# *Lecture 14. Object Recognition*

## *Reading Assignments:*

None

# Next Time

**You** will tell us about your project!

Prepare a 15-minute presentation with:

1) background: what problem are you attacking?
why is it important?
what has been done previously?

2) design: what is your approach?
how is it different from previous ones?
explain what you did.

3) results & conclusion: what did you find? did it work?
what did you learn? was the project useful?

# Next Time

Grading: we'll have a peer-review grading system ;-)

While others present, take critical notes on what is nice/interesting, what sounds bogus, what is missing. Assign a score to the presentation. I will give you guidelines, to evaluate:

- scientific value
- quality of presentation
- significance of results
- overall

Keep this information secret.

Once all the presentations have been made, rank the presentations overall.

Give me your notes.

Final grade will be: quizzes (10%), white paper (20%), scientific value (20%), quality of presentation (20%), significance of results (20%), rank (10%)

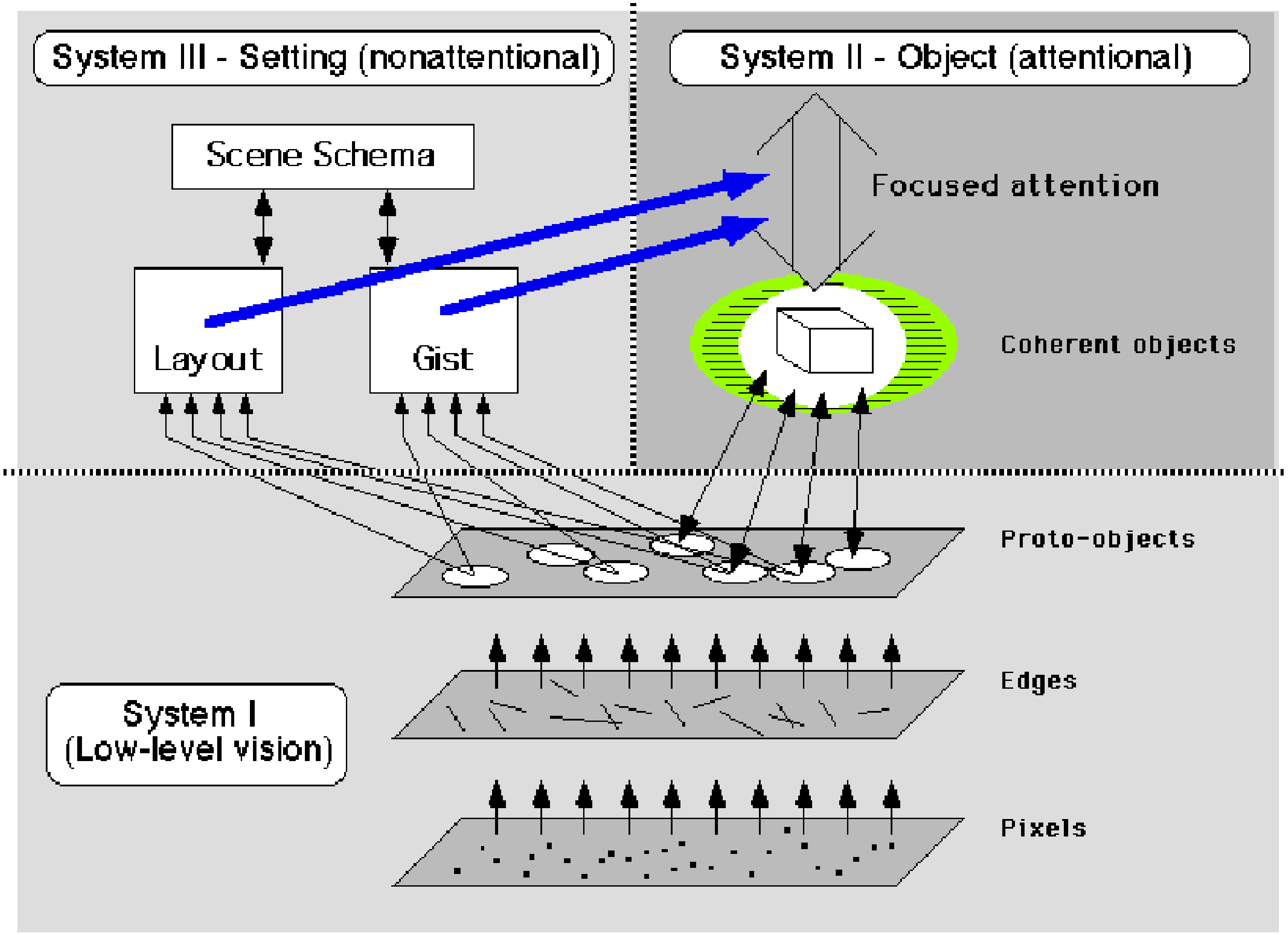


Figure 4

Figure 4. Triadic Architecture. It is suggested that the visual perception of scenes may be carried out via the interaction of three different systems. System I: Early-level processes produce volatile proto-objects rapidly and in parallel across the visual field. System II: Focused attention acts as a hand to "grab" these structures; as long as these structures are held, they form an individuated object with both temporal and spatial coherence. System III: Setting information—obtained via a nonattentional stream—guides the allocation of focused attention to various parts of the scene, and allows priorities to be given to the various possible objects.

# Four stages of representation (Marr, 1982)

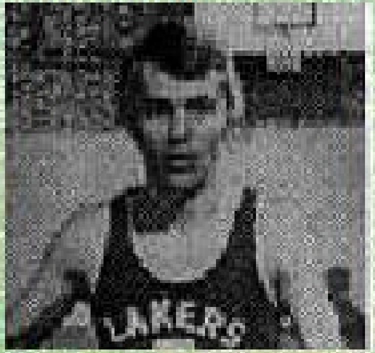
1) pixel-based (light intensity)

2) primal sketch (discontinuities in intensity)

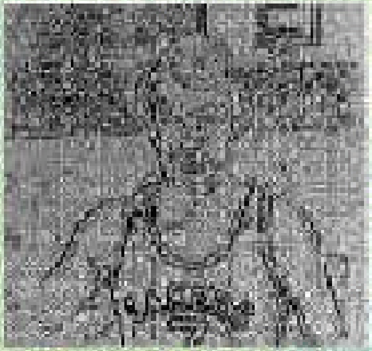
3) 2 ½ D sketch (oriented surfaces, relative depth between surfaces)

4) 3D model (shapes, spatial relationships, volumes)

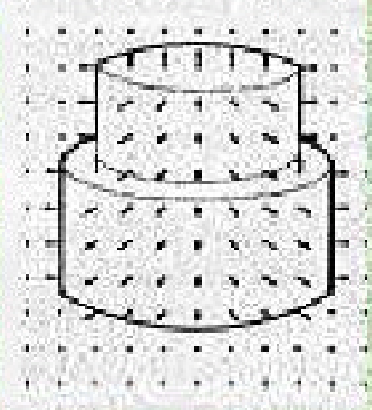
# Real world example of Marr



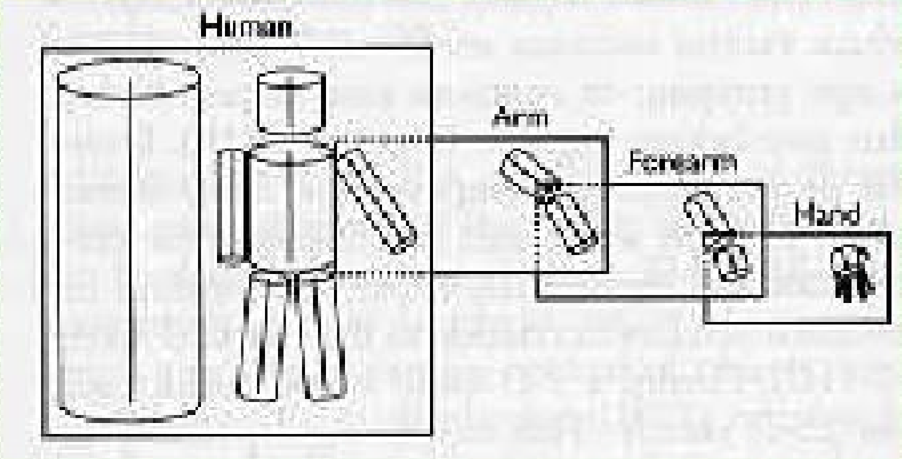
Retinal image



Primal sketch



2 1/2 D sketch



3D model

# Challenges of Object Recognition

The binding problem: binding different features (color, orientation, etc) to yield a unitary percept. (see next slide)

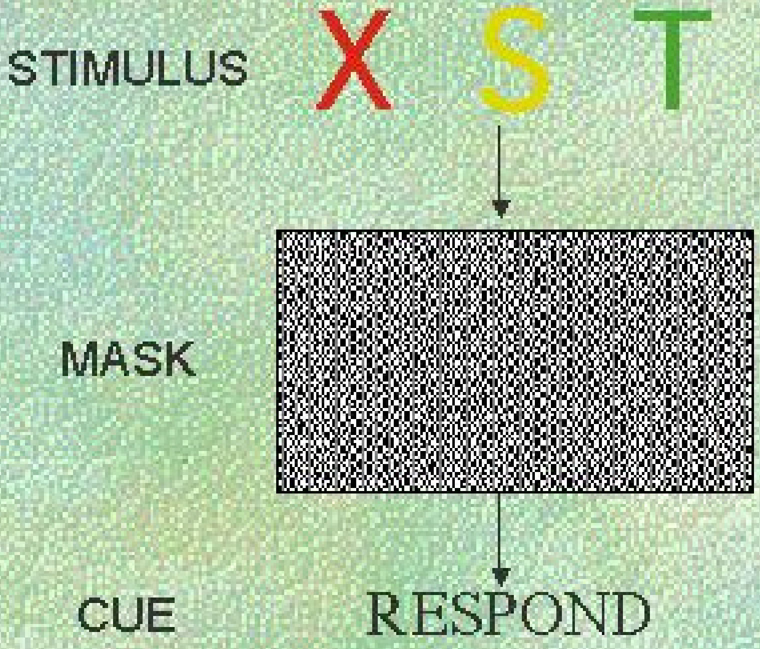
Bottom-up vs. top-down processing: how much is assumed top-down vs. extracted from the image?

Perception vs. recognition vs. categorization: seeing an object vs. seeing is *as* something. Matching views of known objects to memory vs. matching a novel object to object categories in memory.

Viewpoint invariance: a major issue is to recognize objects irrespectively of the viewpoint from which we see them.

# illusory conjunctions

- Present colored letters very briefly, then mask
- On report, subjects correctly report the letters and the color but on ~30% or trials, attach wrong color to letter
    - e.g. Green X



# *Viewpoint Invariance*

Major problem for recognition.

Biederman & Gerhardstein, 1994:

We can recognize two views of an unfamiliar object as being the same object.

Thus, viewpoint invariance cannot only rely on matching views to memory.

# *Models of Object Recognition*

See Hummel, 1995, The Handbook of Brain Theory & Neural Networks

Direct Template Matching:

Processing hierarchy yields activation of view-tuned units.

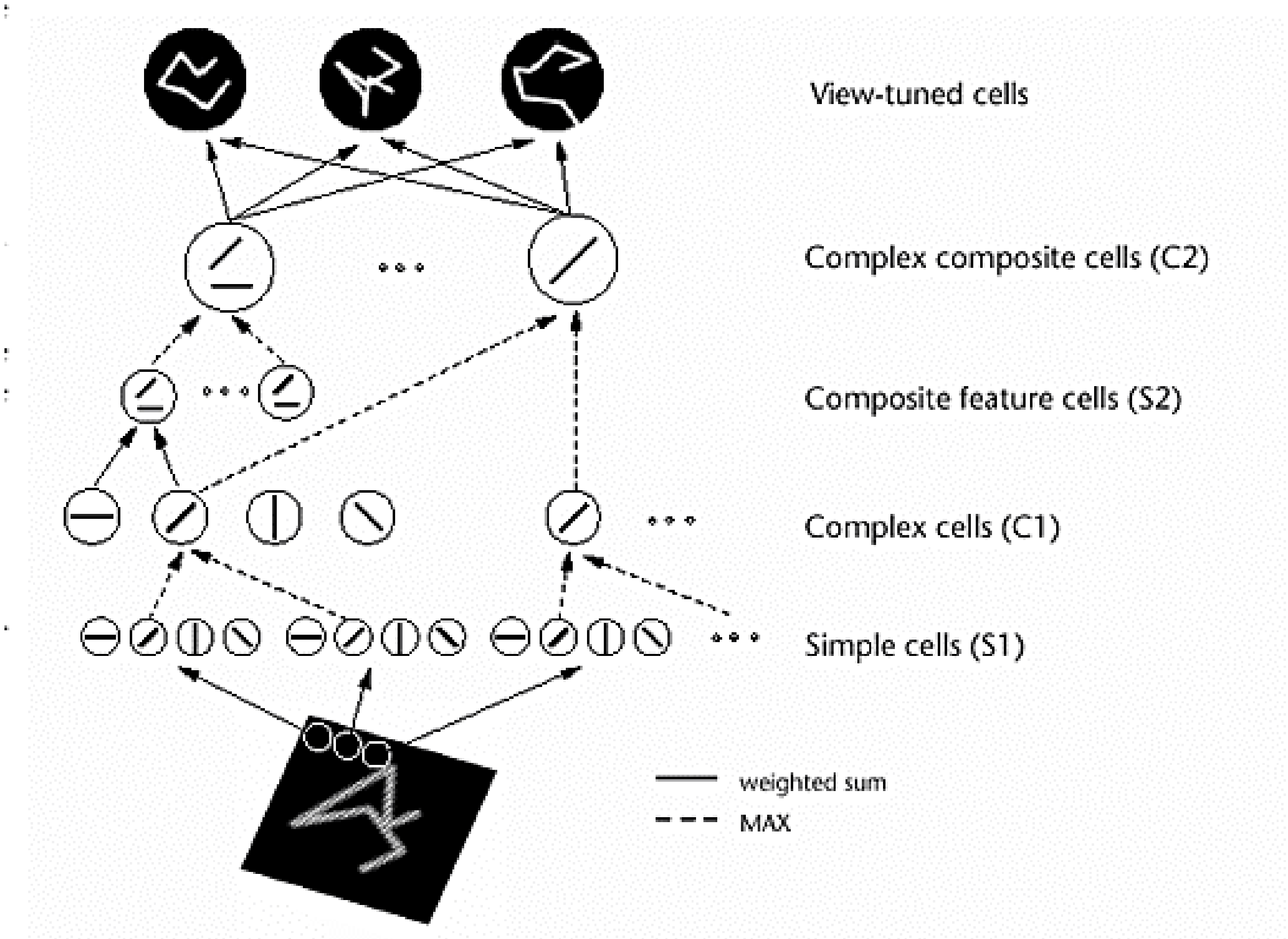
A collection of view-tuned units is associated with one object.

View tuned units are built from V4-like units,
using sets of weights which differ for each object.

e.g., Poggio & Edelman, 1990; Riesenhuber & Poggio, 1999

# Computational Model of Object Recognition

## (Riesenhuber and Poggio, 1999)



the model neurons are tuned for size and 3D orientation of object

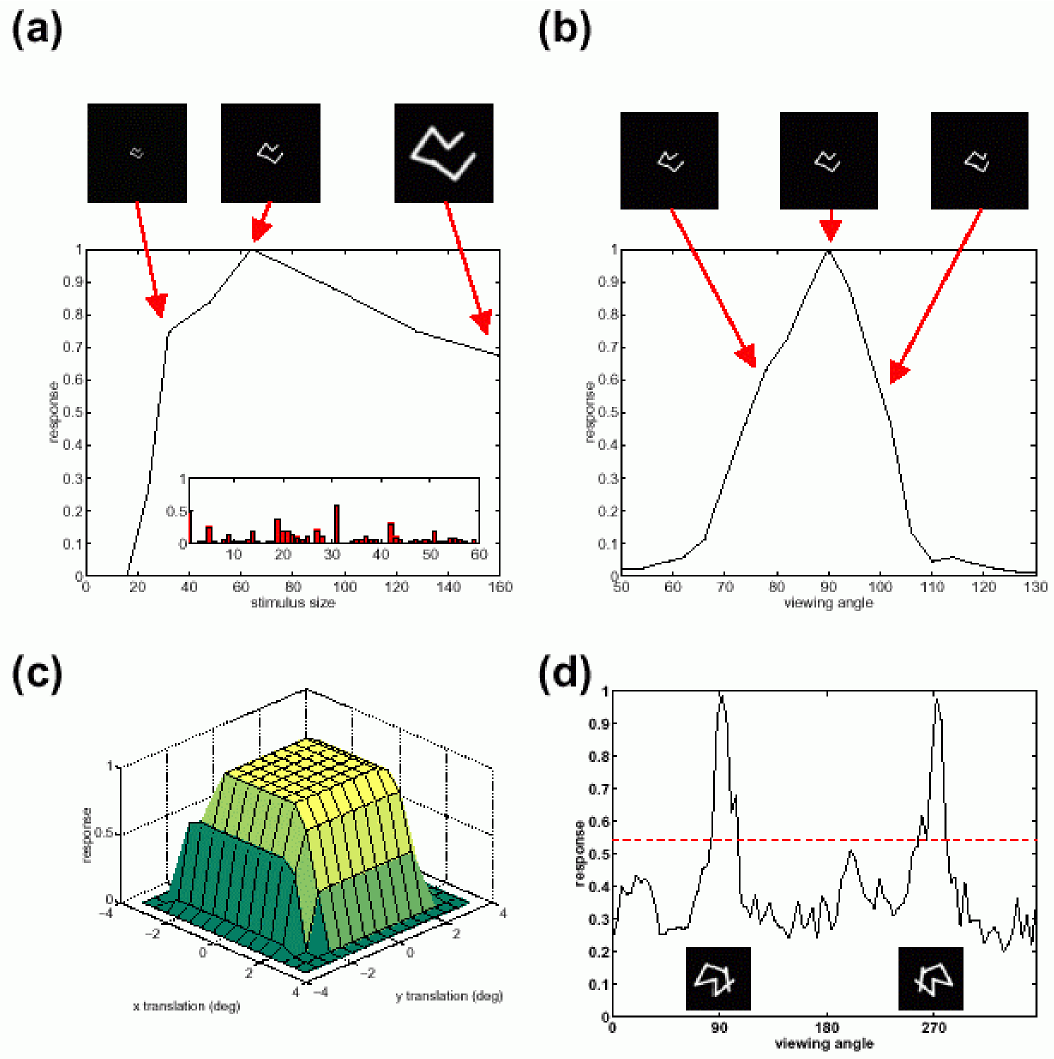


Figure 4: Responses of a sample model neuron to different transformations of its preferred stimulus. The different panels show the same neuron's response to **(a)** varying stimulus sizes (inset shows response to 60 distractor objects, selected randomly from the paperclips used in the physiology experiments[21]), **(b)** rotation in depth and **(c)** translation. Training size was 64 × 64 pixels corresponding to $2^o$ of visual angle. **(d)** shows another neuron's response to pseudo-mirror views (cf. text), with the dashed line indicating the neuron's response to the "best" distractor.

# Models of Object Recognition

## Hierarchical Template Matching:

Image passed through layers of units with progressively more complex features at progressively less specific locations.

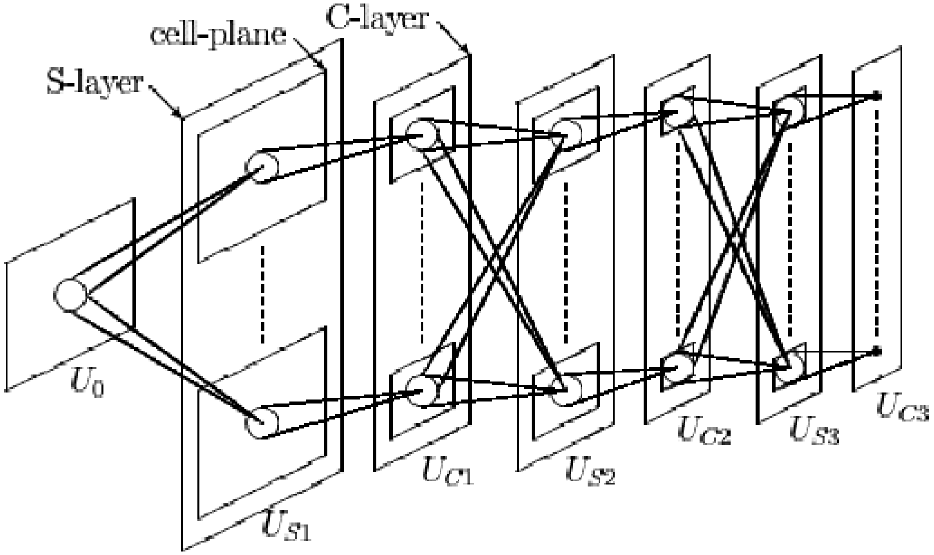
Hierarchical in that features at one stage are built from features at earlier stages.

e.g., Fukushima & Miyake (1982)'s Neocognitron:

Several processing layers, comprising simple (S) and complex (C) cells.

S-cells in one layer respond to conjunctions of C-cells in previous layer.

C-cells in one layer are excited by small neighborhoods of S-cells.



# *Models of Object Recognition*

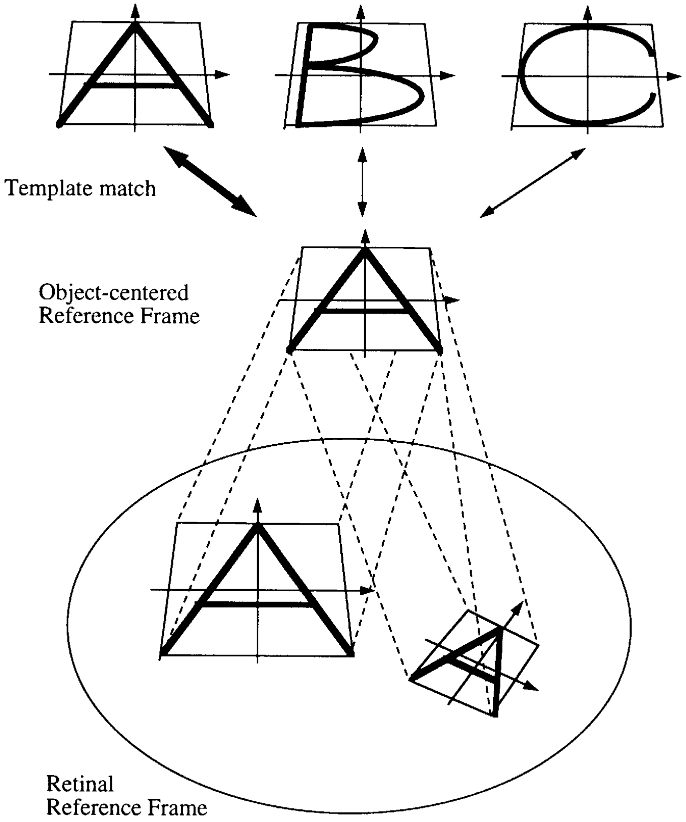
Transform & Match:

First take care of rotation, translation, scale, etc. invariances.

Then recognize based on standardized pixel representation of objects.

e.g., Olshausen et al, 1993,
dynamic routing model

Template match: e.g., with
an associative memory based on
a Hopfield network.



# *Recognition by Components*

Structural approach to object recognition:

Biederman, 1987:

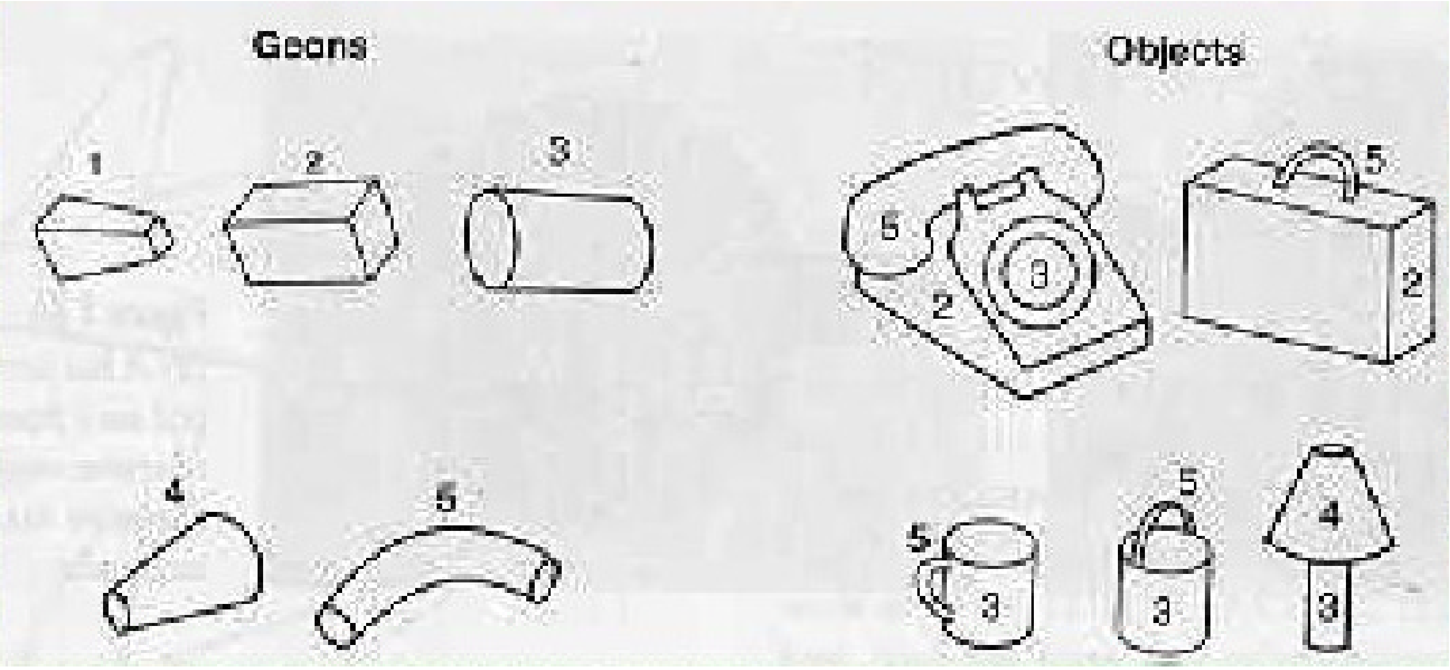
Complex objects are composed so simpler pieces

We can recognize a novel/unfamiliar object by parsing it in terms of its component pieces, then comparing the assemblage of pieces to those of known objects.

# Recognition by components (Biederman, 1987)

**GEONS:** geometric elements of which all objects are composed (cylinders, cones, etc). On the order of 30 different shapes.

Skips 2 ½ D sketch: Geons are directly recognized from edges, based on their nonaccidental properties (i.e., 3D features that are usually preserved by the projective imaging process).

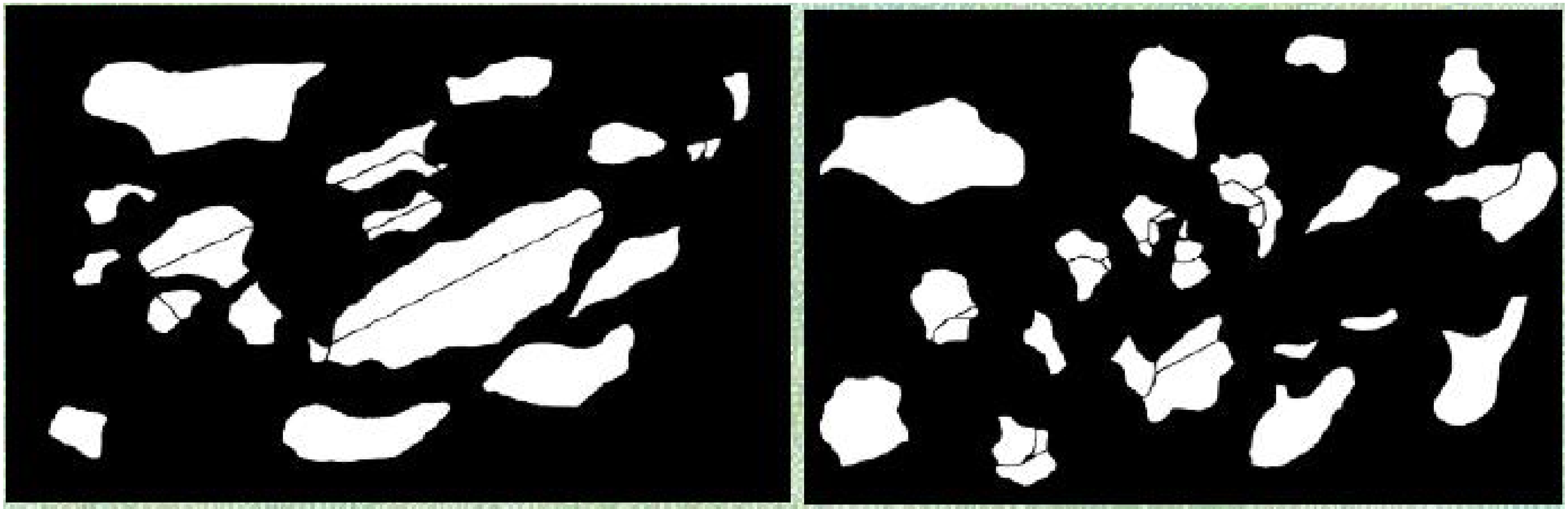


# *Basic Properties of GEONs*

They are sufficiently different from each other to be easily discriminated

They are view-invariant (look identical from most viewpoints)

They are robust to noise (can be identified even with parts of image missing)



Obscured geons

Visible geons

Support for RBC: We can recognize partially occluded objects easily if the occlusions do not obscure the set of geons which constitute the object.

# *Potential difficulties*

A. Structural description not enough, also need metric info

B. Difficult to extract geons from real images

C. Ambiguity in the structural description: most often we have several candidates

D. For some objects, deriving a structural representation can be difficult

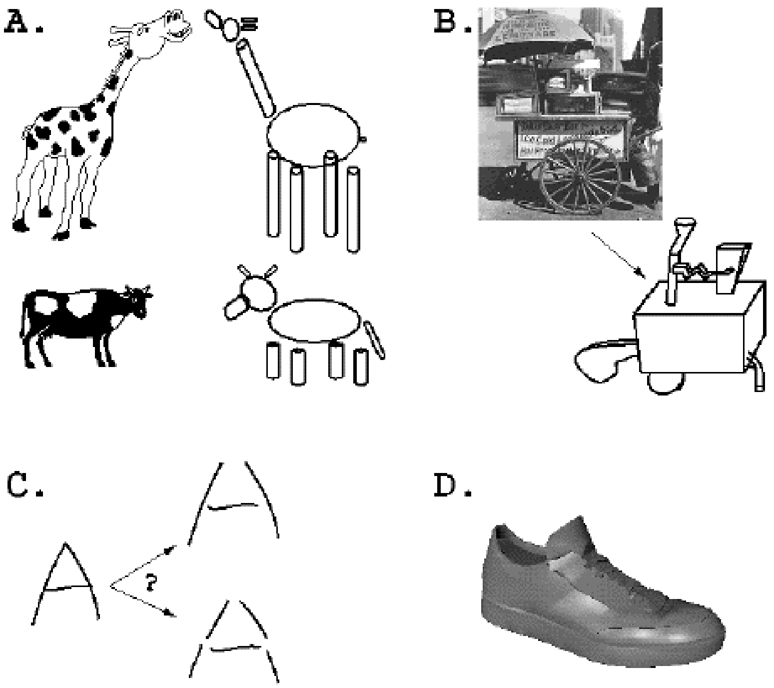
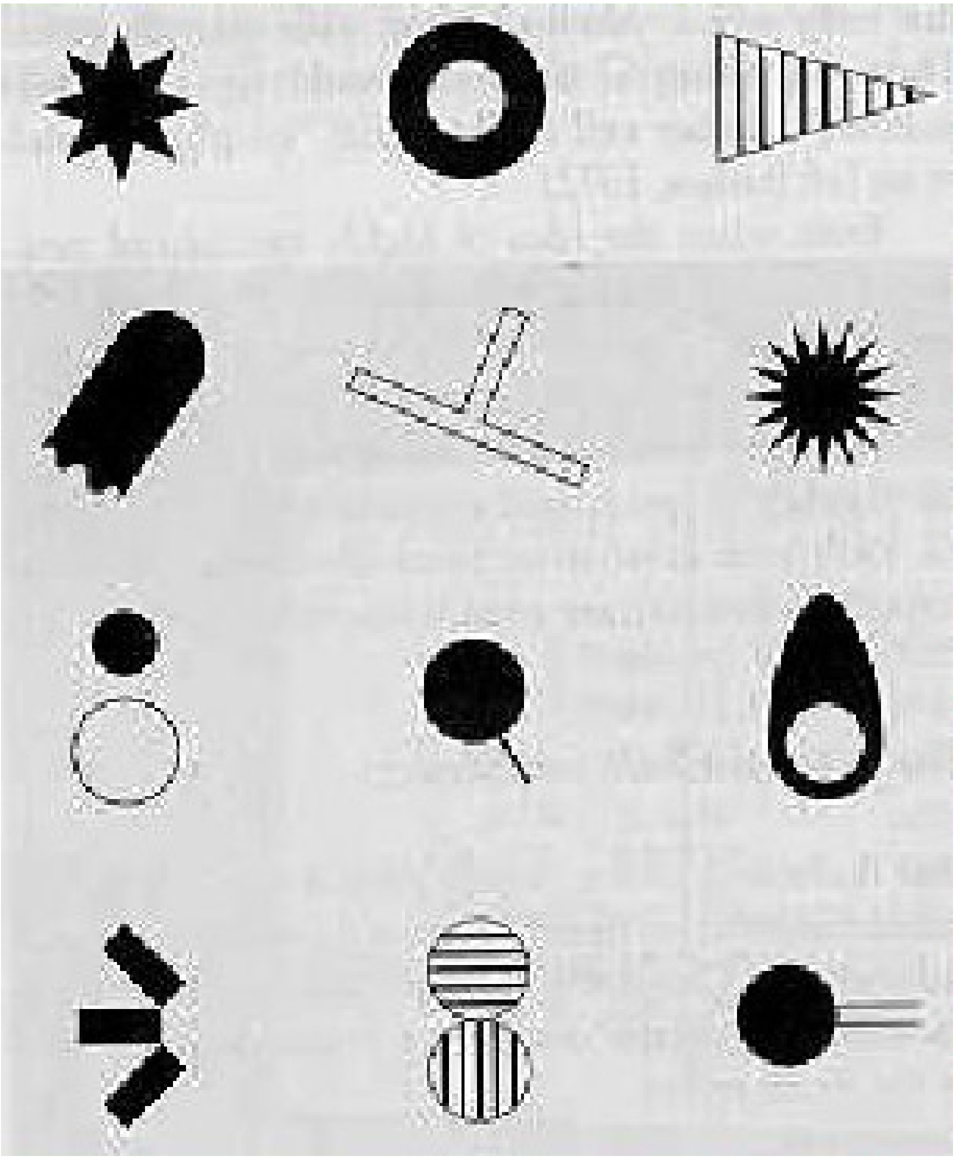


Figure 1: Computational problems with structural representations. A. Structural descriptions must be accompanied by metric information, to represent differences among commonly encountered categories. The inclusion of metric details reduces the ability of structural methods to deal with novel objects. B. A picture of a New York City street-corner hot dog cart, and a stylized object, which, as Biederman [8] suggests, may be described as such following a structural decomposition in the visual system. At present, there is no reliable method for mapping a gray-level image into a collection of (labeled) primitives (lines, corners, etc.) from which RBC's geons are constructed. Thus, although a carefully engineered system such as that described in [22] can form a structural description of the line drawing of a cart-like object, the goal of deriving such a description directly from an image remains elusive. C. Even in simpler tasks (e.g., in character recognition, where the figure is readily separable from the ground), the derivation of a structural description is problematic. The difficulty here stems from the possibility to assign multiple structural descriptions to the same image. D. In some tasks, coming up even with one structural description is problematic: how does one represent a shoe in terms of RBC's geons [7]?

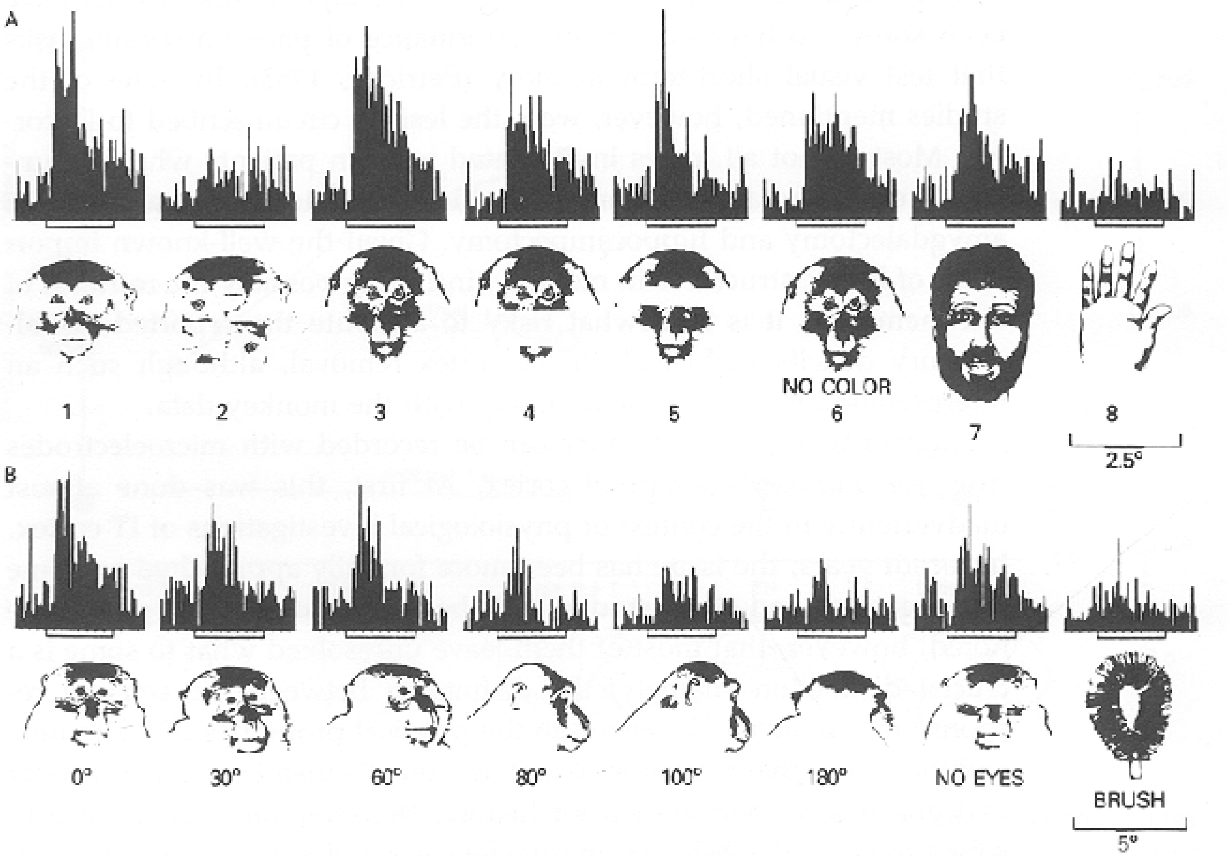
Edelman, 1997

# Geon Neurons in IT?

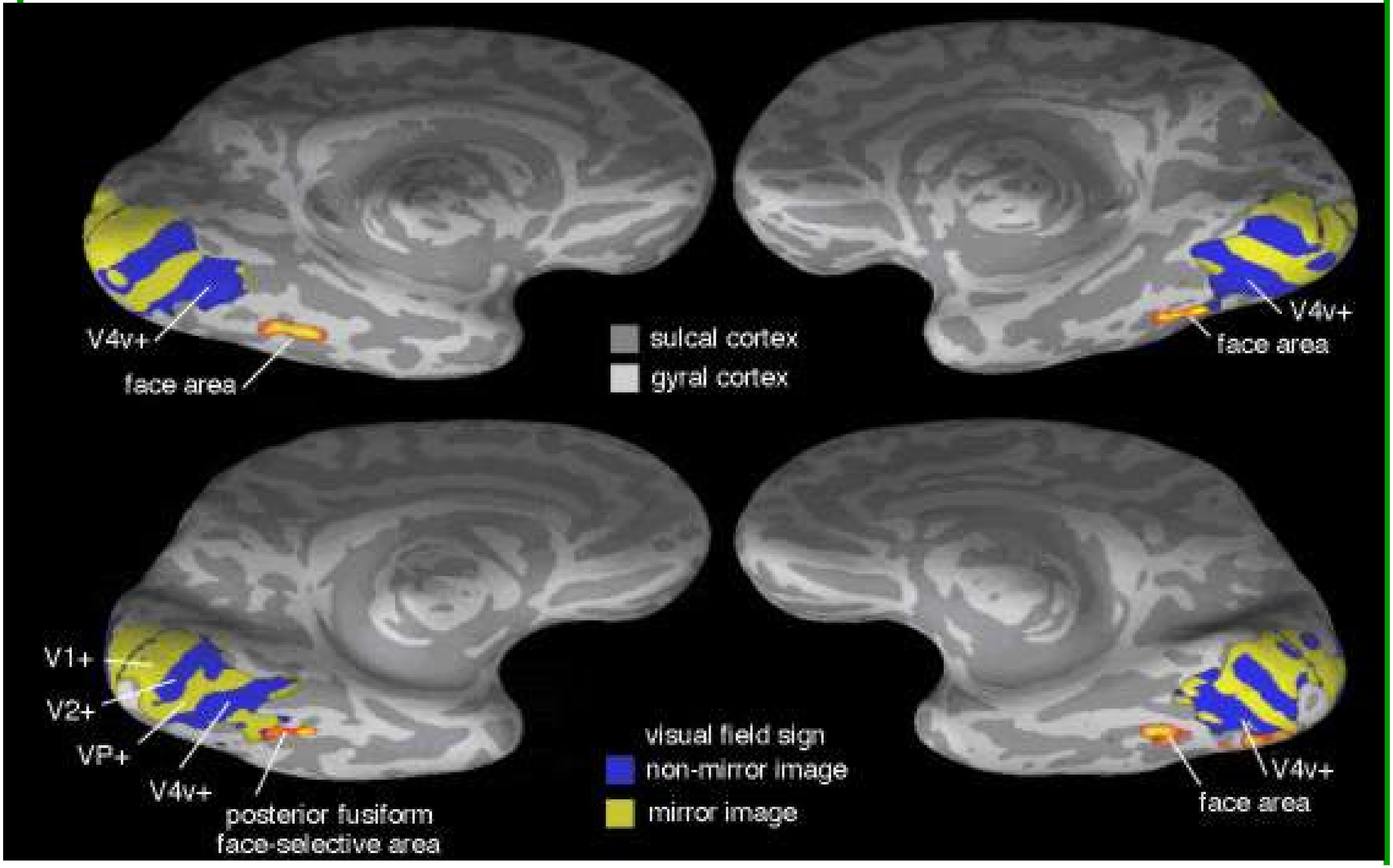
These are preferred stimuli for some IT neurons.



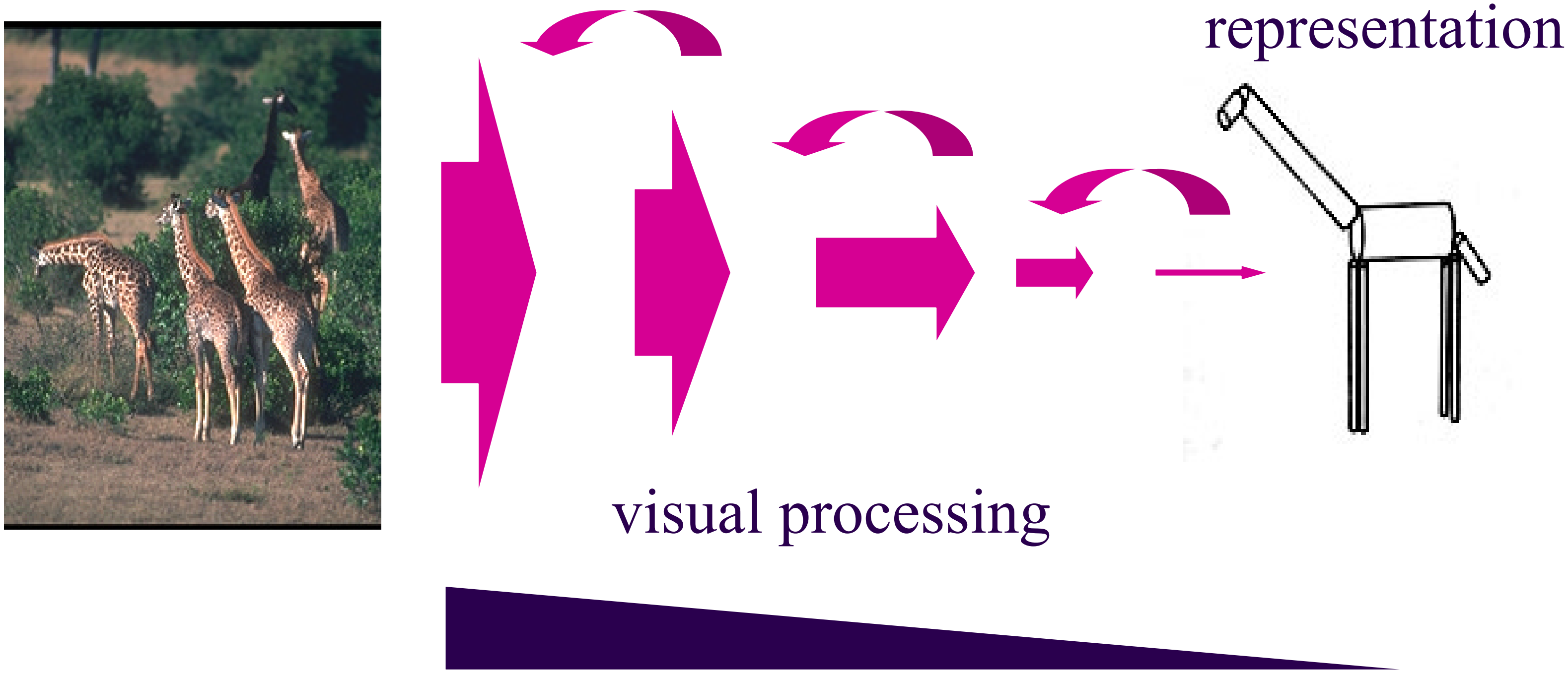
# “Face cells”



# Fusiform Face Area in Humans



# *Standard View on Visual Processing*



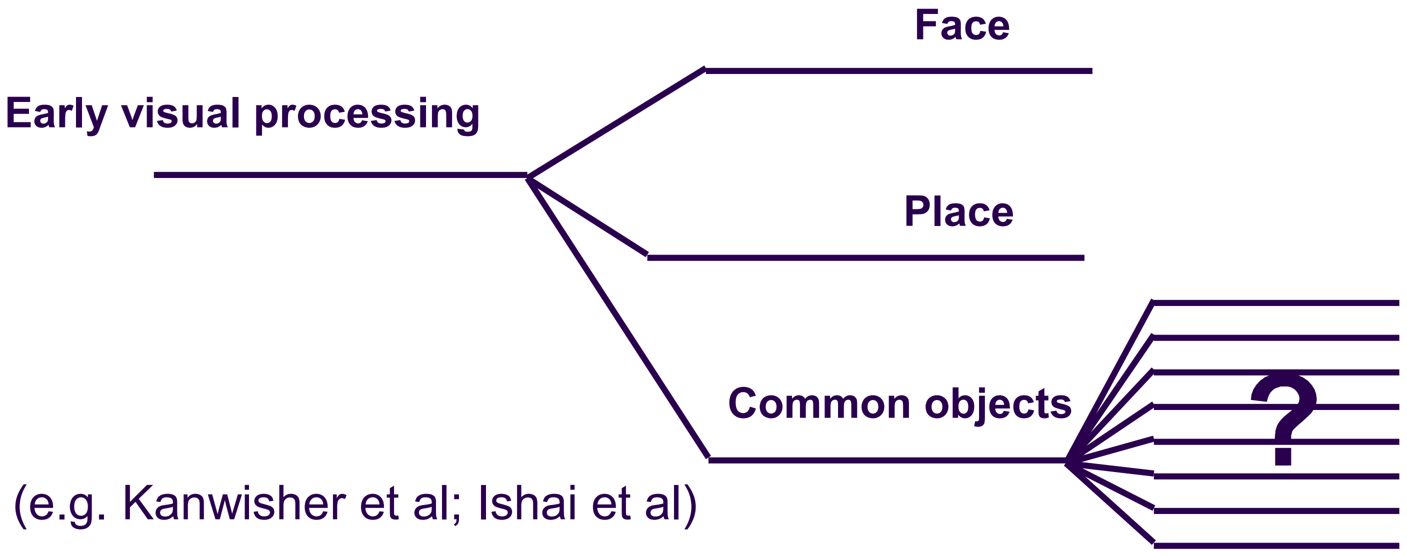
representation

visual processing

- **Image specific**
- **Supports fine discrimination**
- **Noise tolerant**

- **Image invariant**
- **Supports generalization**
- **Noise sensitive**

Tjan, 1999



**Early visual processing**

**Face**

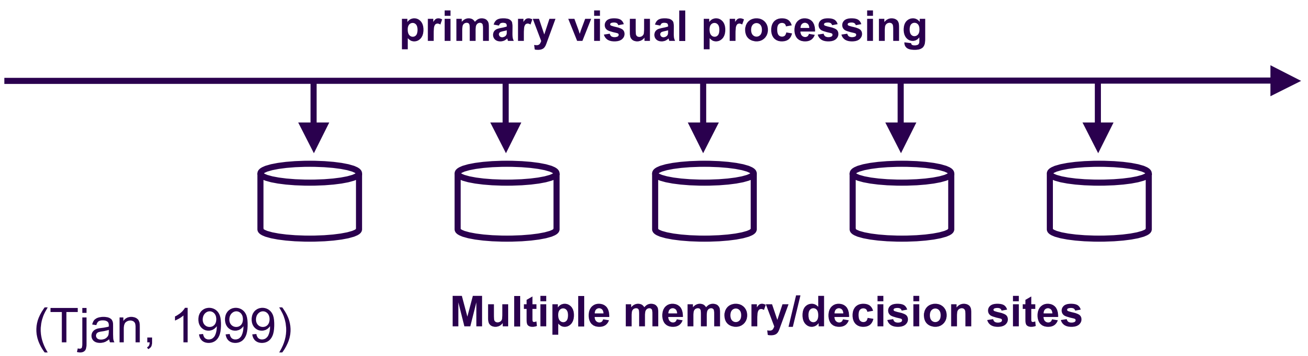
**Place**

**Common objects**

**?**

(e.g. Kanwisher et al; Ishai et al)

---

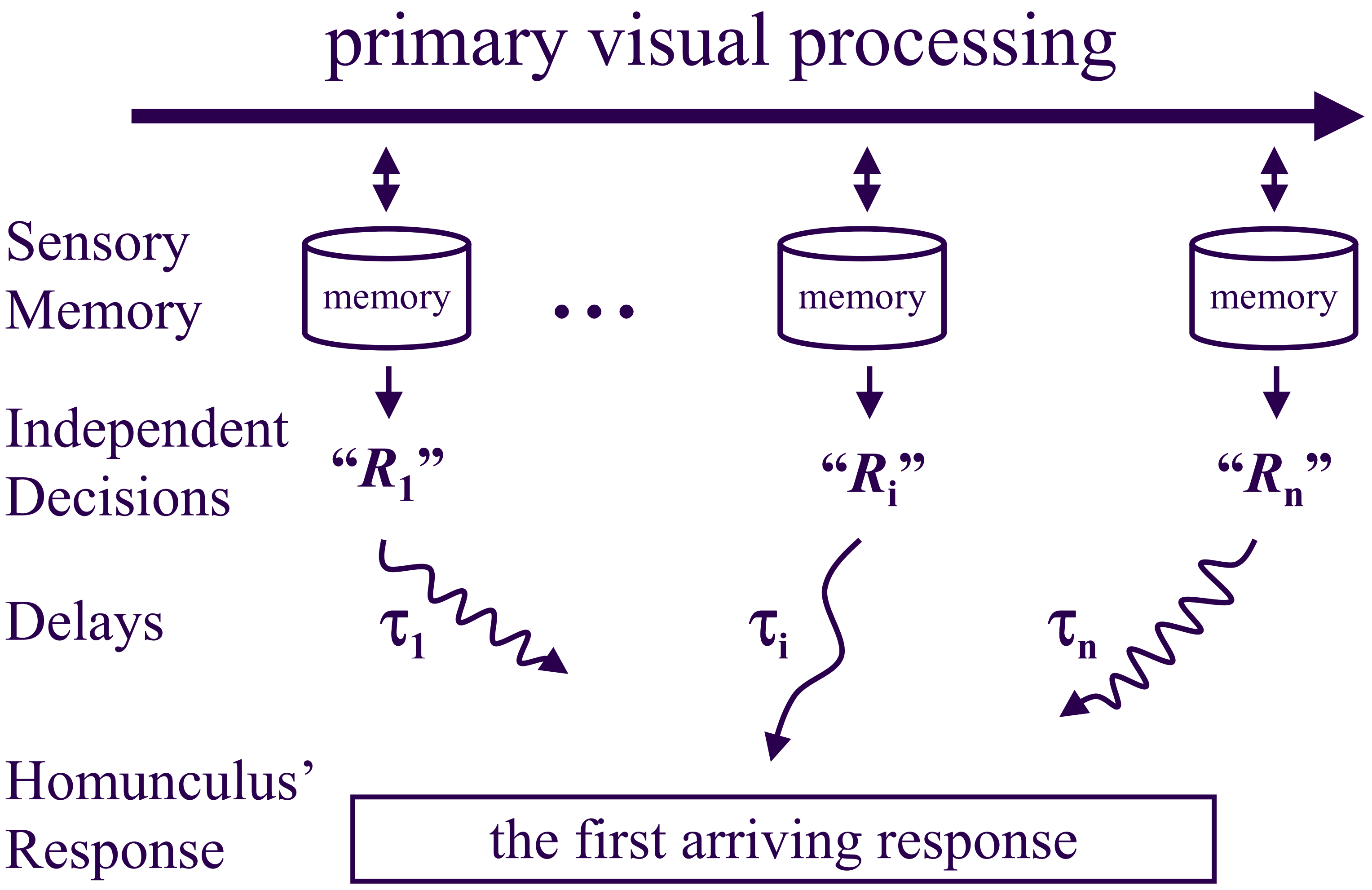


**primary visual processing**

**Multiple memory/decision sites**

(Tjan, 1999)

# *Tjan's "Recognition by Anarchy"*



primary visual processing

Sensory Memory

memory … memory memory

Independent Decisions

"$R_1$" "$R_i$" "$R_n$"

Delays

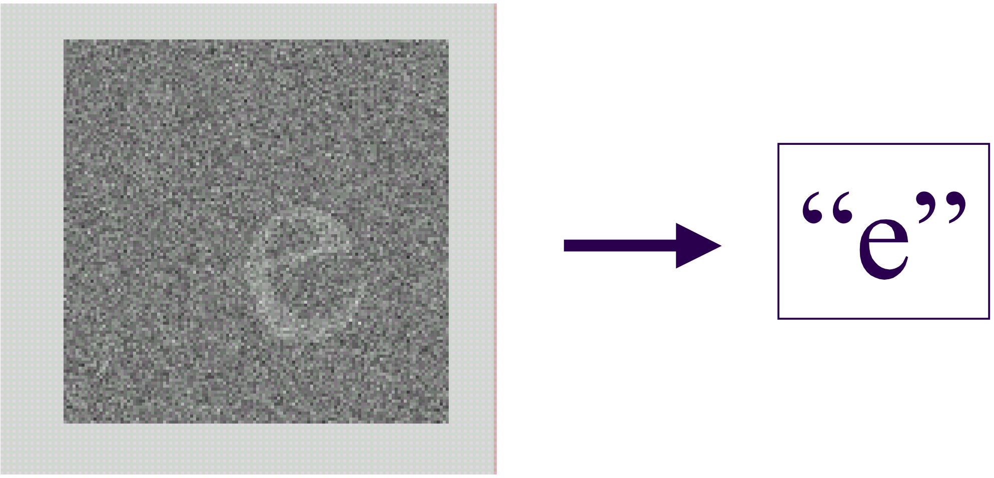
$\tau_1$ $\tau_i$ $\tau_n$

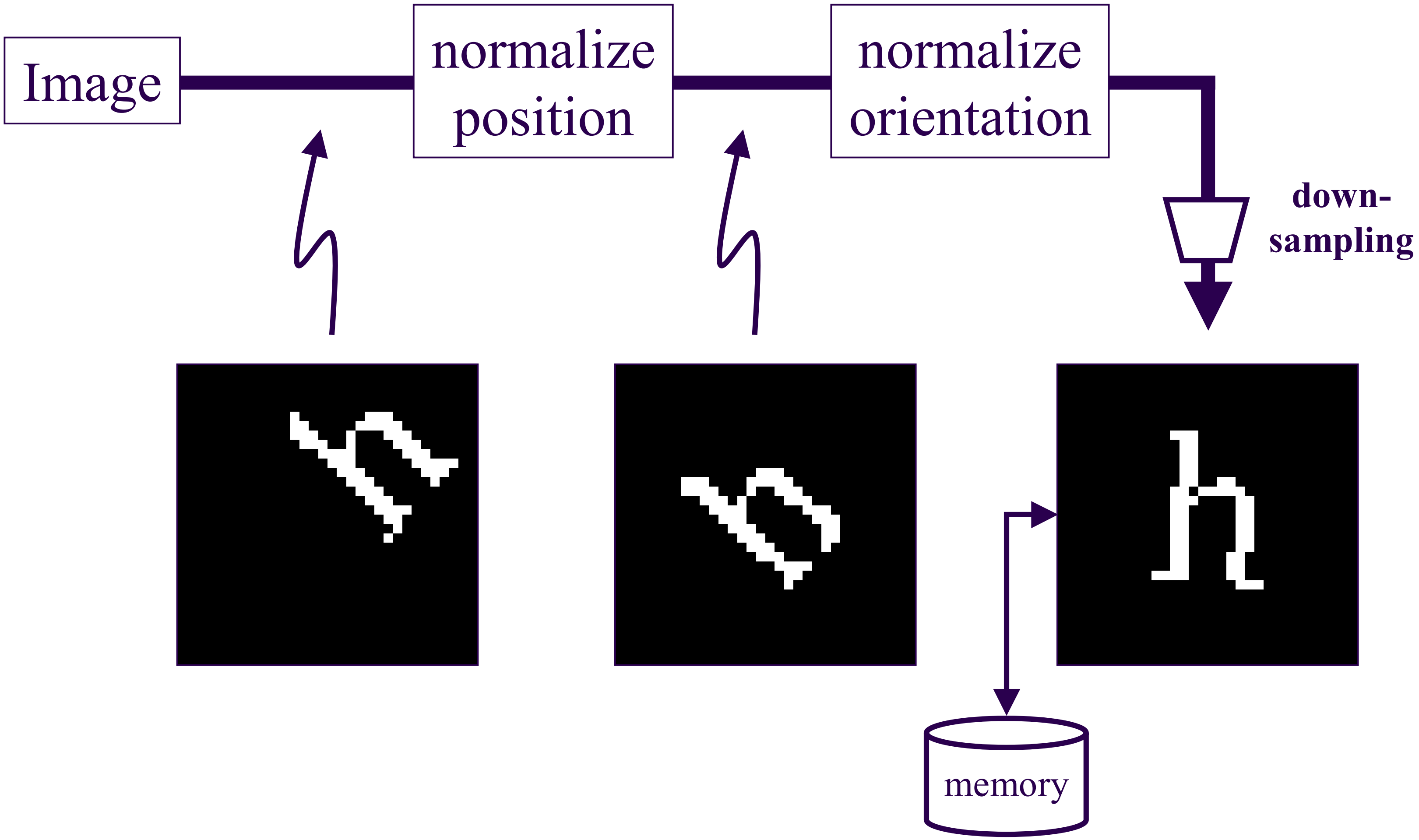
Homunculus' Response

the first arriving response

# *A toy visual system*

***Task:*** ***Identify letters from arbitrary positions & orientations***





Image

normalize position

normalize orientation

**down-sampling**

memory

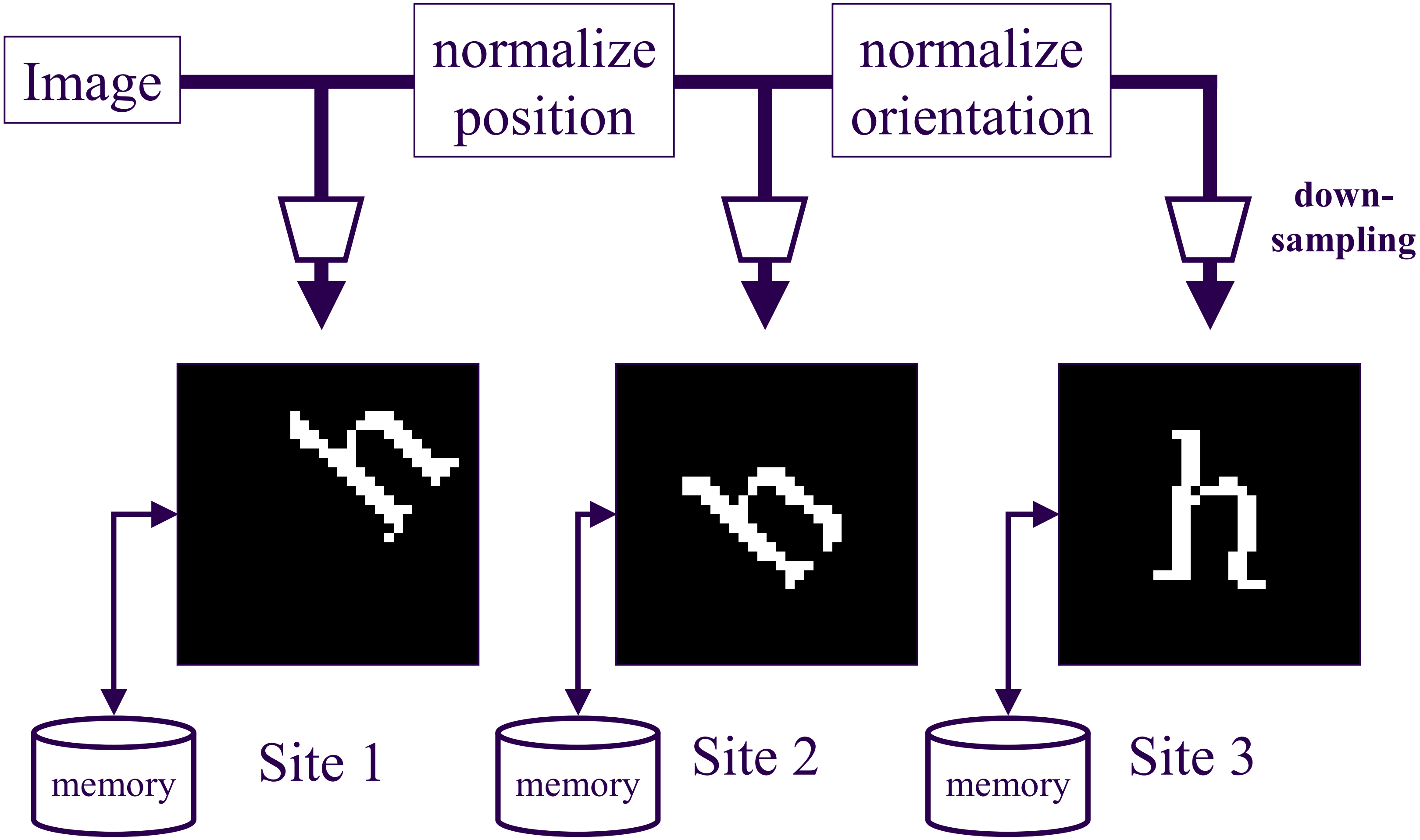


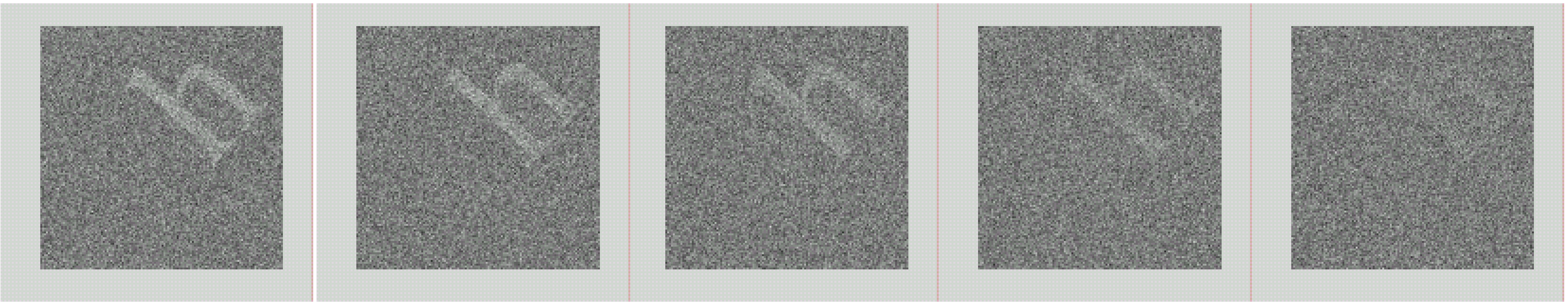
Image
normalize position
normalize orientation
down-sampling
memory
Site 1
memory
Site 2
memory
Site 3

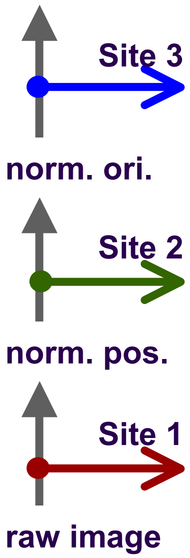
## *Study stimuli:*

***5 orientations × 20 positions at high SNR***

## *Test stimuli:*

***1) familiar (studied) views,***
***2) new positions,***
***3) new position & orientations***





Site 3

norm. ori.

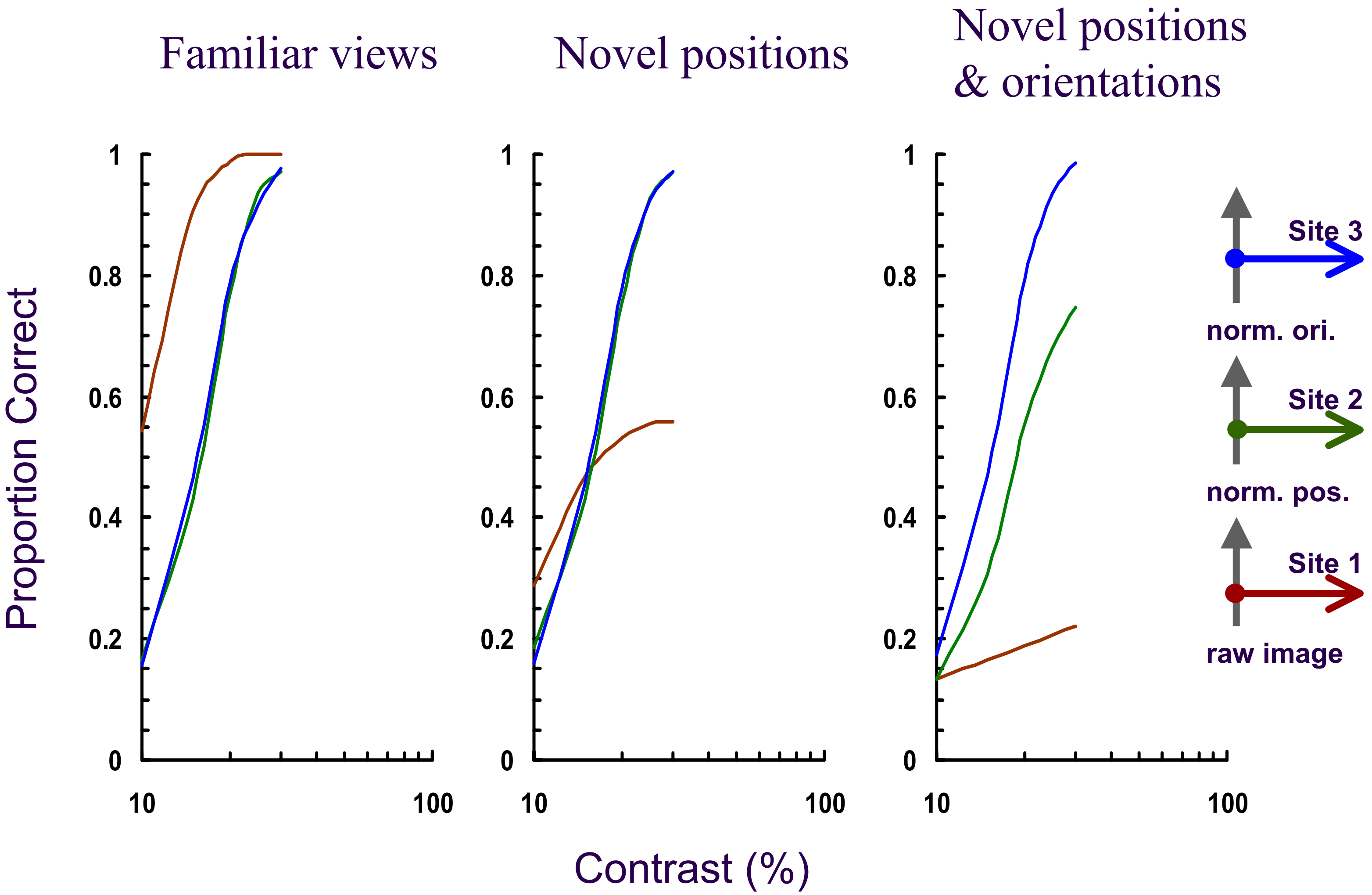
Site 2

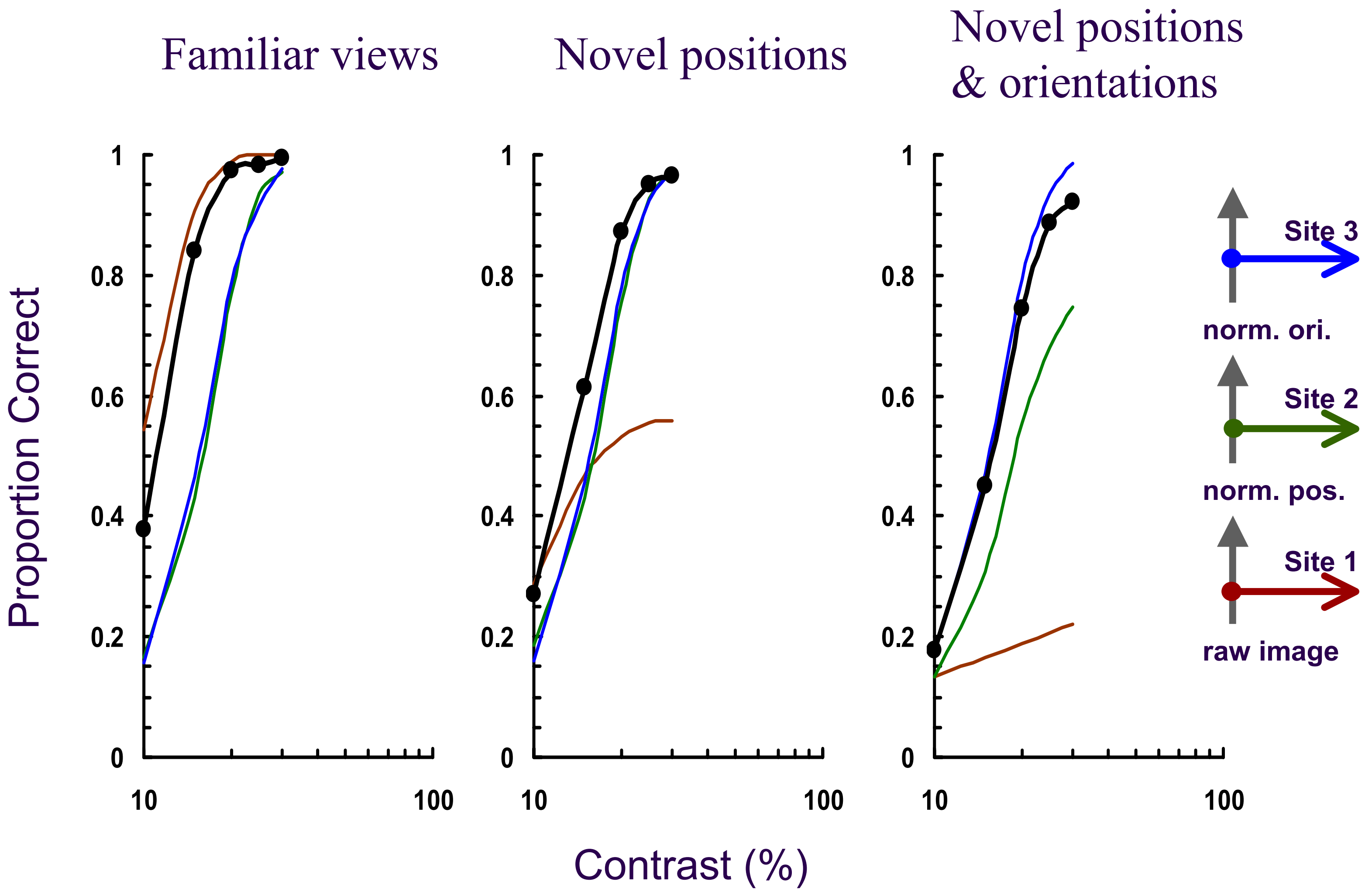
norm. pos.

Site 1

raw image

Processing speed for each recognition module depends on recognition difficulty by that module.


Familiar views
Novel positions
Novel positions & orientations
Proportion Correct
1
0.8
0.6
0.4
0.2
0
10
100
Contrast (%)
Site 3
norm. ori.
Site 2
norm. pos.
Site 1
raw image



Black curve: full model in which recognition is based on the fastest of the responses from the three stages.

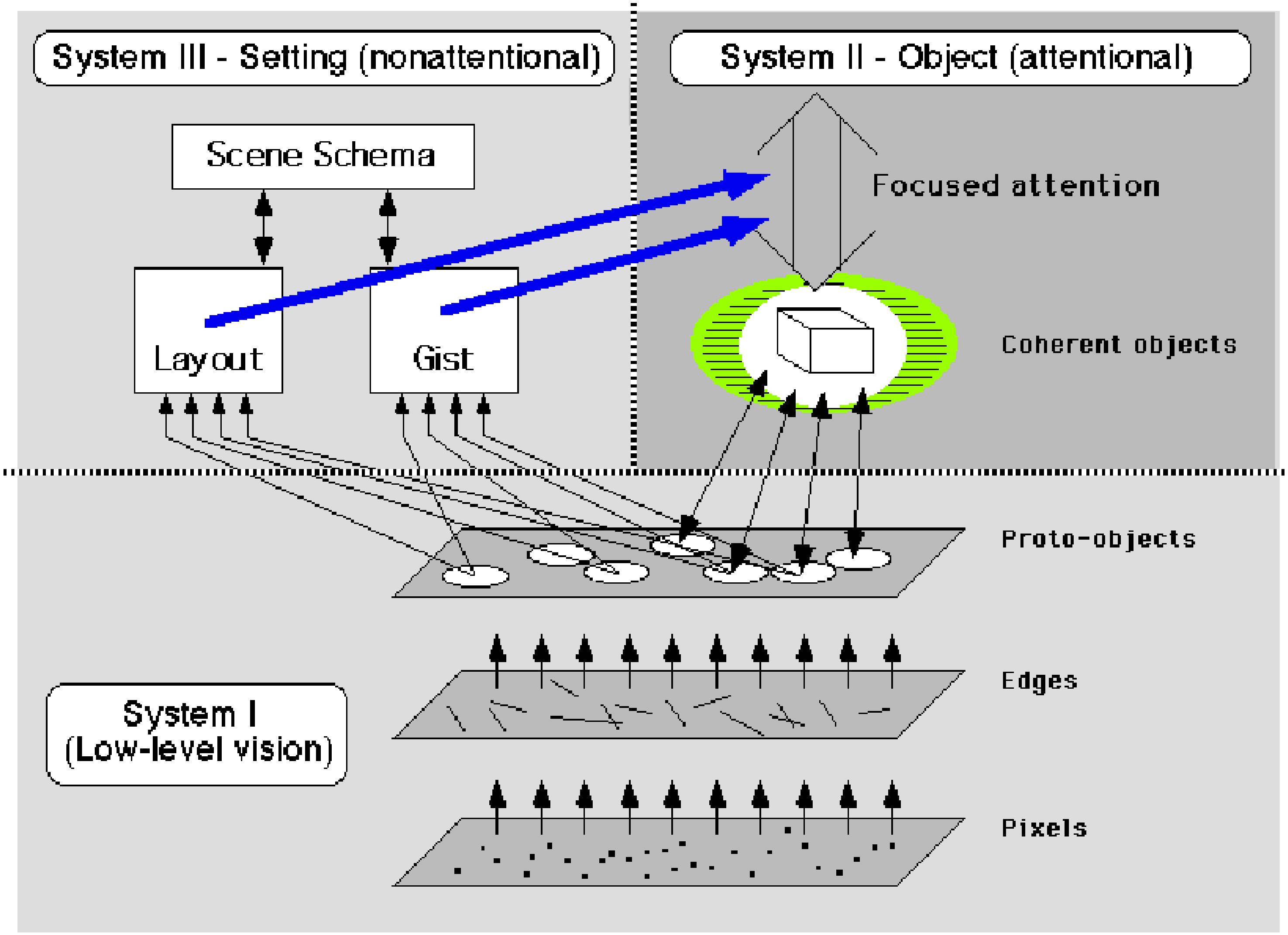


**Figure 4**

Figure 4. Triadic Architecture. It is suggested that the visual perception of scenes may be carried out via the interaction of three different systems. System I: Early-level processes produce volatile proto-objects rapidly and in parallel across the visual field. System II: Focused attention acts as a hand to "grab" these structures; as long as these structures are held, they form an individuated object with both temporal and spatial coherence. System III: Setting information—obtained via a nonattentional stream—guides the allocation of focused attention to various parts of the scene, and allows priorities to be given to the various possible objects.

# Syllabus Overview

*This is tentative and still open to suggestions!*

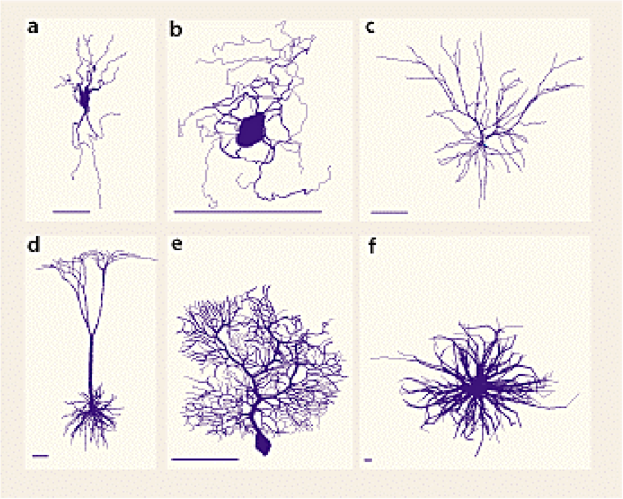
- ***Course Overview and Fundamentals of Neuroscience.***
- ***Neuroscience basics.***
- ***Experimental techniques in visual neuroscience.***
- ***Introduction to vision.***
- ***Low-level processing and feature detection.***
- ***Coding and representation.***
- ***Stereoscopic vision.***
- ***Perception of motion.***
- ***Color perception.***
- ***Visual illusions.***
- ***Visual attention.***
- ***Shape perception and scene analysis.***
- ***Object recognition.***
- ***Computer graphics, virtual reality and robotics.***

# Syllabus Overview

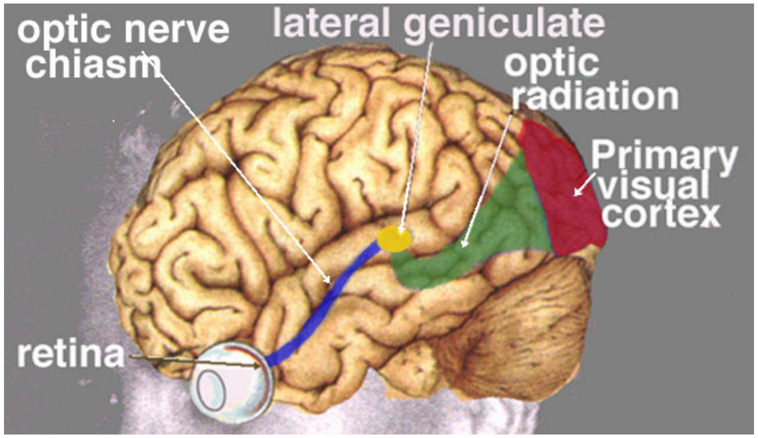
- ***Course Overview and Fundamentals of Neuroscience.***
  - why is vision hard while it seems so naturally easy?
  - why is half of our brain primarily concerned with vision?
  - Towards domestic robots: how far are we today?
  - What can be learned from the interplay between biology and computer science?

# Syllabus Overview

◆ ***Neuroscience basics.***

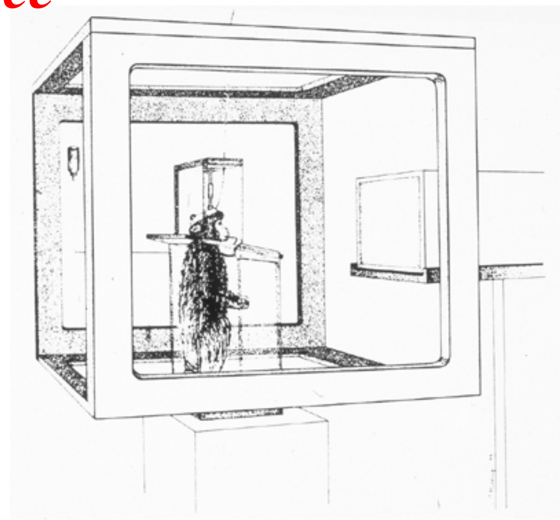


- The brain, its gross anatomy
- Major anatomical and functional areas
- The spinal cord and nerves
- Neurons, different types
- Support machinery and glial cells
- Action potentials
- Synapses and inter-neuron communication
- Neuromodulation
- Power consumption and supply
- Adaptability and learning

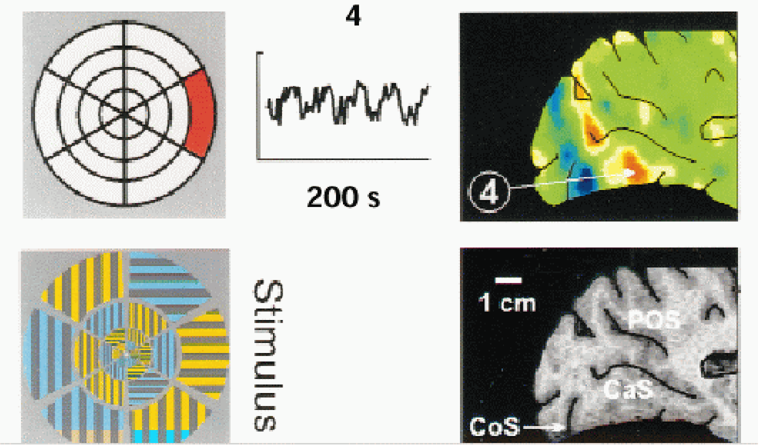


# Syllabus Overview

◆ ***Experimental techniques in visual neuroscience***



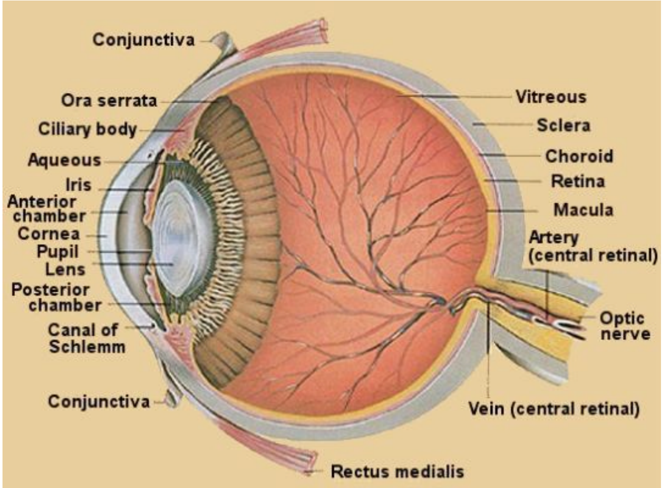
- Recording from neurons: electrophysiology
- Multi-unit recording using electrode arrays
- Stimulating while recording
- Anesthetized vs. awake animals
- Single-neuron recording in awake humans
- Probing the limits of vision: visual psychophysics
- Functional neuroimaging: Techniques
- Experimental design issues
- Optical imaging
- Transcranial magnetic stimulation

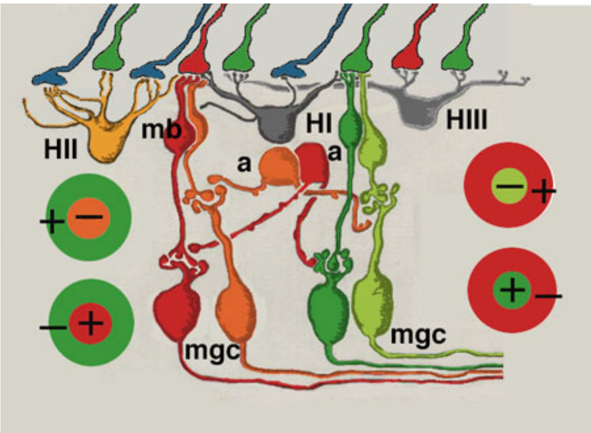


# Syllabus Overview

◆***Introduction to vision.***

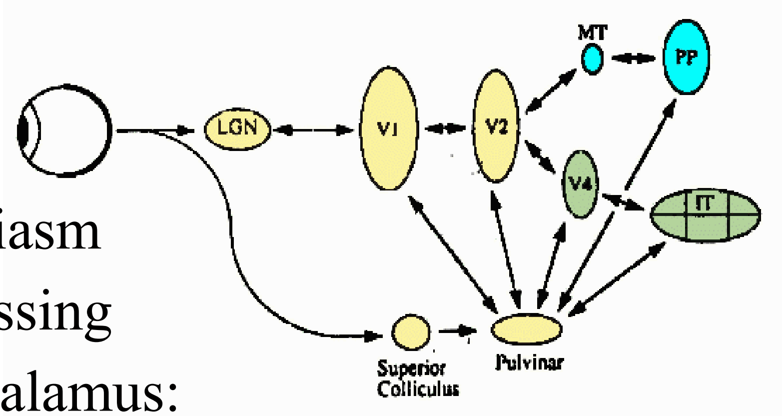
- Biological eyes compared to cameras and VLSI sensors
- Different types of eyes
- Optics
- Theoretical signal processing limits
- Introduction to Fourier transforms, applicability to vision
- The Sampling Theorem
- Experimental probing of theoretical limits
- Phototransduction
- Retinal organization
- Processing layers in the retina
- Adaptability and gain control.



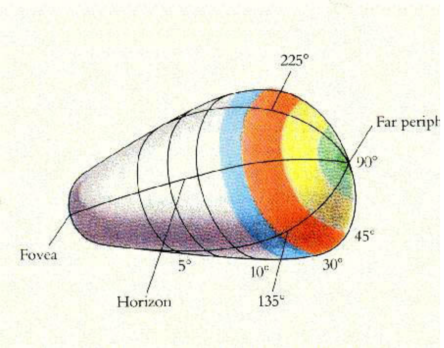


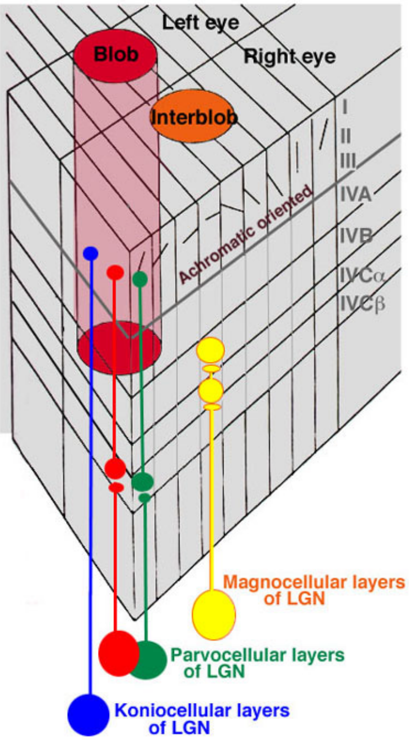
# *Syllabus Overview*

## ◆*More Introduction to Vision.*

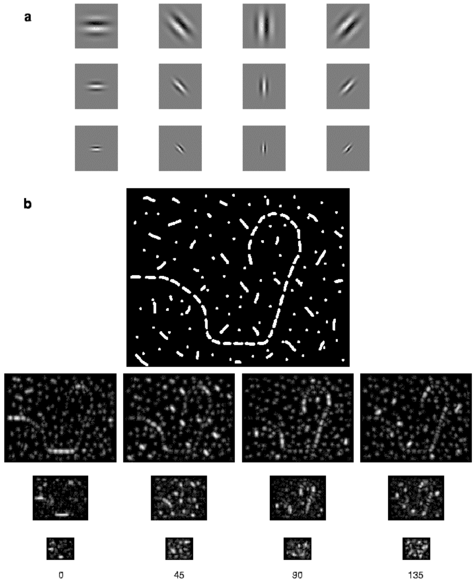


- Leaving the eyes: optic tracts, optic chiasm
- Associated pathology and signal processing
- The lateral geniculate nucleus of the thalamus: the first relay station to cortical processing
- Image processing in the LGN
- Notion of receptive field
- Primary visual cortex
- Cortical magnification
- Retinotopic mapping
- Overview of higher visual areas
- Visual processing pathways



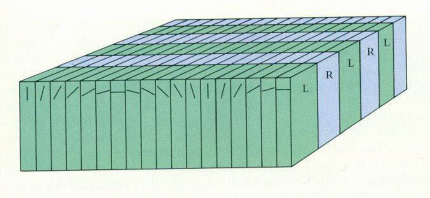


# Syllabus Overview



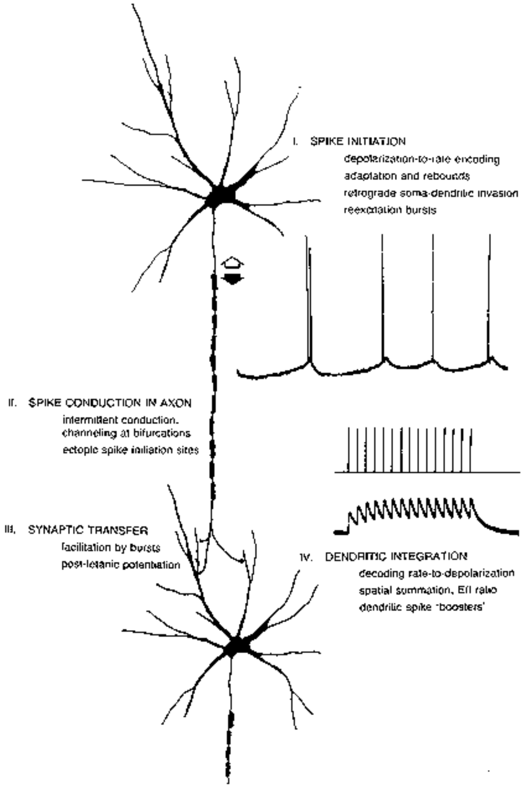
◆ ***Low-level processing and feature detection.***

- Basis transforms; wavelet transforms; jets
- Optimal coding
- Texture segregation
- Grouping
- Edges and boundaries; optimal filters for edge detection
- Random Markov fields and their relevance to biological vision
- Simple and complex cells
- Cortical gain control
- Columnar organization & short-range interactions
- Long-range horizontal connections and non-classical surround
- How can artificial vision systems benefit from these recent advances in neuroscience?



# *Syllabus Overview*

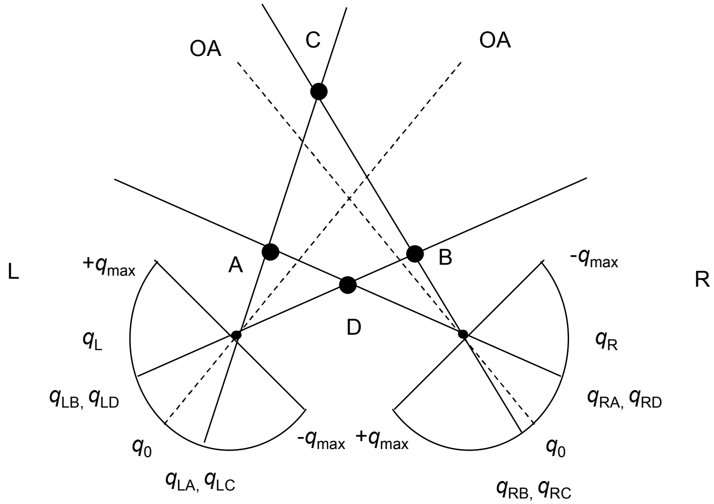
◆ ***Coding and representation.***



- Spiking vs. mean-rate neurons
- Spike timing analysis
- Autocorrelation and power spectrum
- Population coding; optimal readout
- Neurons as random variables
- Statistically efficient estimators
- Entropy & mutual information
- Principal component analysis (PCA)
- Independent component analysis (ICA)
- Application of these neuroscience analysis tools to engineering problems where data is inherently noisy (e.g., consumer-grade video cameras, VLSI implementations, computationally efficient approximate implementations).

# *Syllabus Overview*

## ◆ *Stereoscopic vision*

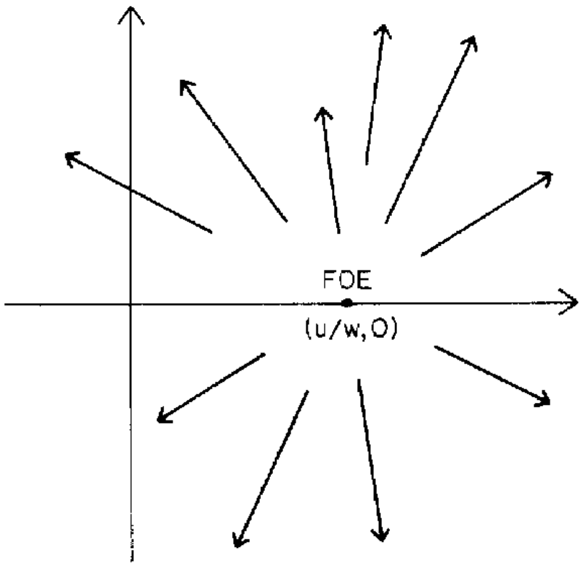


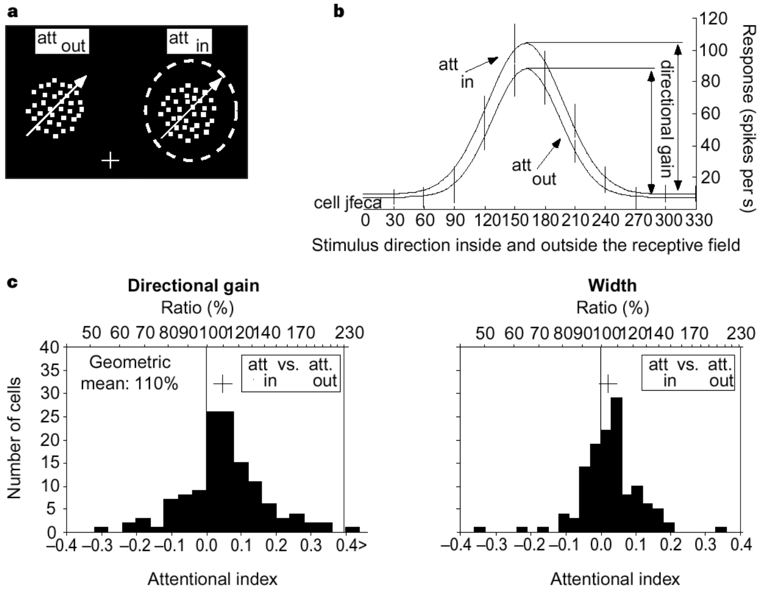
- Challenges in stereo-vision
- The Correspondence Problem
- Inferring depth from several 2D views
- Several cameras vs. one moving camera
- Brief overview of epipolar geometry and depth computation
- Neurons tuned for disparity
- Size constancy
- Do we segment objects first and then match their projections on both eyes to infer distance?
- Random-dot stereograms ("magic eye"): how do they work and what do they tell us about the brain?

# *Syllabus Overview*

## ◆*Perception of motion*

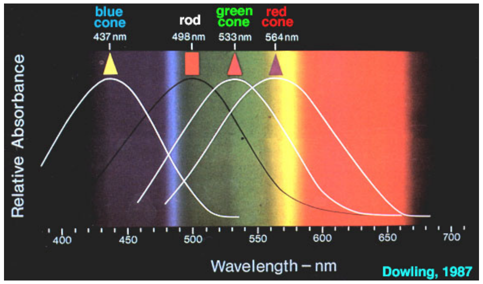
- Optic flow
- Segmentation and regularization
- Efficient algorithms
- Robust algorithms
- The spatio-temporal energy model
- Computing the focus of expansion and time-to-contact
- Motion-selective neurons in cortical areas MT and MST



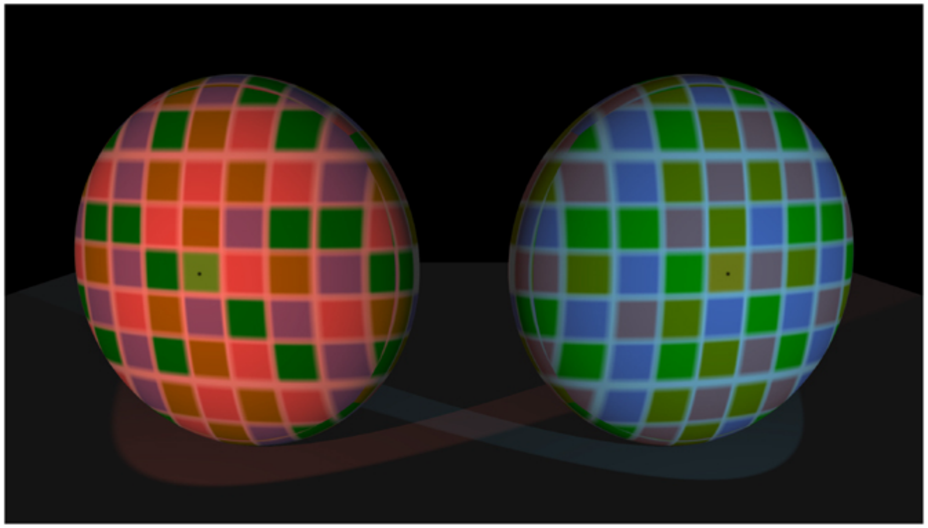


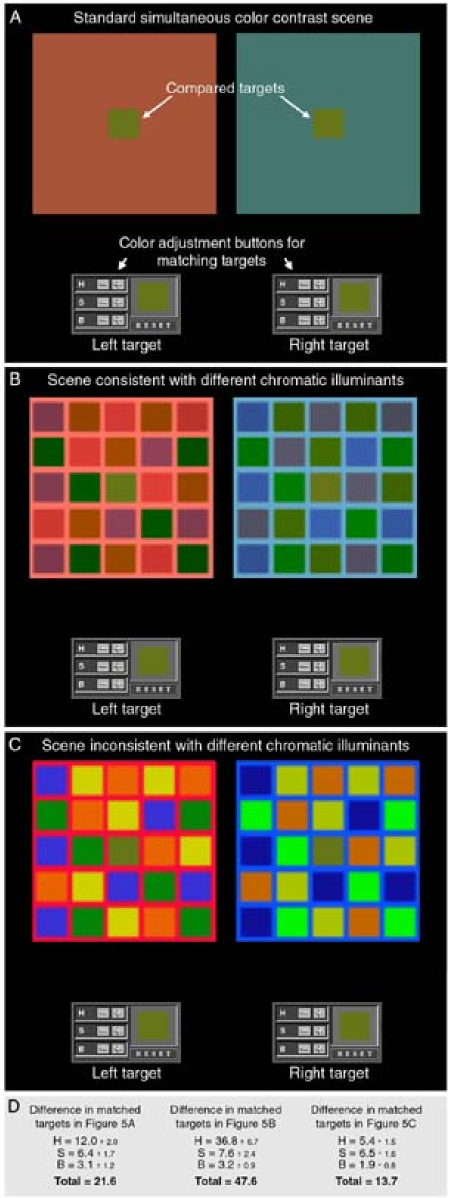
# *Syllabus Overview*

## ◆ *Color perception*



- Color-sensitive photoreceptors (cones)
- Visible wavelengths and light absorption
- The Color Constancy problem: how can we build stable percepts of colors despite variations in illumination, shadows, etc

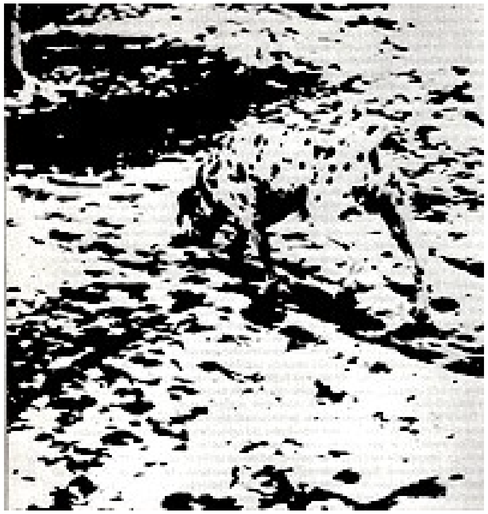


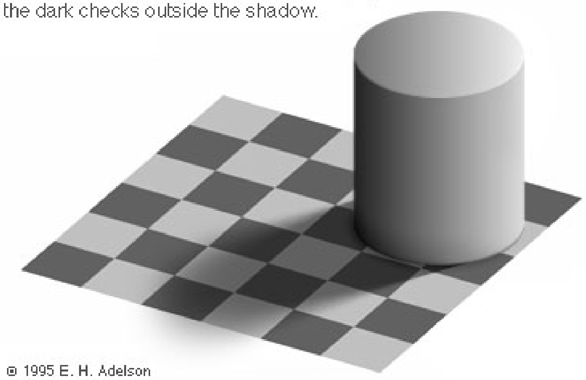


# Syllabus Overview

## ◆ *Visual illusions*

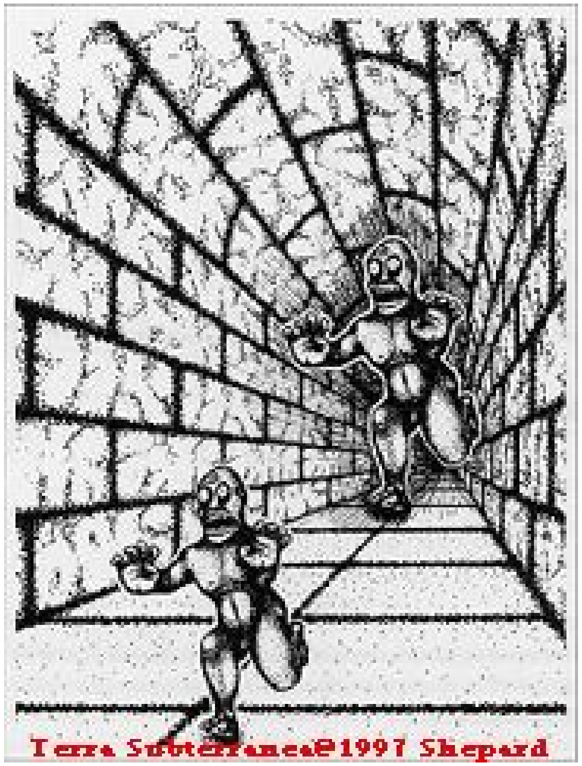
- What can illusions teach us about the brain?
- Examples of illusions
- Which subsystems studied so far do various illusions tell us about?
- What computational explanations can we find for many of these illusions?





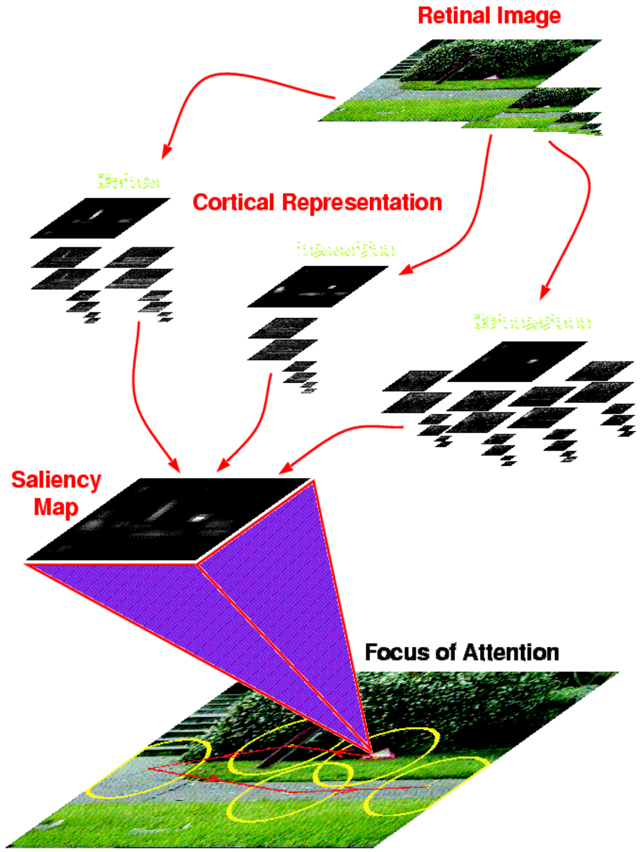
The light check in the shadow is the same gray as the dark checks outside the shadow.

© 1995 E. H. Adelson



# Syllabus Overview

## ◆ *Visual attention*

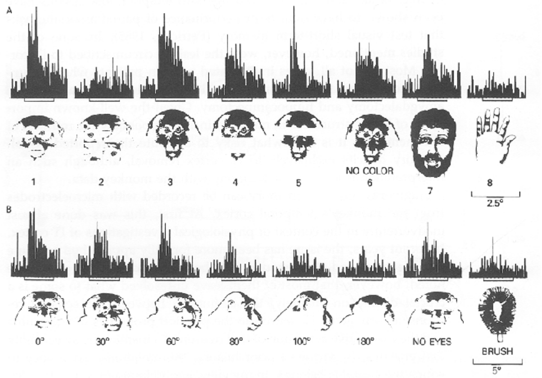


- Several kinds of attention
- Bottom-up and top-down
- Overt and covert
- Attentional modulation
- How can understanding attention contribute to computer vision systems?
- Biological models of attention
- Change blindness
- Attention and awareness
- Engineering applications of attention: image compression, target detection, evaluation of advertising, etc...

# Syllabus Overview

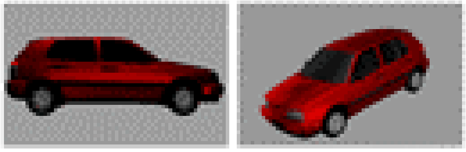
## ◆*Shape perception and scene analysis*

"Face cells"



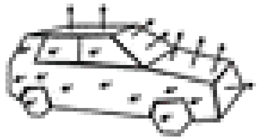
- Shape-selective neurons in cortex
- Coding: one neuron per object or population codes?
- Biologically-inspired algorithms for shape perception
- The "gist" of a scene: how can we get it in 100ms or less?
- Visual memory: how much do we remember of what we have seen?
- The world as an outside memory and our eyes as a lookup tool

1)

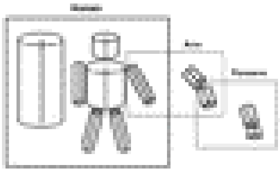


2)

3)

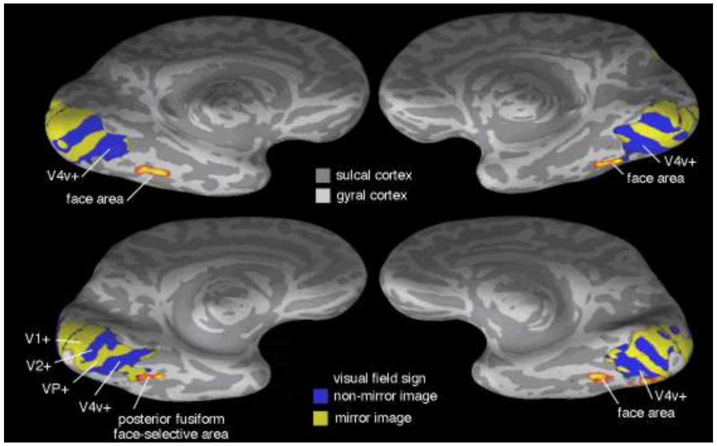


4)



# *Syllabus Overview*

## ◆ *Object recognition*



- The basic issues
- Translation and rotation invariance
- Neural models that do it
- 3D viewpoint invariance (data and models)
- Classical computer vision approaches: template matching and matched filters; wavelet transforms; correlation; etc.
- Examples: face recognition.
- More examples of biologically-inspired object recognition systems which work remarkably well