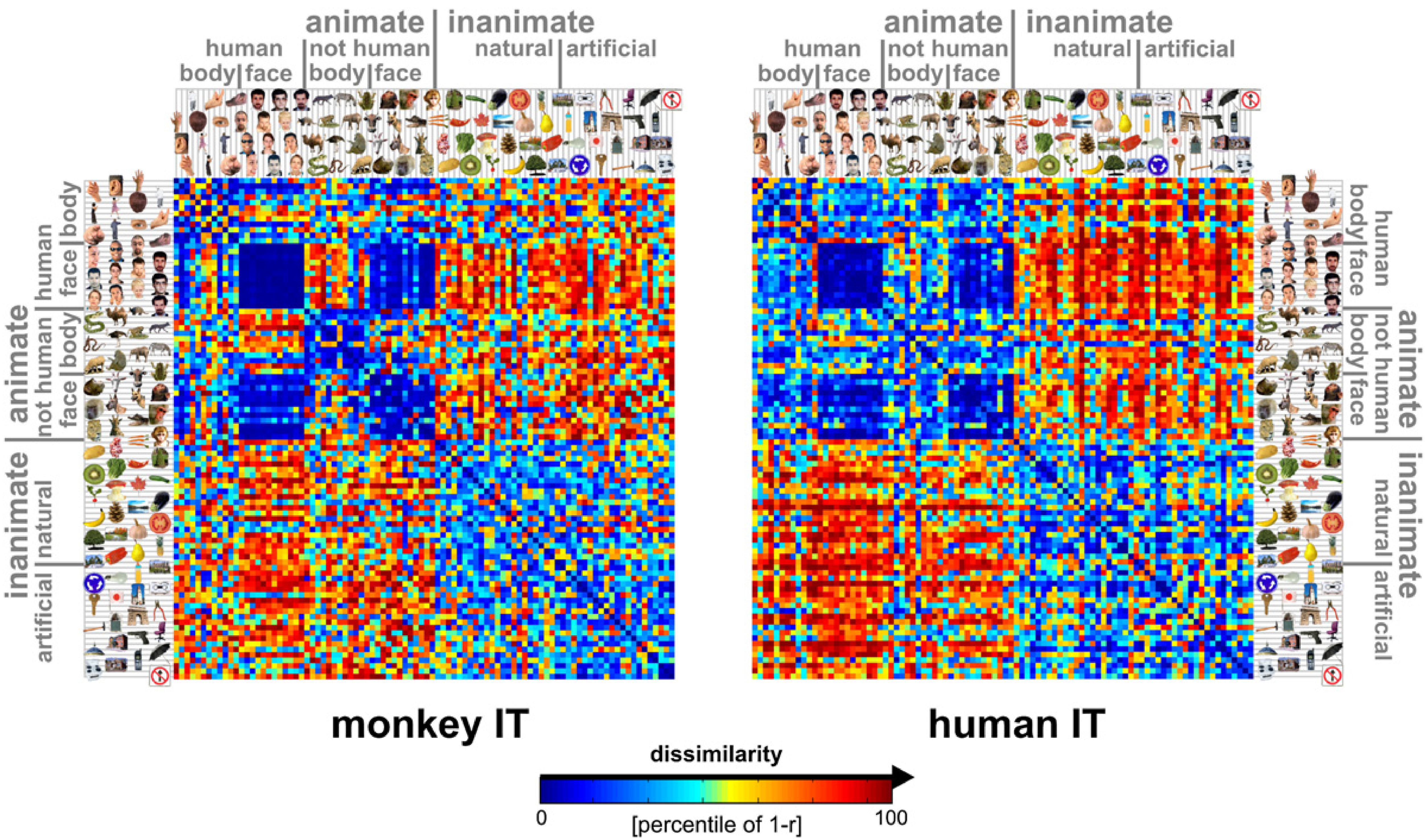




Supervised category learning in PFC

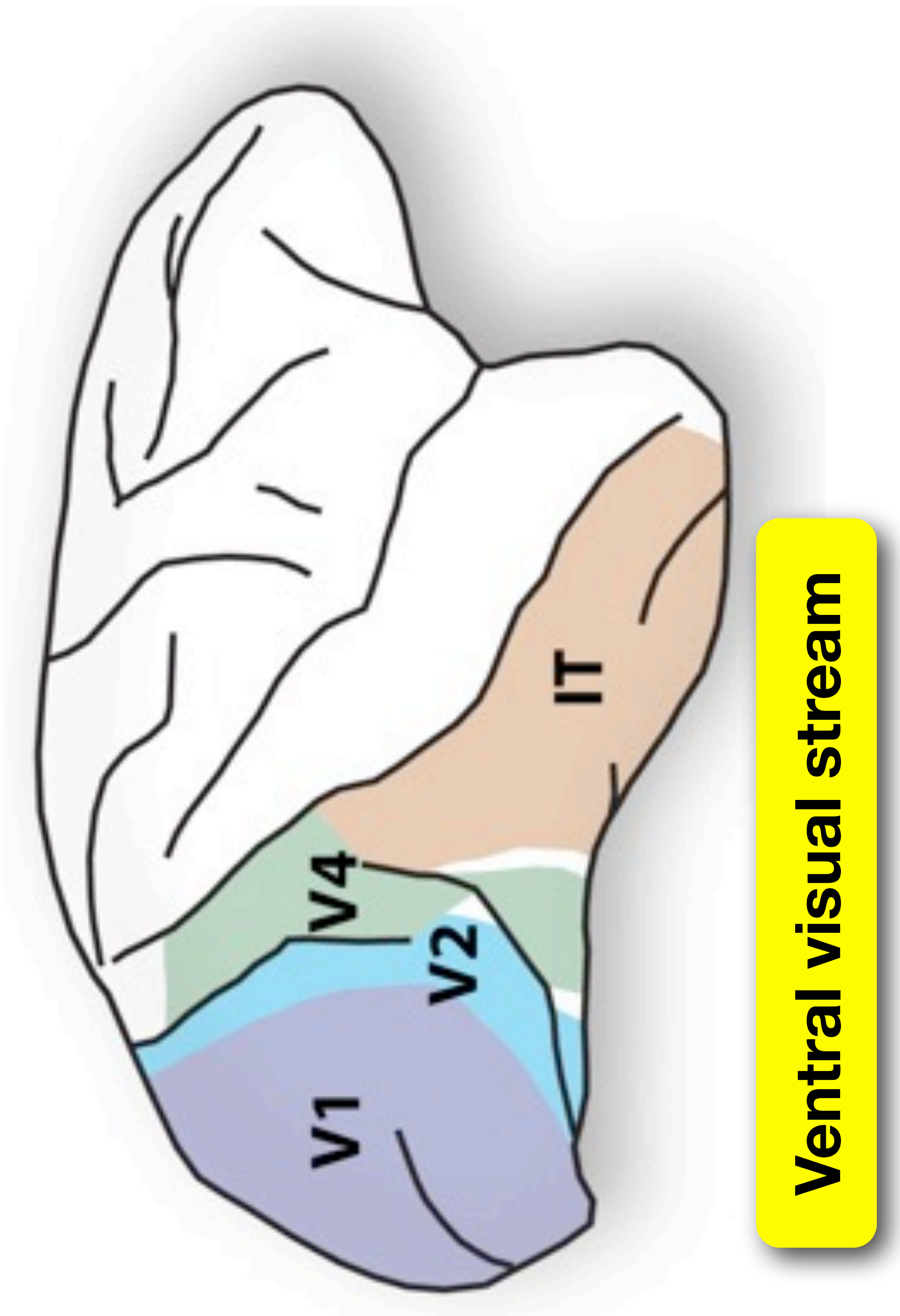
Learning and plasticity

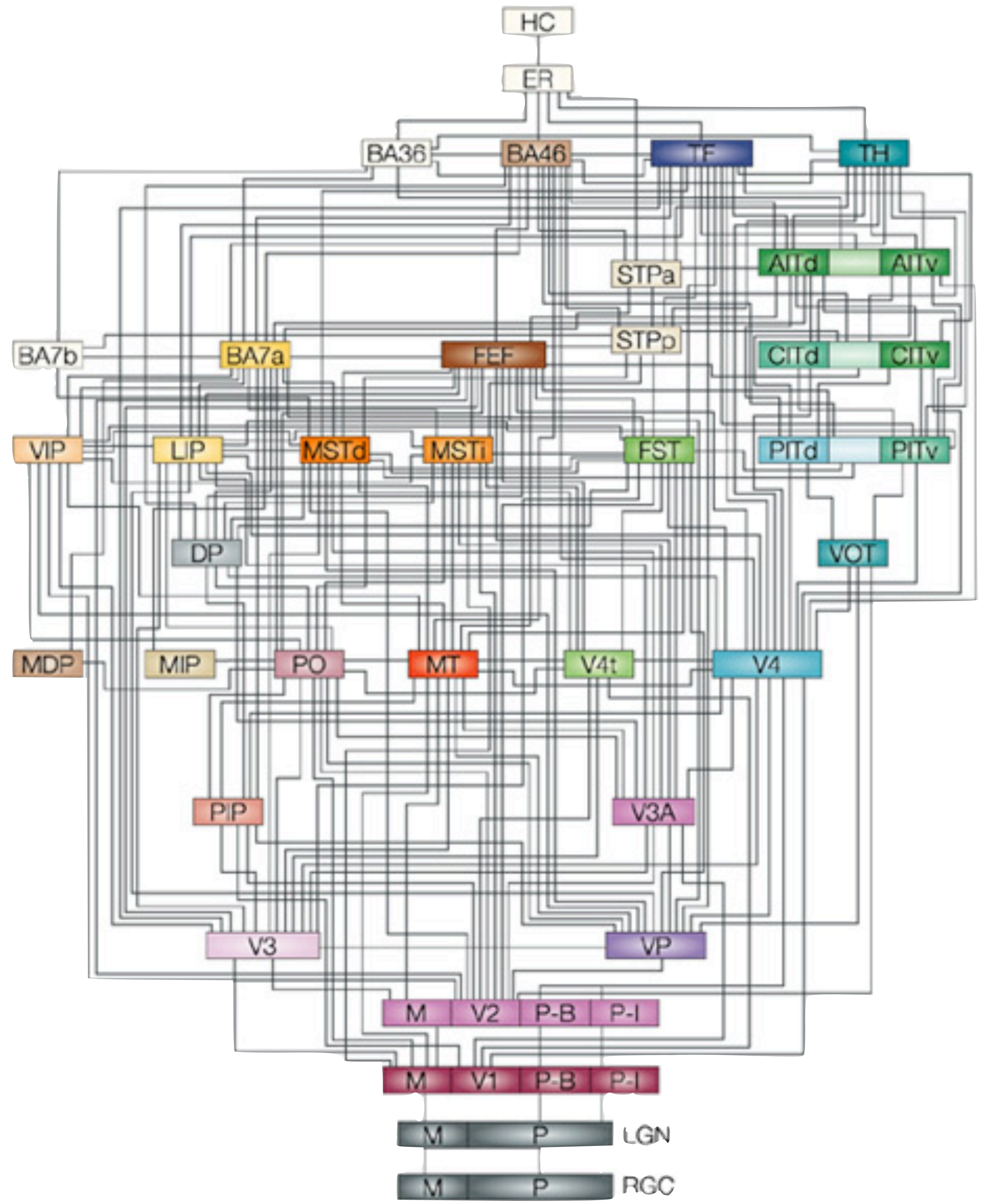
Li & DiCarlo '08



Matching object representations in man and monkey

Kriegeskorte et al ’08

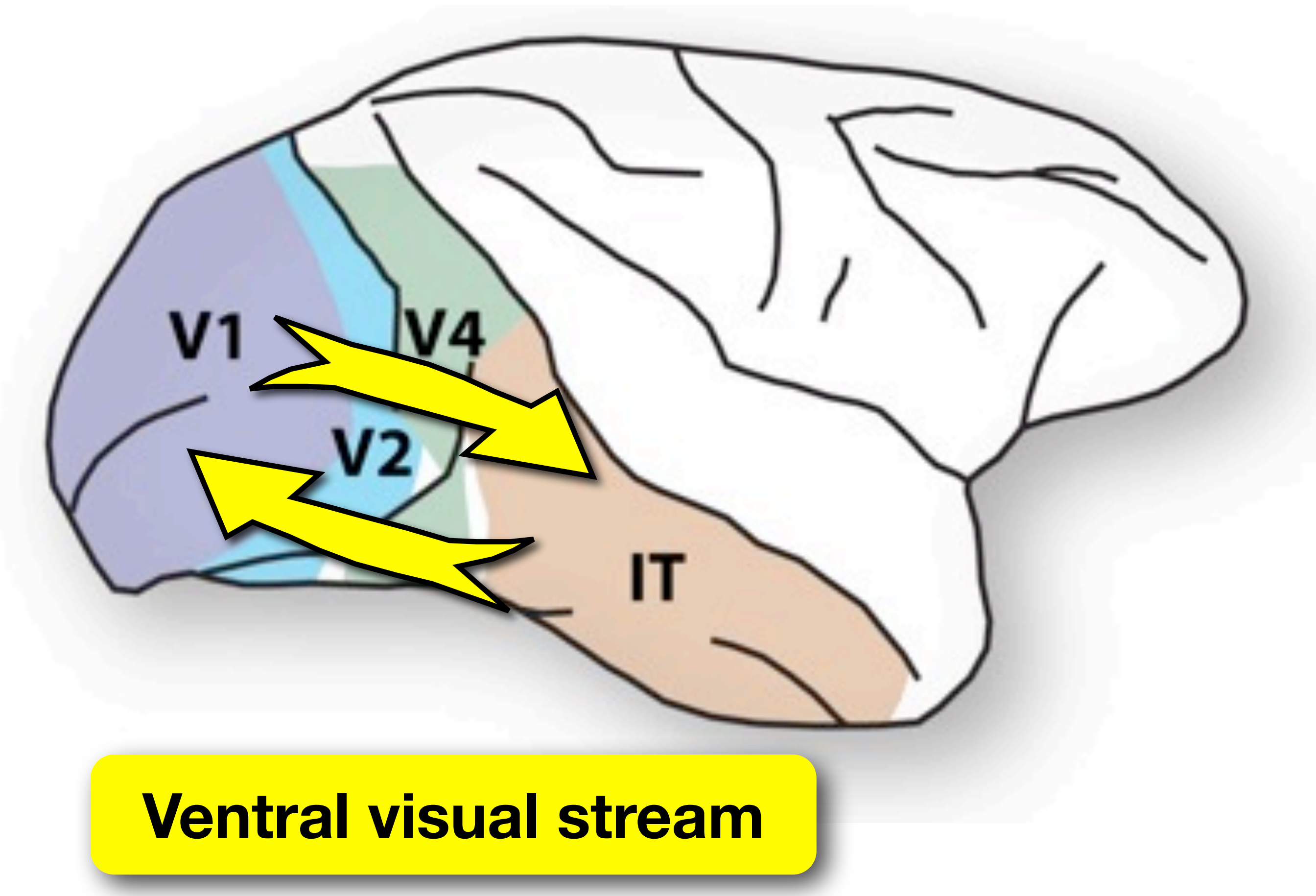




Summary

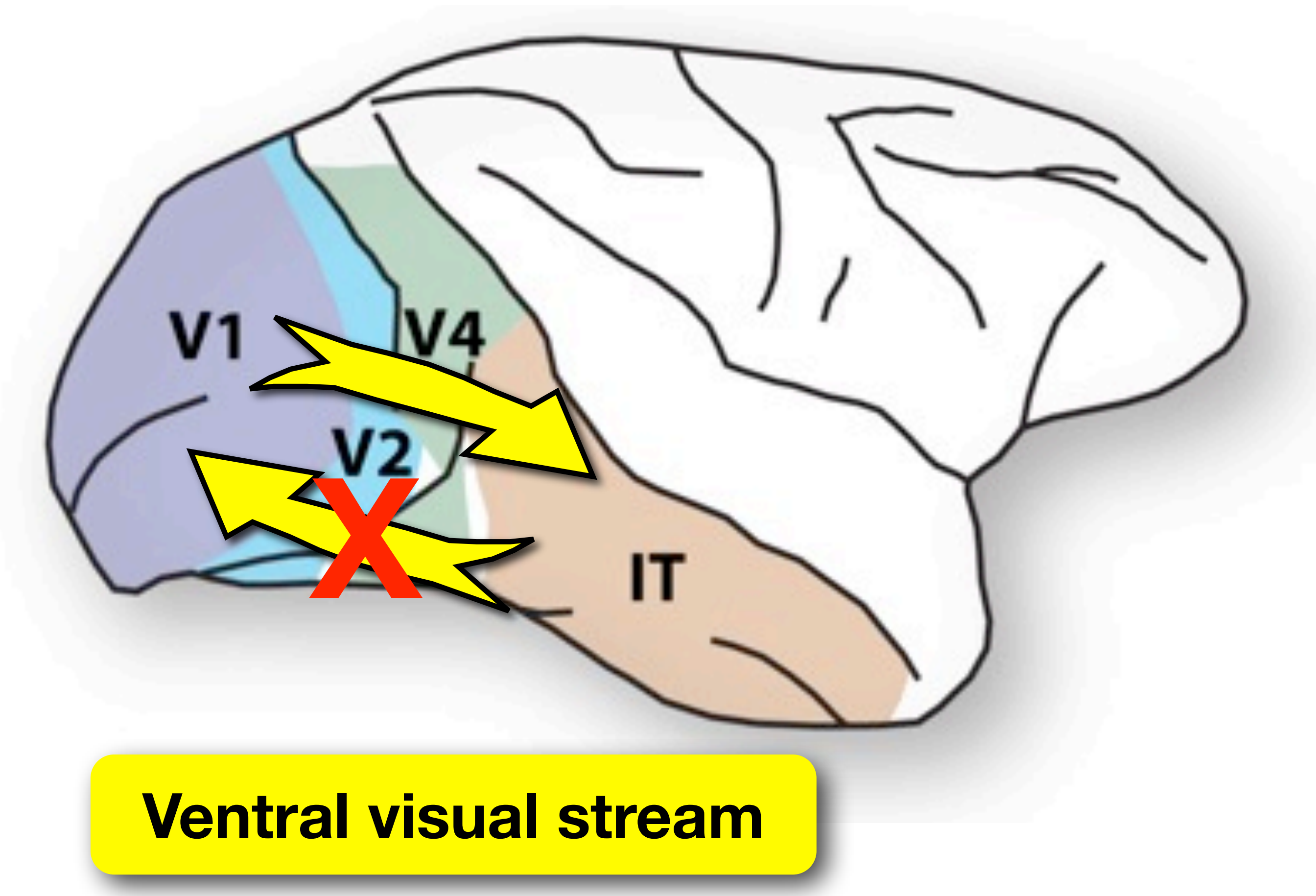
Hierarchies are ubiquitous:
Anatomy, function & latencies

source: Felleman & VanEssen ’90



Summary

Two modes of processing:
Bottom-up vs. recurrent



Summary

Two modes of processing:
Bottom-up vs. recurrent

Deep learning in the visual cortex

Thomas Serre
Brown University

I. Fundamentals of primate vision

II. Computational mechanisms of rapid recognition and feedforward processing

III. Beyond feedforward processing: Attentional mechanisms and cortical feedback

Vision is complex but the solution might be simple...

Cavanagh '95