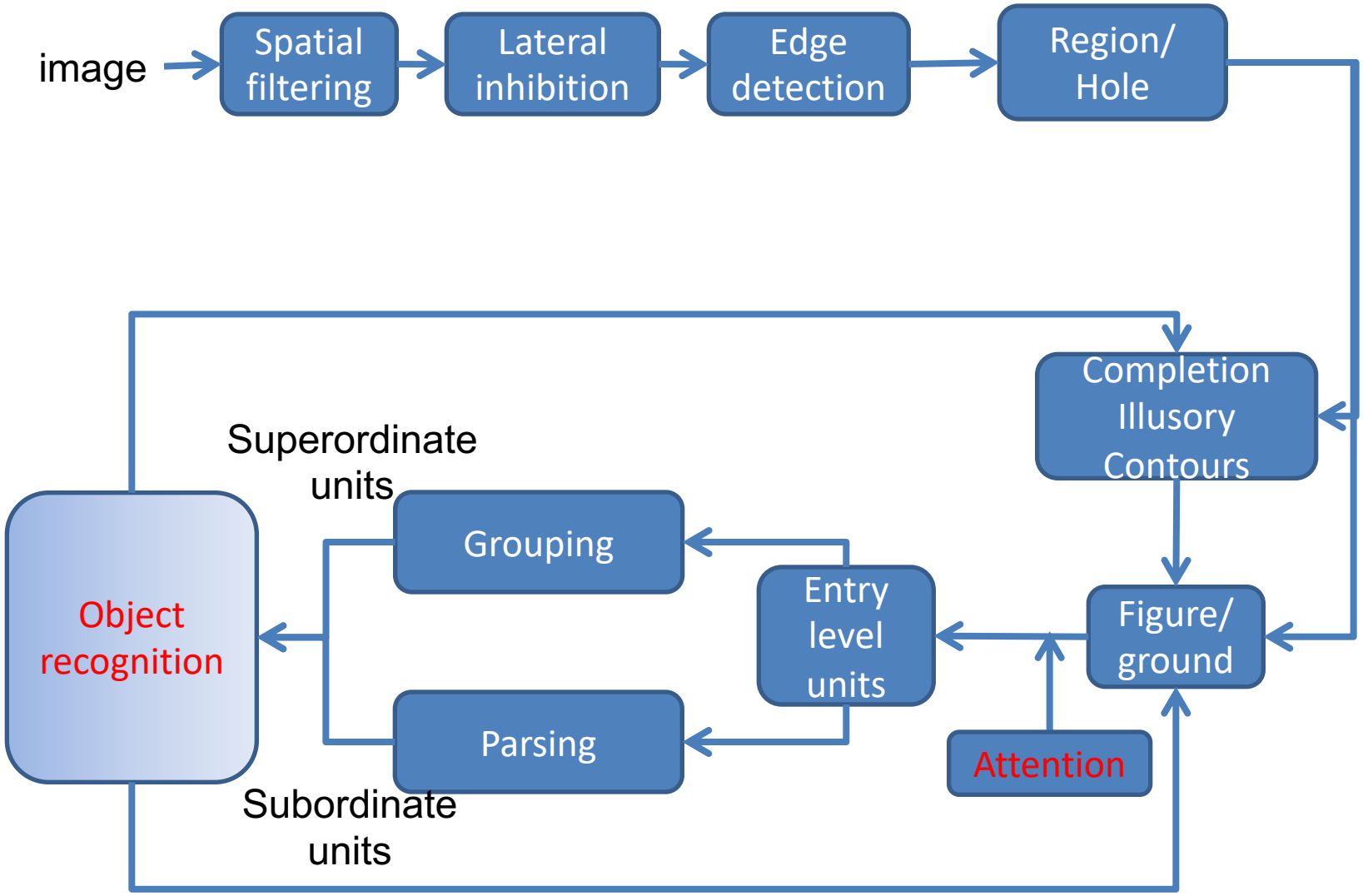


image
Spatial filtering
Lateral inhibition
Edge detection
Region/ Hole
Completion Illusory Contours
Superordinate units
Grouping
Object recognition
Entry level units
Figure/ ground
Parsing
Attention
Subordinate units

Object Perception

Wu Xihong

Peking University

The Organized Visual System

- Organization is especially important in the visual system because of the tasks the visual system faces.
  - One task is to process information about various characteristics, or features, of objects, such as size, shape, orientation, color, movement, and location in space.
  - Once this information is processed, how are all of an object's features combined?
- Organization plays a central role in achieving the tasks of both processing specific information and combining information to create coherent perceptions.

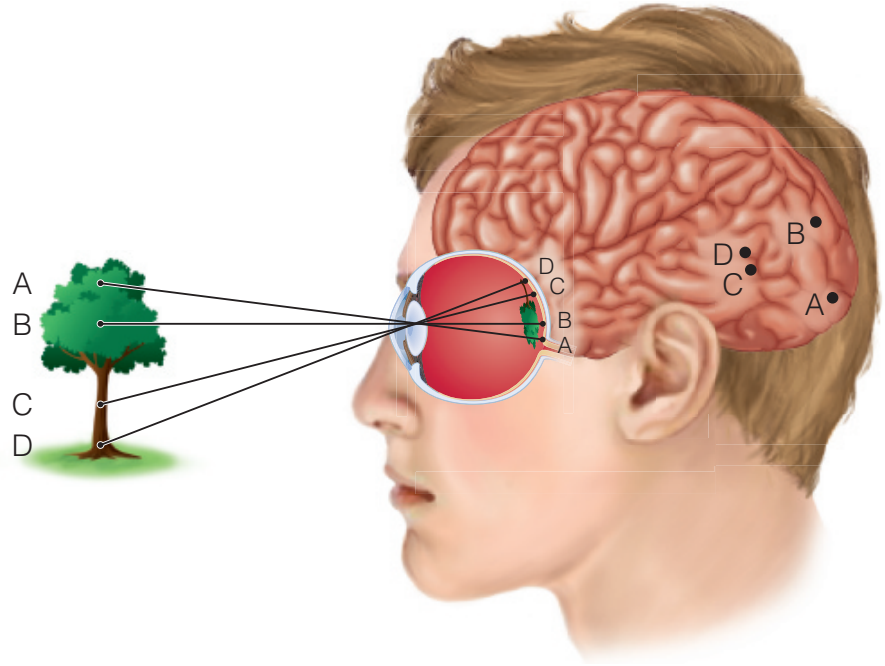
1 Spatial Organization

- Spatial organization refers to the way stimuli at specific locations in the environment are represented by activity at specific locations in the nervous system.
  - There are objects to the left and right, high and low. This organization in visual space then becomes transformed into organization in the eye, when an image of the scene is created on the retina.
  - It is easy to appreciate spatial organization at the level of the retinal image because this image is essentially a picture of the scene.

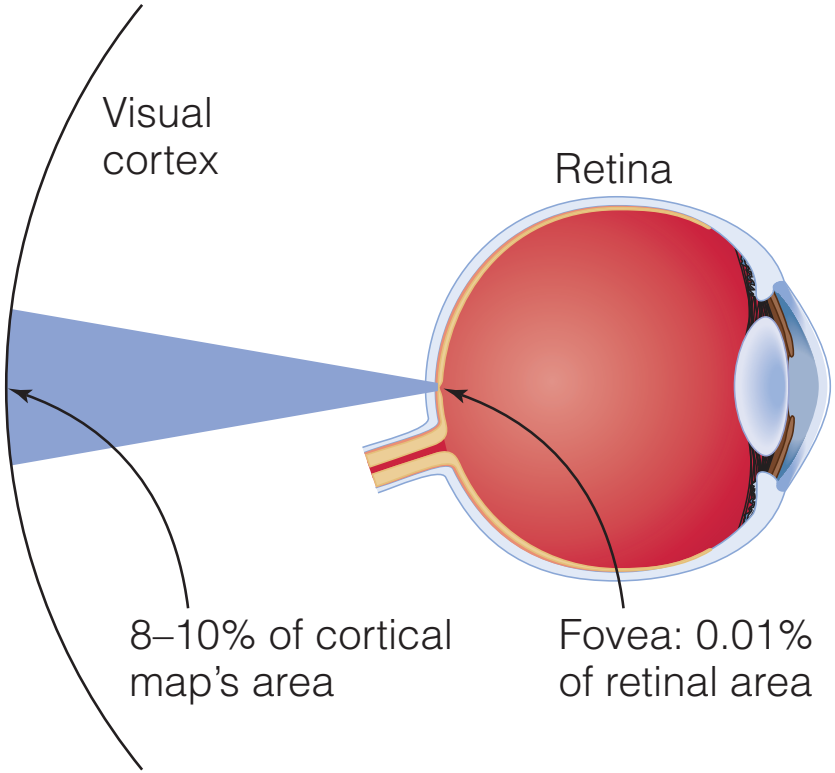
1.1 The Electronic Map on V1

- How points in the retinal image are represented *spatially* on the striate cortex.
  - Stimulating various places on the retina and determining where neurons fire in the cortex.
  - We can also reverse the process by recording from a neuron in the cortex and determining the location of its receptive field on the retina

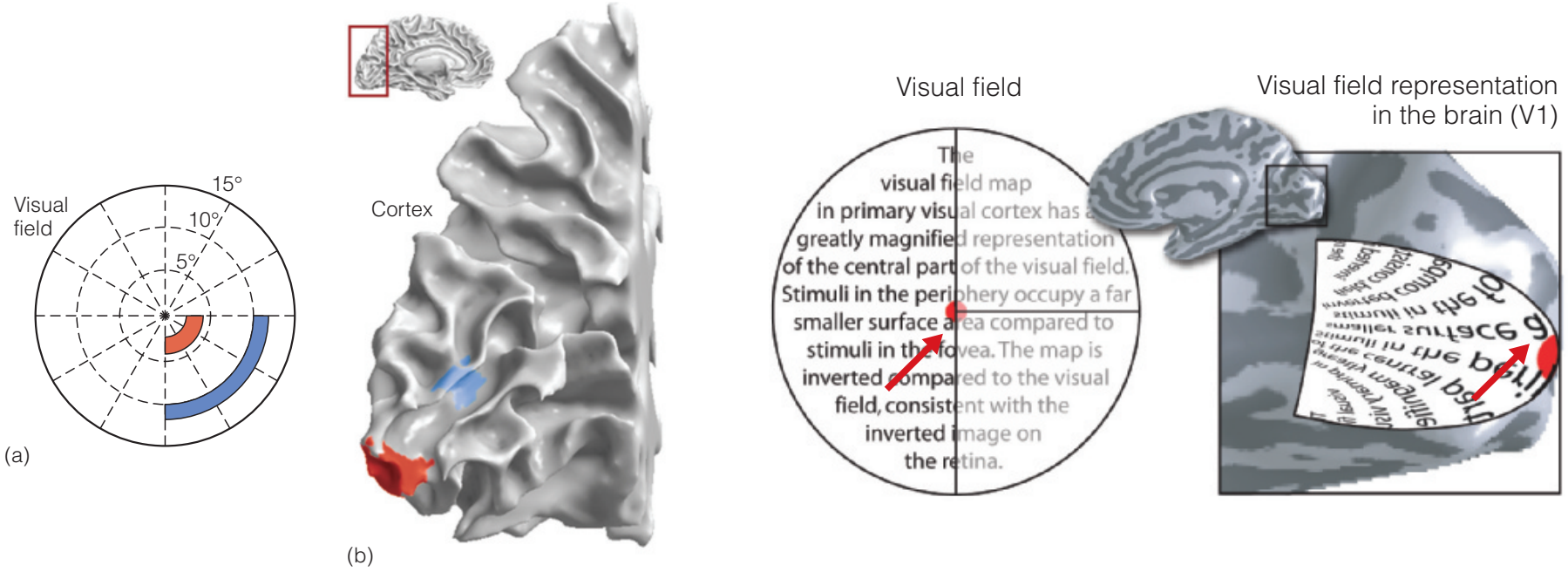
Retinotopic map: means that two points that are close together on an object and on the retina will activate neurons that are close together in the brain.



- Cortical Magnification: The map on the cortex is distorted, with more space being allotted to locations near the fovea than to locations in the peripheral retina.



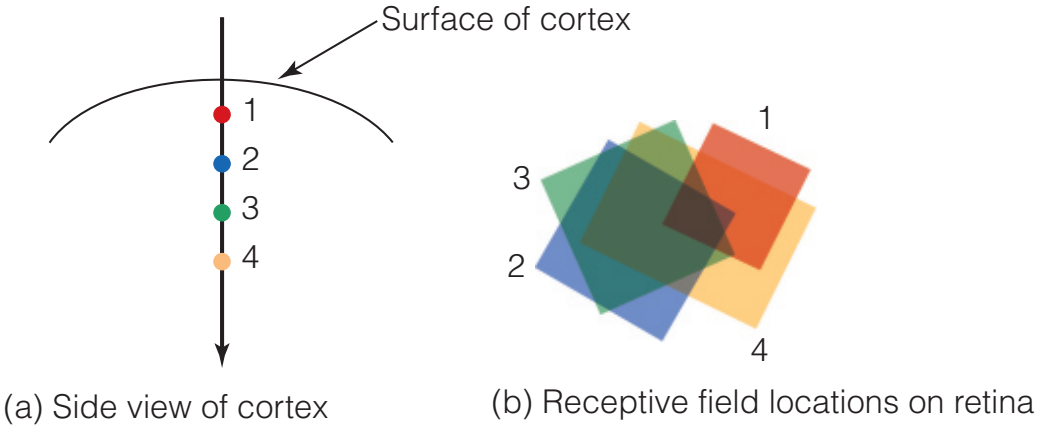
The fovea accounts for only 0.01 percent of the retina's area, signals from the fovea account for 8 to 10 percent of the retinotopic map on the cortex.

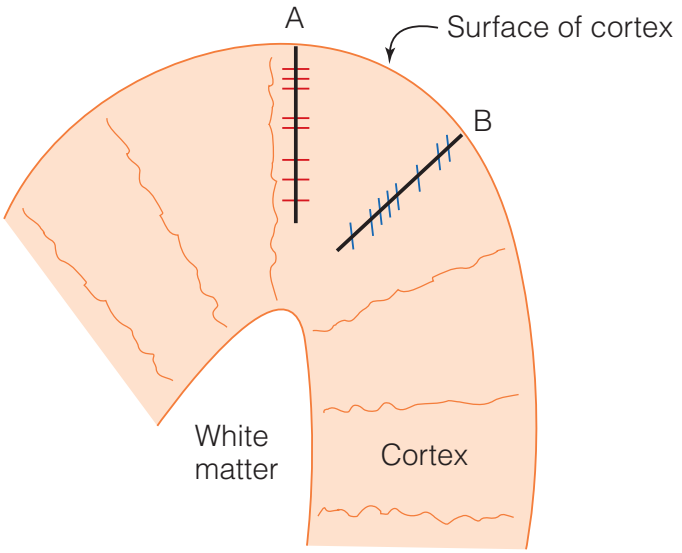


The fact that more space on the cortex translates into better detail vision rather than larger size is an example of the fact that *what we perceive doesn't exactly match the "picture" in the brain.*

1.2 The Cortex Is Organized in Columns

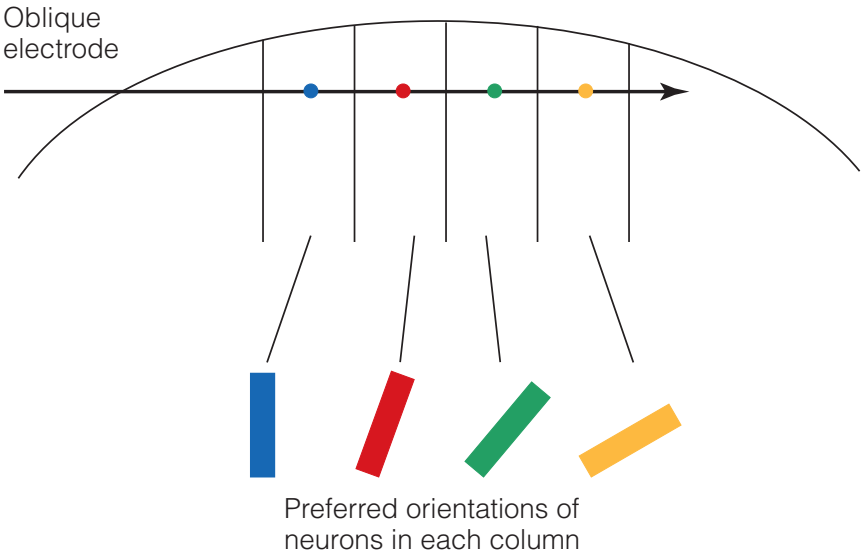
- What is happening below the cortex surface.

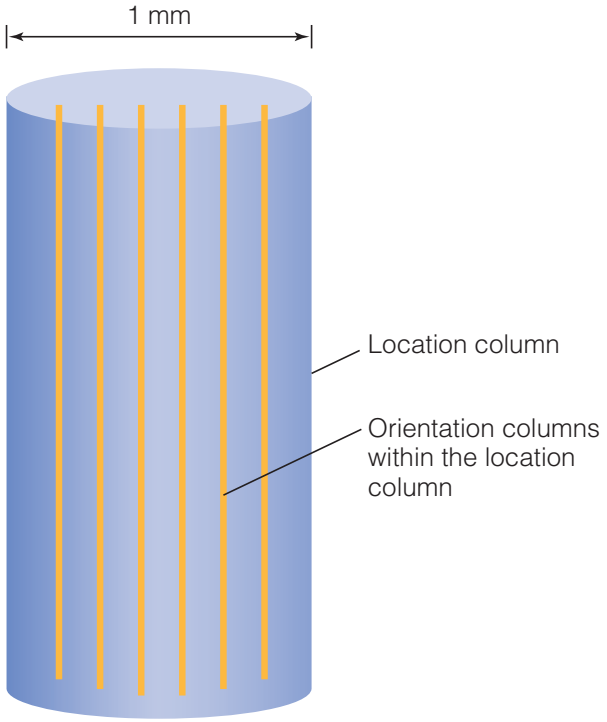




All of the neurons within a location column have their receptive fields at the same location on the retina.

The cortex is organized into orientation columns, with each column containing cells that respond best to a particular orientation.





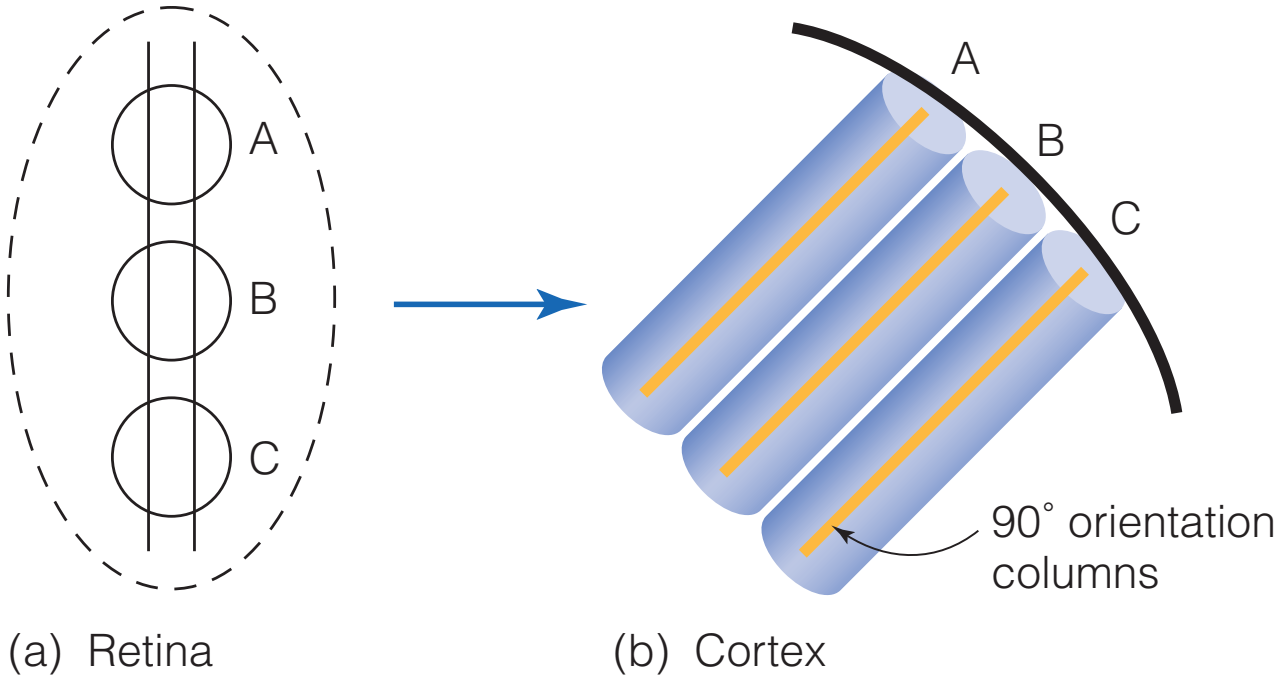
The neurons' preferred orientations changed in an orderly fashion. 1-mm dimension is the size of one location column.

Hypercolumn： a location column that contains the full range of orientation columns.

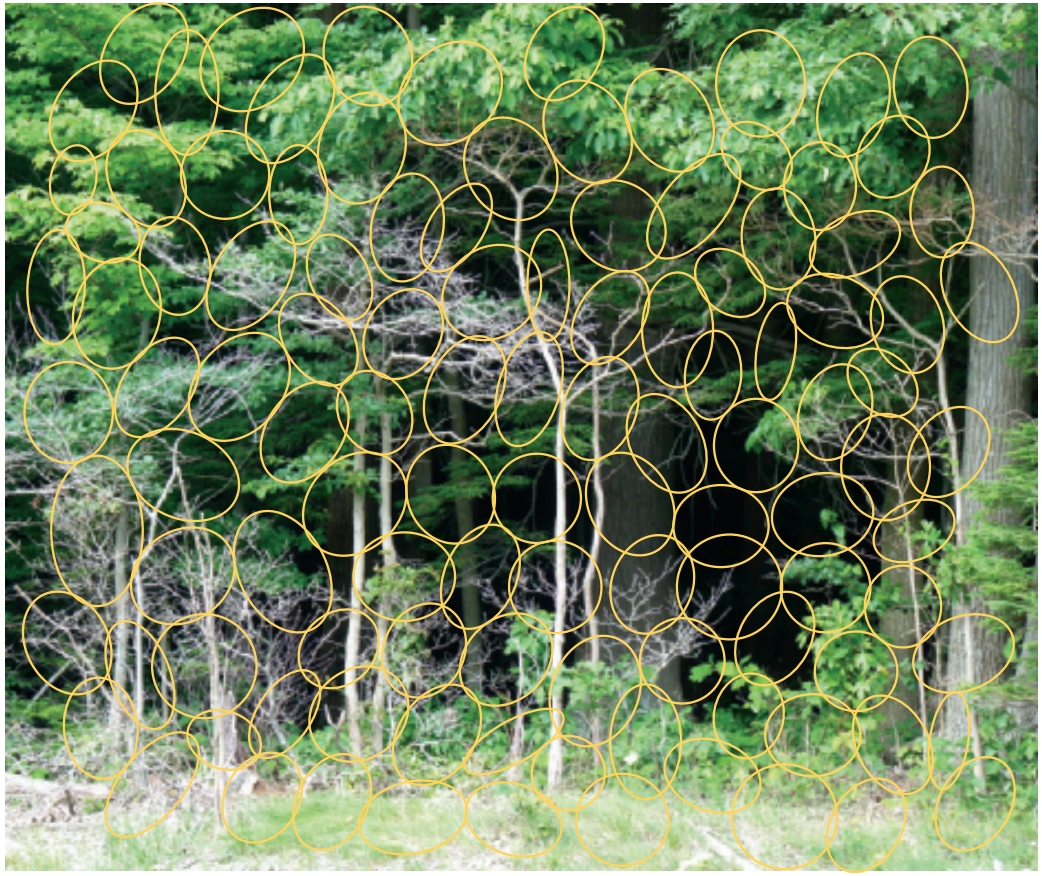
1.3 How Do Feature Detectors Respond to a Scene?

(a) (b)

Each of the circles represents the area served by a location column.



The representation of the tree in the visual cortex is contained in the *firings* of neurons in separate cortical columns.



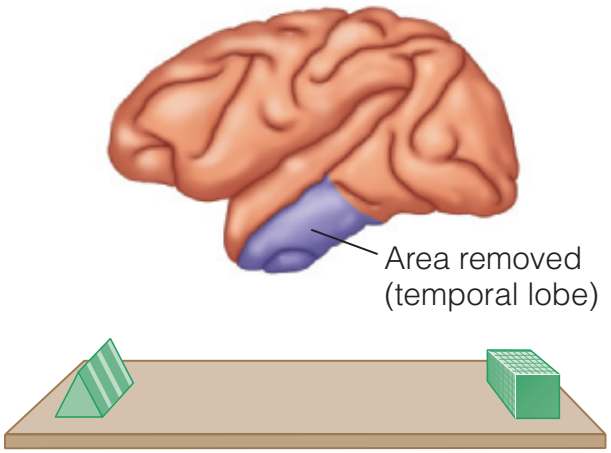
This representation in the striate cortex is only the *first step* in representing the tree. As we will now see, signals from the striate cortex travel to a number of other places in the cortex for further processing.

2 Streams: Pathways for What, Where, and How

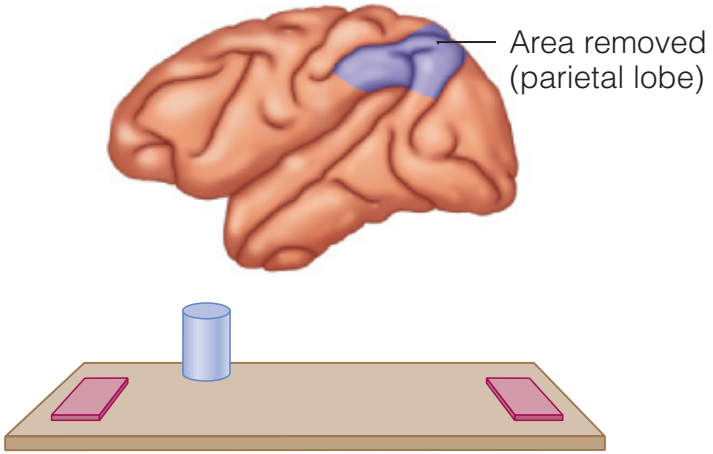
- Investigating how visual stimulation causes activity in areas far beyond the striate cortex.
  - There are pathways, or “streams”, that transmit information from the striate cortex to other areas in the brain.
  - Distinguished two streams that served different functions.

2.1 Streams for Information About **What** and **Where**

- Ungerleider and Mishkin (1982) used a technique called ***ablation***, which refers to the destruction or removal of tissue in the nervous system. Two tasks:
  - an object discrimination problem
  - a landmark discrimination problem.



(a) Object discrimination



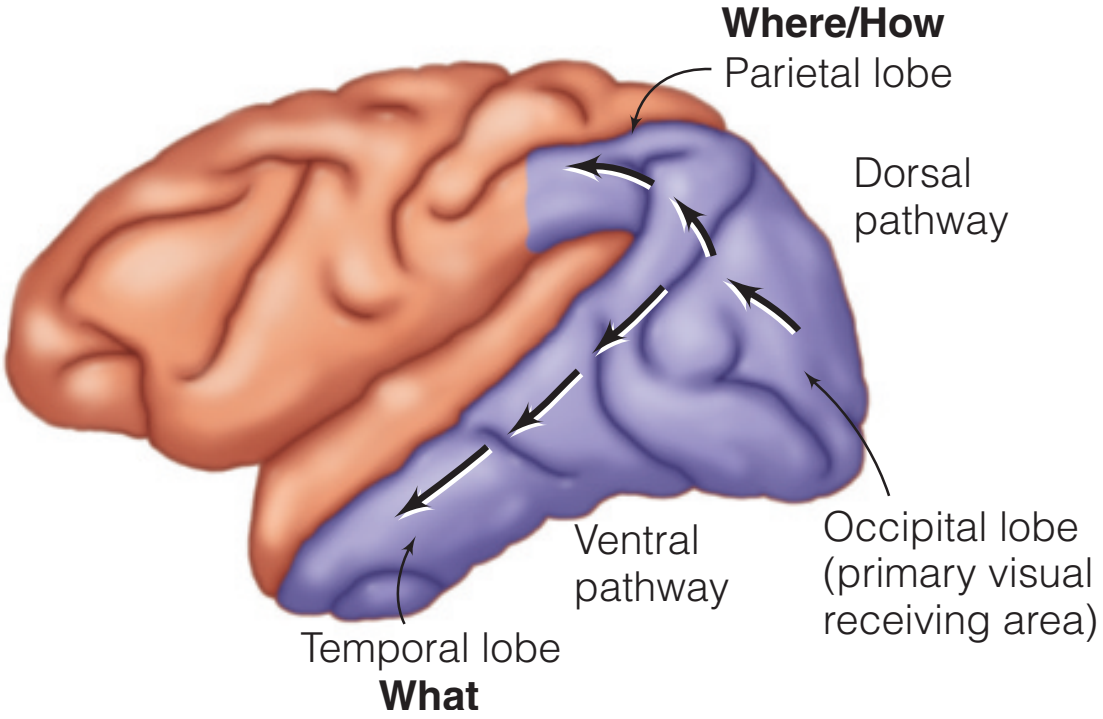
(b) Landmark discrimination

Double Dissociation Experiment

	NAMING OBJECTS	DETERMINING OBJECT'S LOCATION
(a) ALICE: Temporal lobe damage (ventral stream)	NO	YES
(b) BERT: Parietal lobe damage (dorsal stream)	YES	NO

A double dissociation and enable us to conclude that recognizing objects and locating objects operate independently of each other.

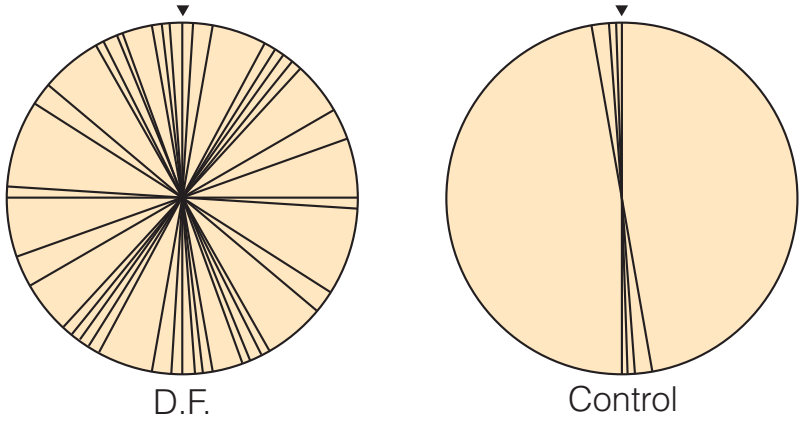
- **Ventral pathway(*What)*** : the pathway that reaches the temporal lobe is responsible for determining an object's identity.
- **Dorsal pathway(*Where*)**: the pathway that leads to the parietal lobe is responsible for determining an object's location.



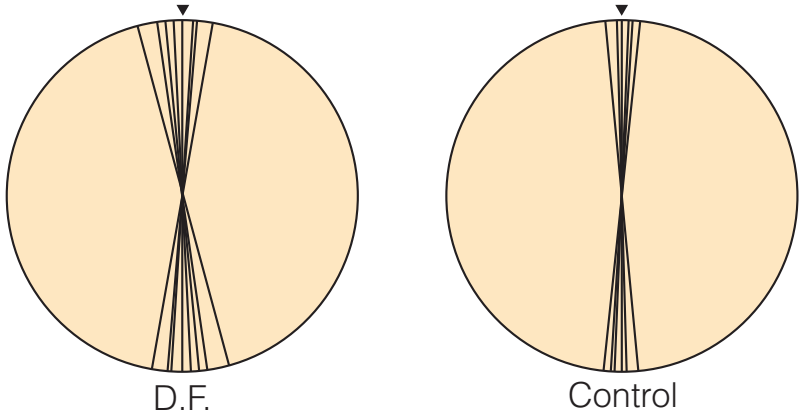
- The cortical ventral and dorsal streams can actually be traced back to the retina and LGN.
  - The properties of the ventral and dorsal streams are established by two different types of ganglion cells in the retina, which transmit signals to different layers of the LGN.
- The ventral and dorsal pathways serve different functions, note that
  - The pathways are not totally separated, but have ***connections*** between them;
  - Signals flow not only “up” the pathway toward the parietal and temporal lobes, but “***back***” as well.

2.2 Streams for Information About **What** and **How**

- Taking an action would involve knowing the location of the object, consistent with the idea of *where*, but it goes beyond *where* to involve a physical interaction with the object.
- According to this idea, the dorsal stream provides information about ***how*** *(action)* to direct action with regard to a stimulus.



(a) Static orientation matching



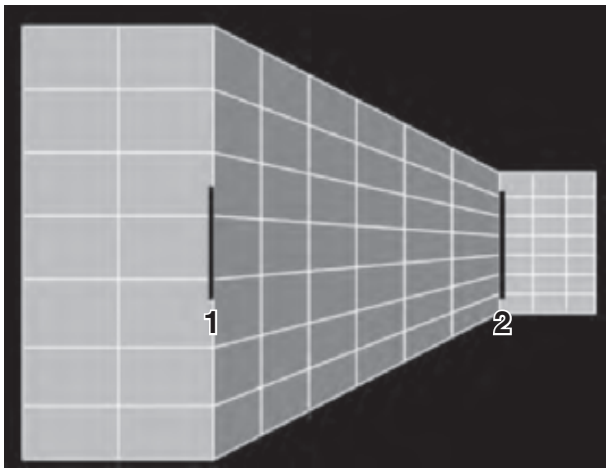
(b) Active “posting”

Performance of D.F. and a person without brain damage on two tasks:

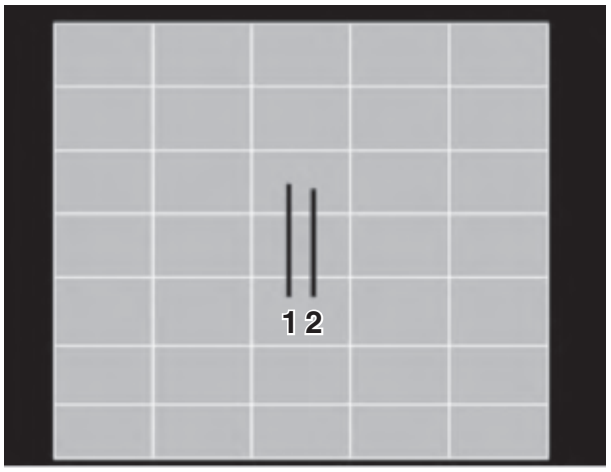
(a) judging the orientation of a slot;

(b) placing a card through the slot;

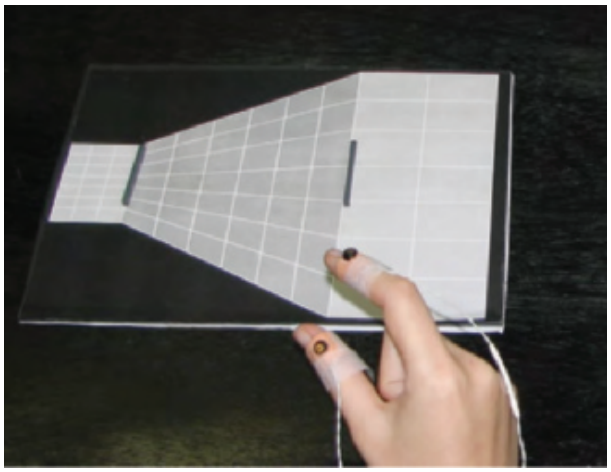
The Behavior of People Without Brain Damage



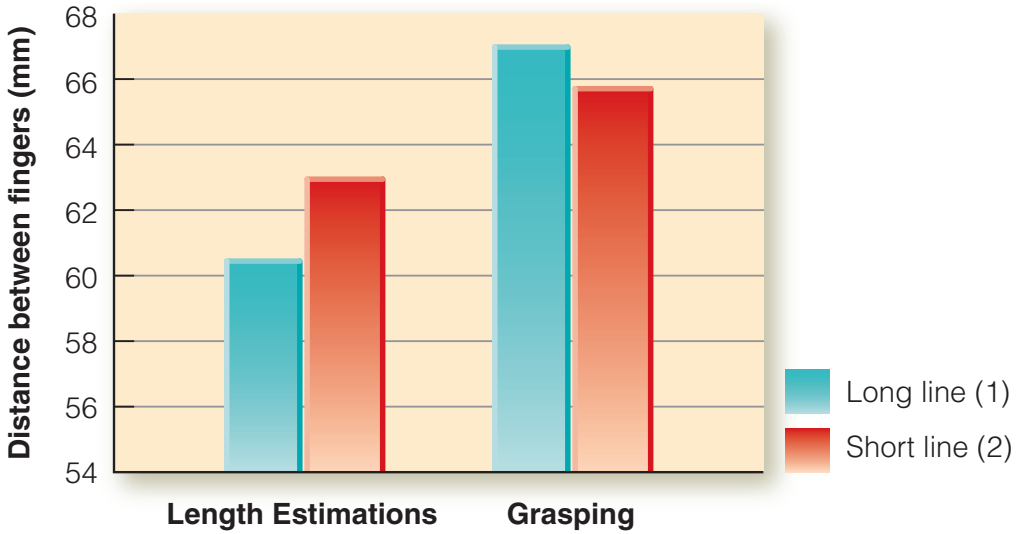
(a)



(b)

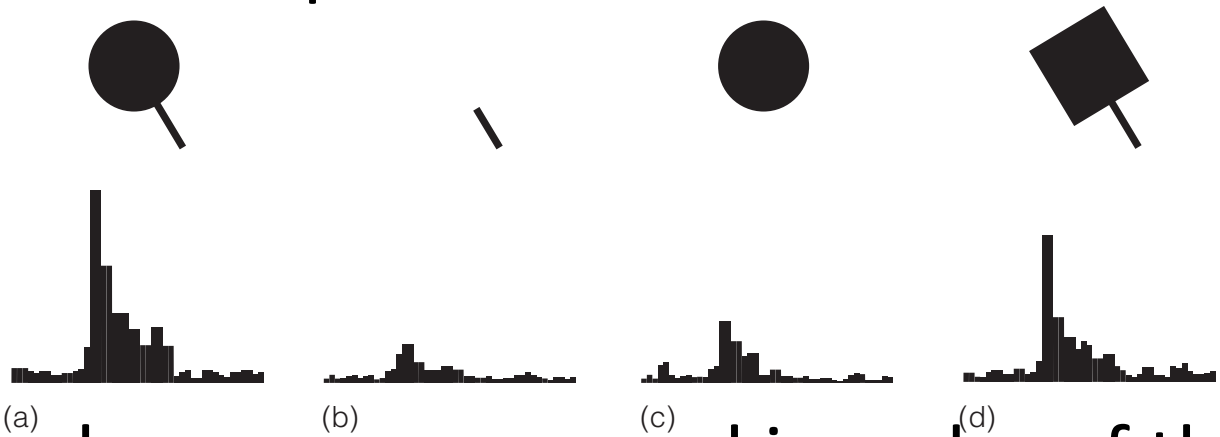


(c)

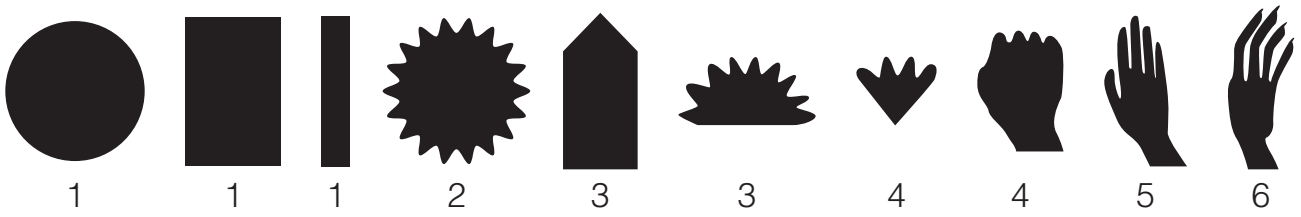


3. Modularity: Structures for Faces, Places, and Bodies

- Neurons in the temporal cortex that responded best to complex stimuli.

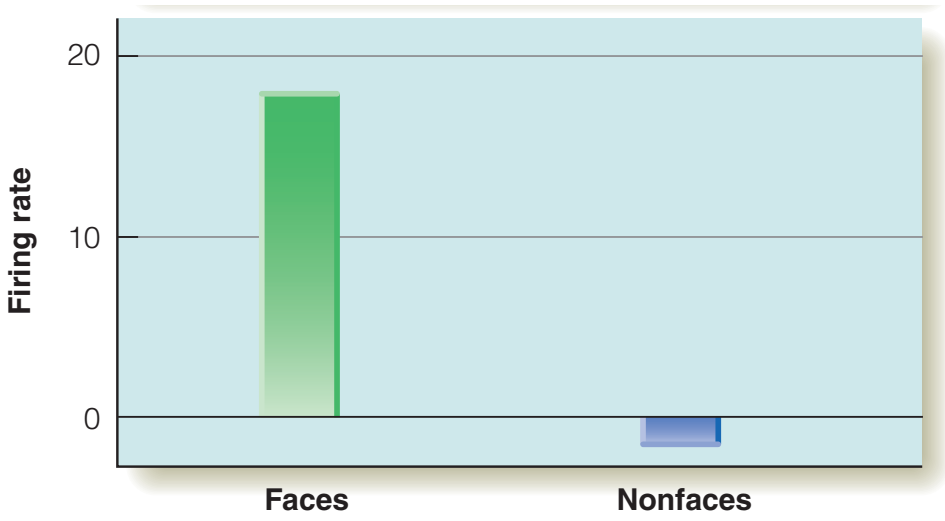


- The shapes are arranged in order of their ability to cause the neuron to fire.



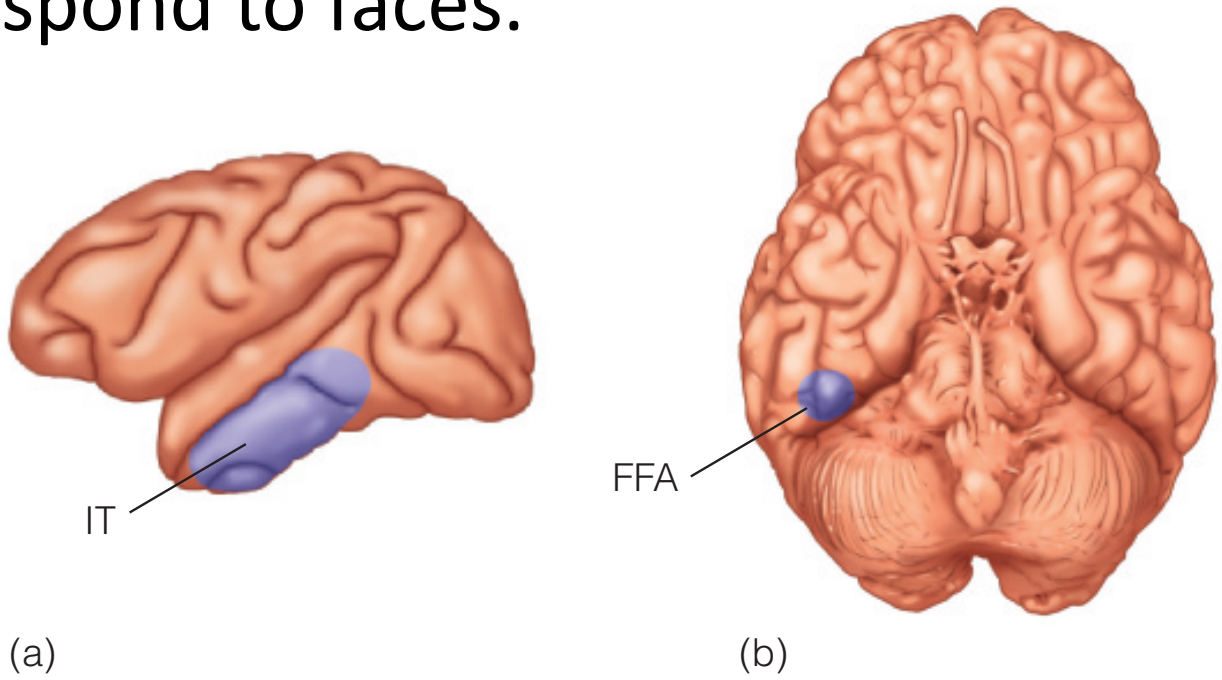
3.1 Face Neurons in the Monkey's IT Cortex

- Neurons in the monkey's inferotemporal (IT) cortex responded best to faces, when presented pictures of faces and pictures of nonface stimuli.

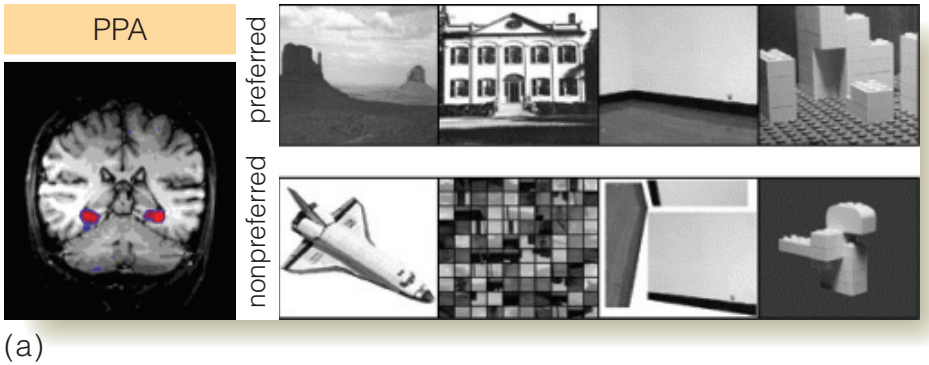


3.2 Areas for Faces, Places, and Bodies in the Human Brain

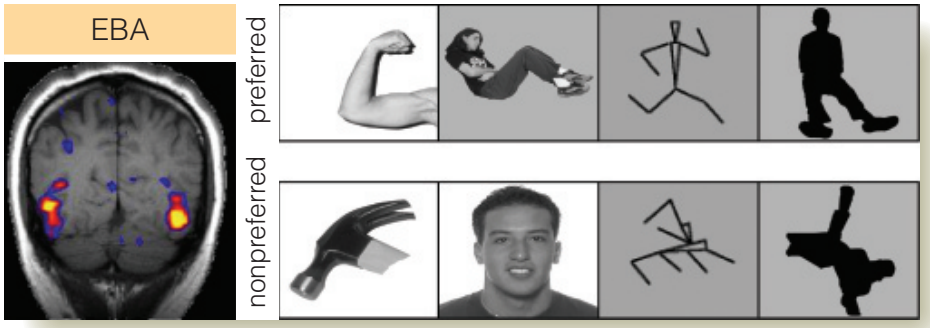
- The fusiform face area (FFA), which is located in **the fusiform gyrus** on the underside of the brain directly below the IT cortex, specialized to respond to faces.



- The parahippocampal place area (**PPA**) is activated by pictures depicting indoor and outdoor scenes.
  - what is important for this area is information about spatial layout.
- The extrastriate body area (**EBA**), is activated by pictures of bodies and parts of bodies.

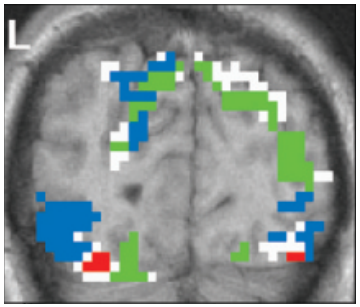


(a)



(b)

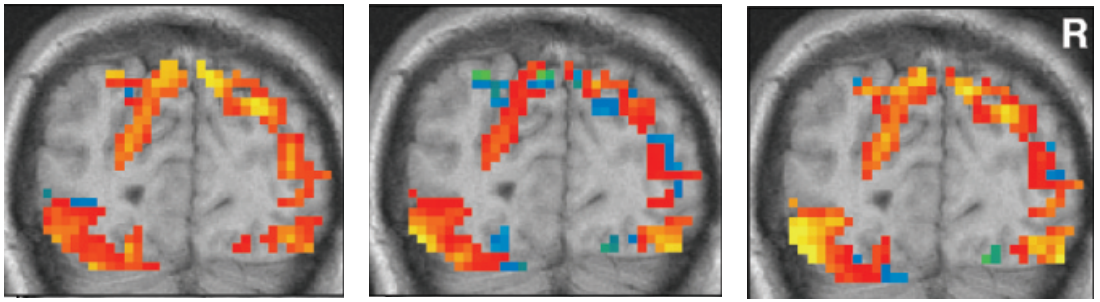
- Our perception of faces, landmarks, and people's bodies depends on specifically tuned neurons in areas such as the FFA, PPA, and EBA.
- But it is also important to recognize that even though stimuli like faces and buildings activate specific areas of the brain, *these stimuli also activate other areas of the brain as well*.



Segregation by category

Maximal Respose to:

Houses
Faces
Chairs
No difference



Houses Faces Chairs

Response magnitude

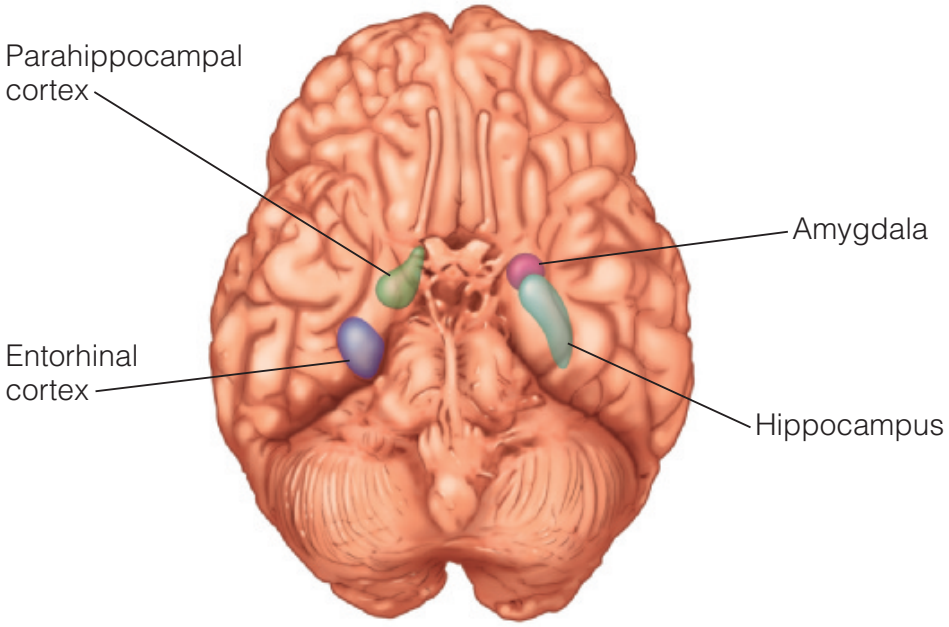
Percent Activation

−1 0 +1 +2

- The visual system is organized ***spatially***.
  - The spatial map is retinotopic, which means that points on the LGN or cortex correspond to specific points on the retina or in a scene.
  - Spatial organization becomes *weaker* as we move to higher cortical areas, because in areas such as IT cortex, neurons have very large receptive fields that extend over large areas of the retina and visual field.
- The visual system is organized ***functionally***.
  - Different streams for *what* and *where*/*how* and with specific cortical areas that are rich in neurons that respond to specific types of stimuli such as *faces*, *places*, and *bodies*.

4 Where Vision Meets **Memory**

- Some of the signals leaving the IT cortex reach structures in the medial temporal lobe (MTL), such as the parahippocampal cortex, the entorhinal cortex, and the hippocampus.



- Neurons in the hippocampus that respond to faces of specific people, to specific structures.

Halle Berry

- The possible role of these neurons in memory is supported by the way they respond to many different *views* of the stimulus, different *modes* of depiction, and even *words* signifying the stimulus.
- These neurons are not responding to visual features of the pictures, but to *concepts*.

5 Connecting Neural Activity and Object Perception

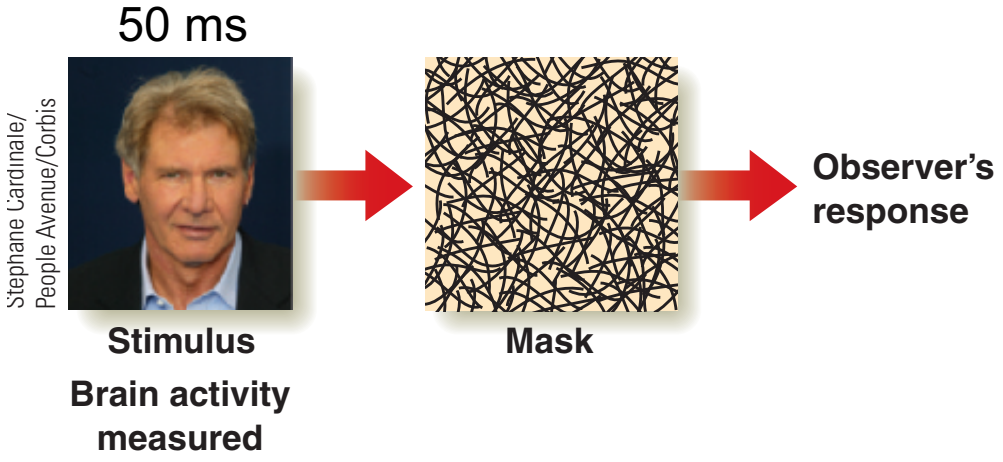
- So far we have focused on how perception is determined by aspects of *stimuli*.
- Now it is time to consider the relationship between physiological processes and the perception of objects.

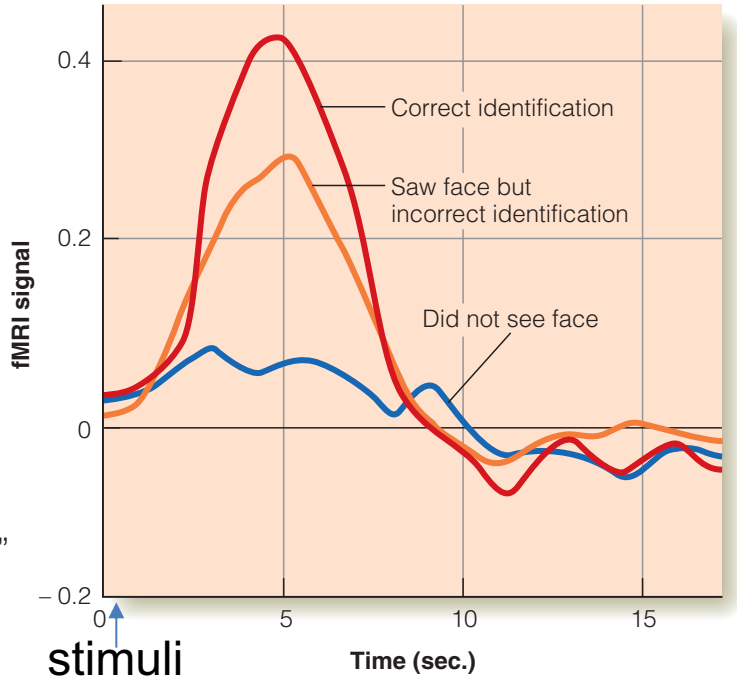
Two major themes on the physiology of vision

- Describing the types of stimuli that cause neurons at different levels of the visual system to respond;
  - Show how neurons at higher levels of the visual system respond to more and more complex stimuli.
- Describing how neurons in the visual system are organized.
  - Do electrical signals that represent objects at different places in a scene go to different places in the brain?
  - Are there separate brain areas that determine our perception of different qualities?

5.1 Brain Activity and Identifying a Picture

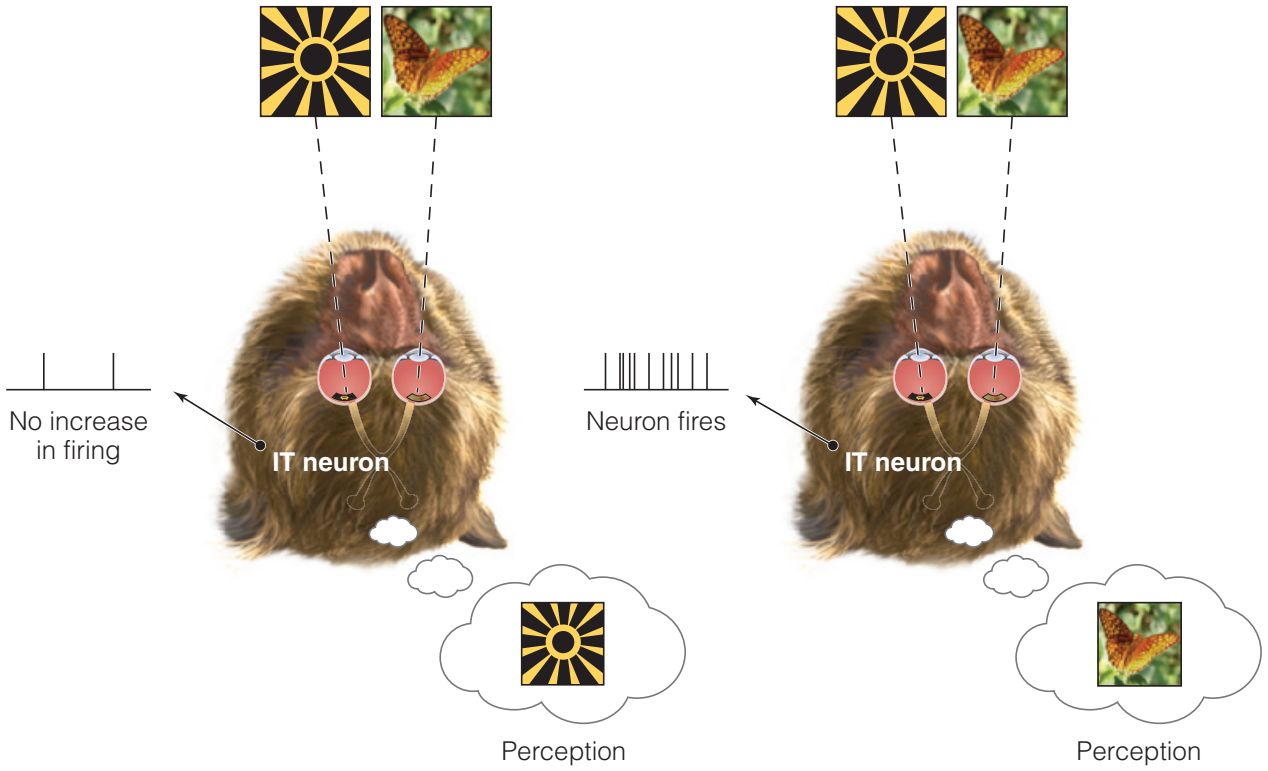
- Relationship between the brain activation that occurs when looking at an object and a person's ability to identify the object.

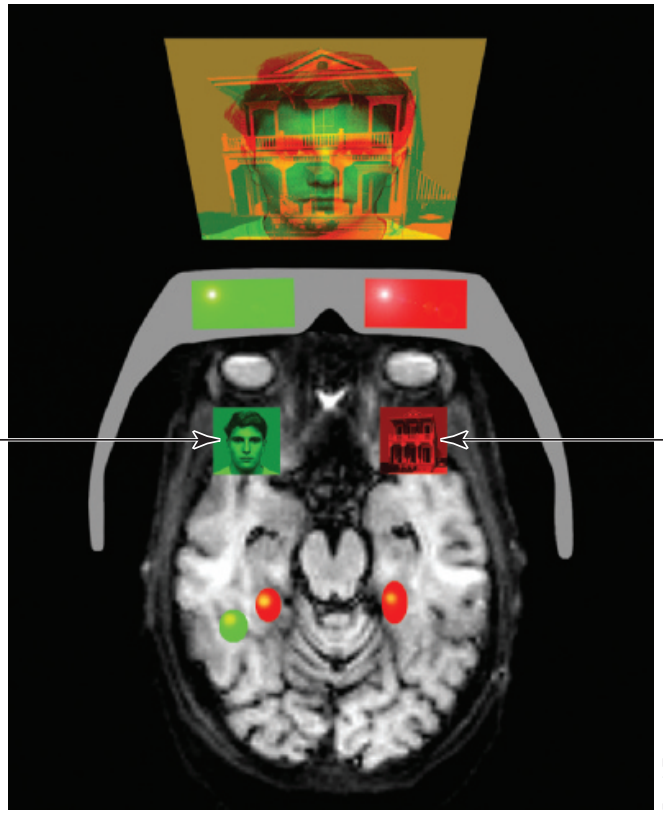




5.2 Brain Activity and Seeing

- But if each eye receives totally different images, the brain can't fuse the two images and a condition called **binocular rivalry** occurs.





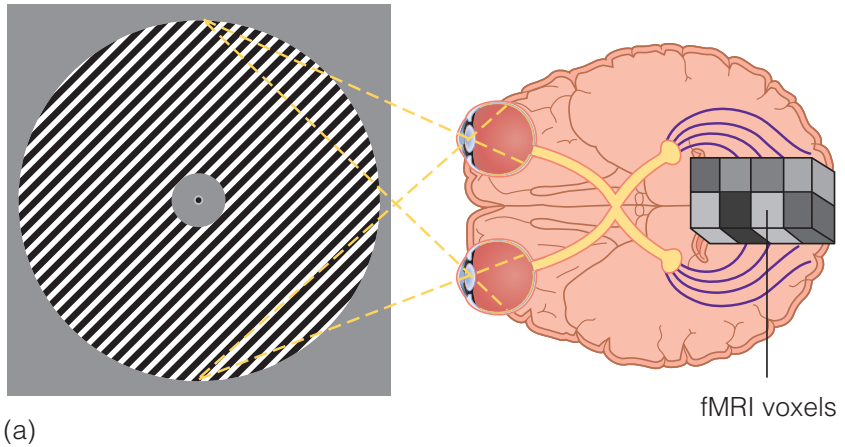
Experiment viewed the overlapping red house and green face through red-green glasses, so the house image was presented to the right eye and the face image to the left eye.

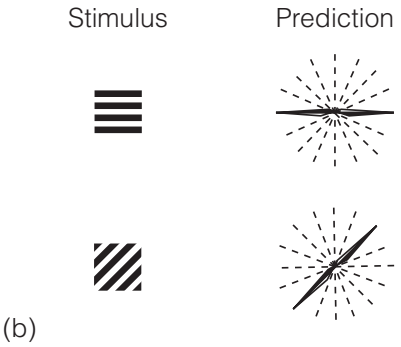
- Activity in the brain changed depending on what the person was experiencing.
- Demonstrate a dynamic relationship between perception and brain activity in which changes in perception and changes in brain activity ***mirrored*** each other.

5.3 Reading the Brain

- Whether it is possible to determine what a person is seeing by analyzing the pattern of activity in the brain?

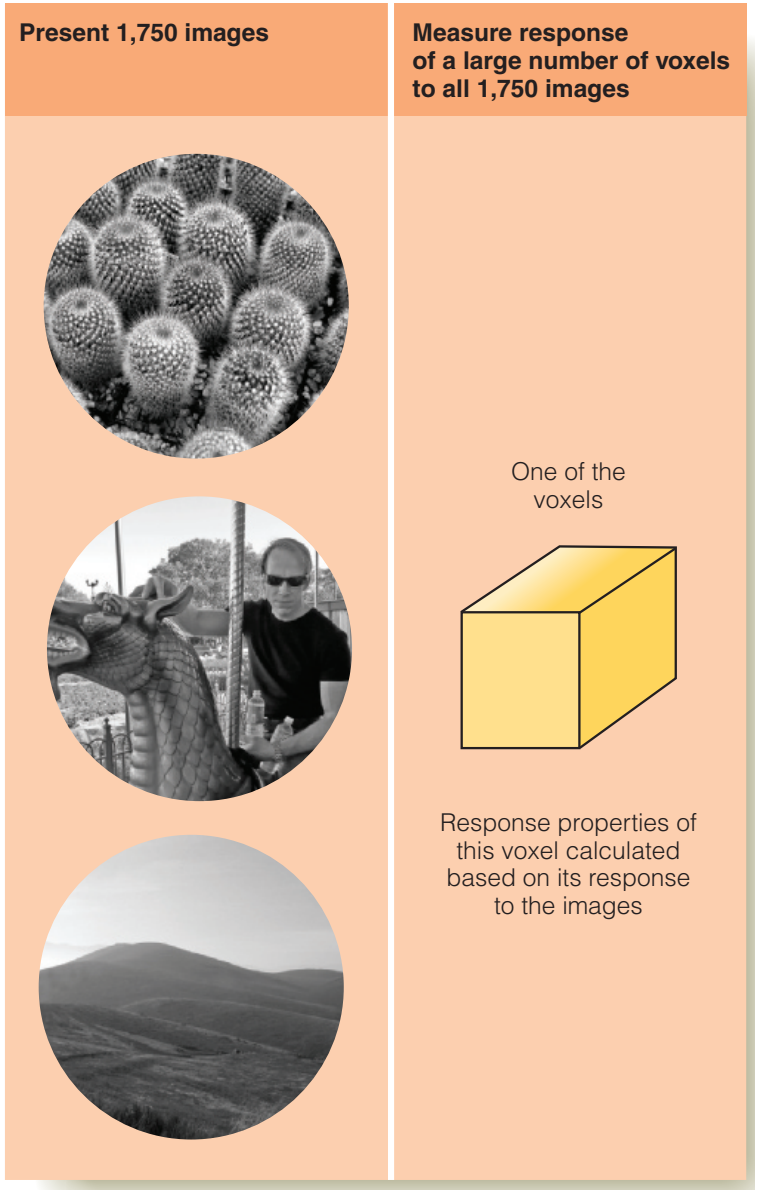
Used the relationship between voxel activity and orientation to create an "orientation decoder".





Structural Encoding

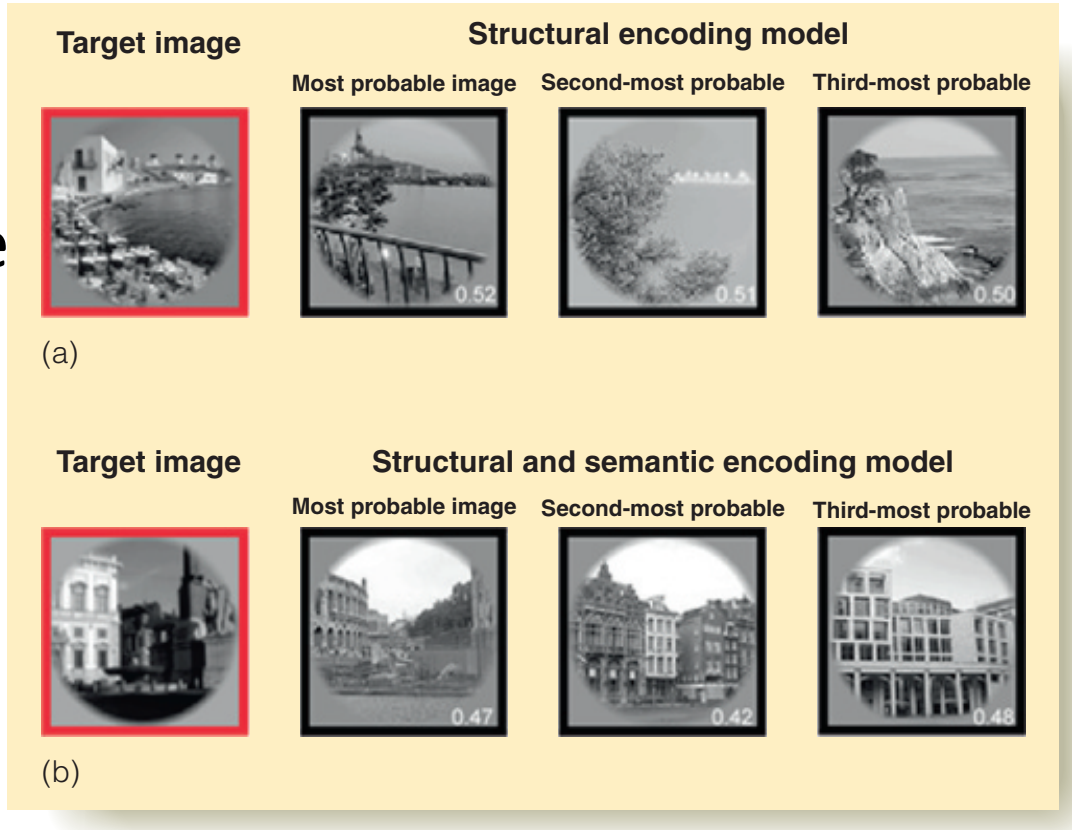
- Determining how a large number of voxels responded to specific features of each scene, such as line orientation, detail, and the position of the image.
- Once the structural encoder was calibrated, it was “reversed” to make predictions in the opposite direction.



Semantic Encoding

- The relationship between voxel activation and the *meaning* or *category* of a scene.

Encoder is calibrated by measuring the pattern of voxel activation to a large number of images that have previously been classified into categories such as “crowd”, “portrait”, “vehicle”, and “outdoor”.



- Of course, the ultimate decoder won't need to compare its output to huge image databases.
- It will just analyze the voxel activation patterns and recreate the image of the scene.
- Presently, there is only one "decoder" that has achieved this, and that is your own brain!

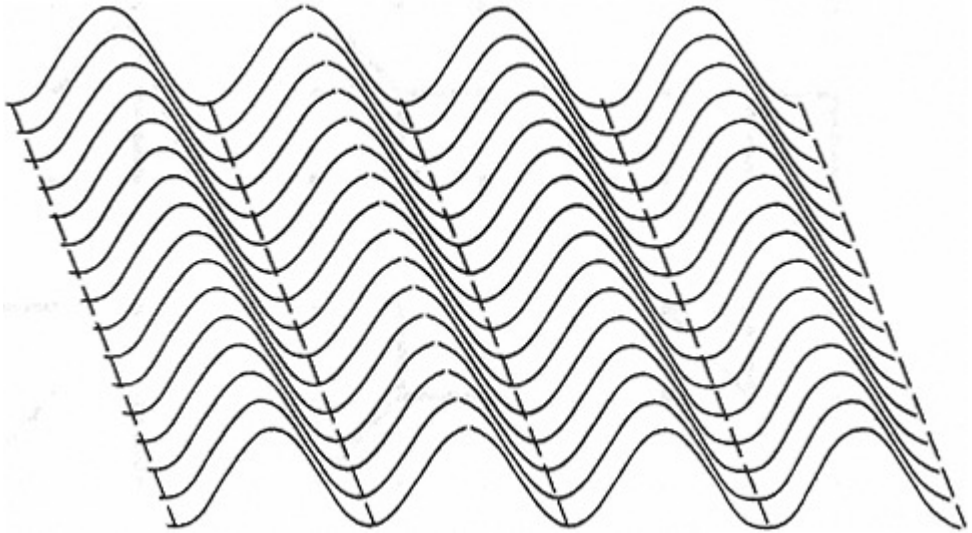
6 Parts

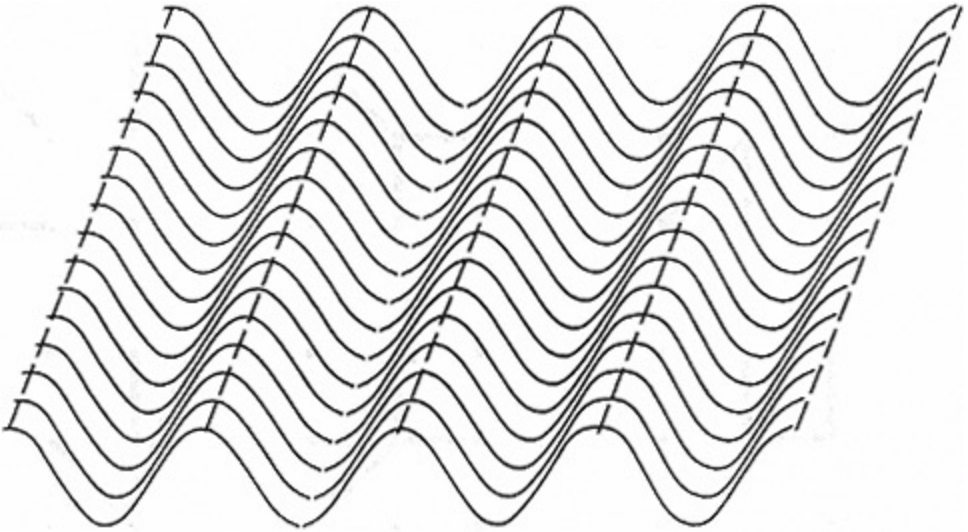
- Most complex objects are perceived as being composed of distinct parts.
  - A human body is perceived as being composed of a head, a torso, two arms, and two legs;
  - A chair is perceived as containing a seat, a back, and four legs;
- Object perceptions include the spatial relations among the parts.

6.1 Evidence for Perception of Parts

- Linguistic Evidence
  - The structure of language seems reflect the underlying structure of perception.
  - We have separate words to refer to the salient parts of many familiar objects, words such as head and torso for bodies, legs and seat for chairs.
  - ***Perception specifies the parts, for which language provides names.***

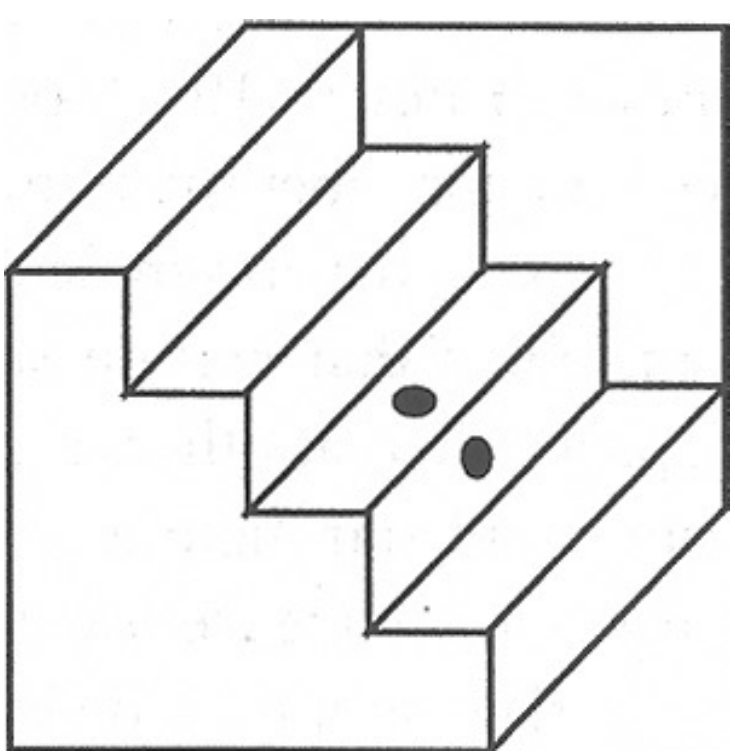
Phenomenon Evidence

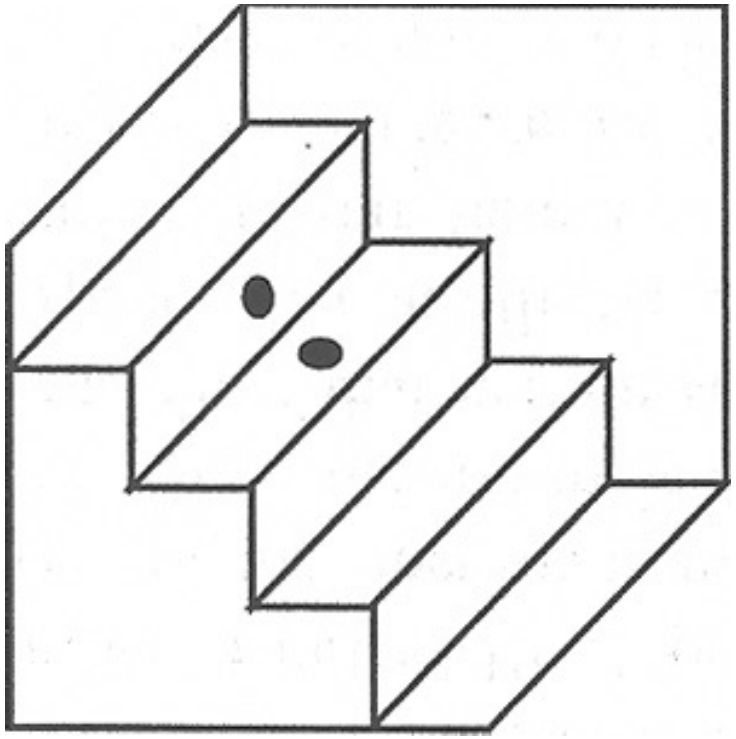




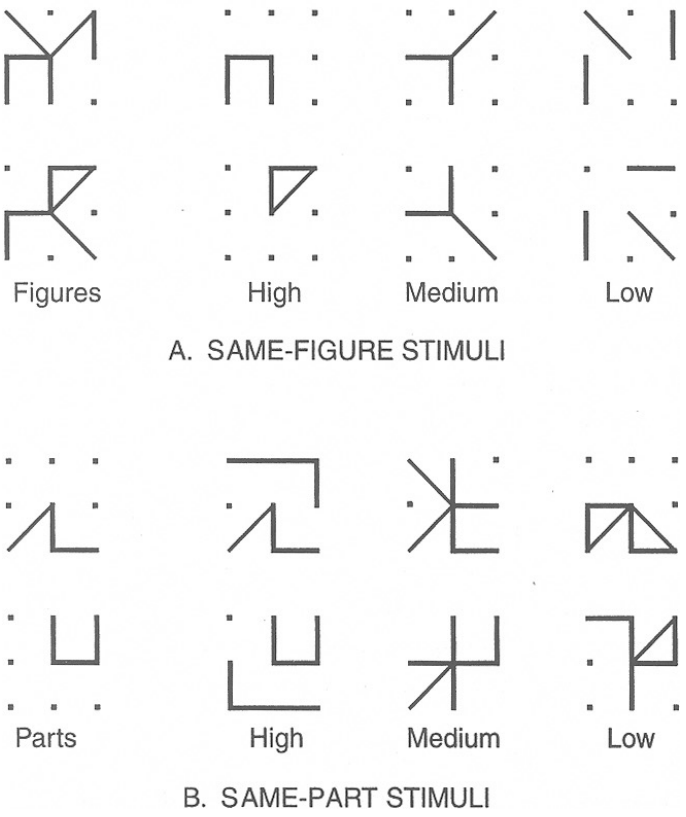
- Show not only that this surface is perceived as being made up of parts, but also something about how the parts are determined.

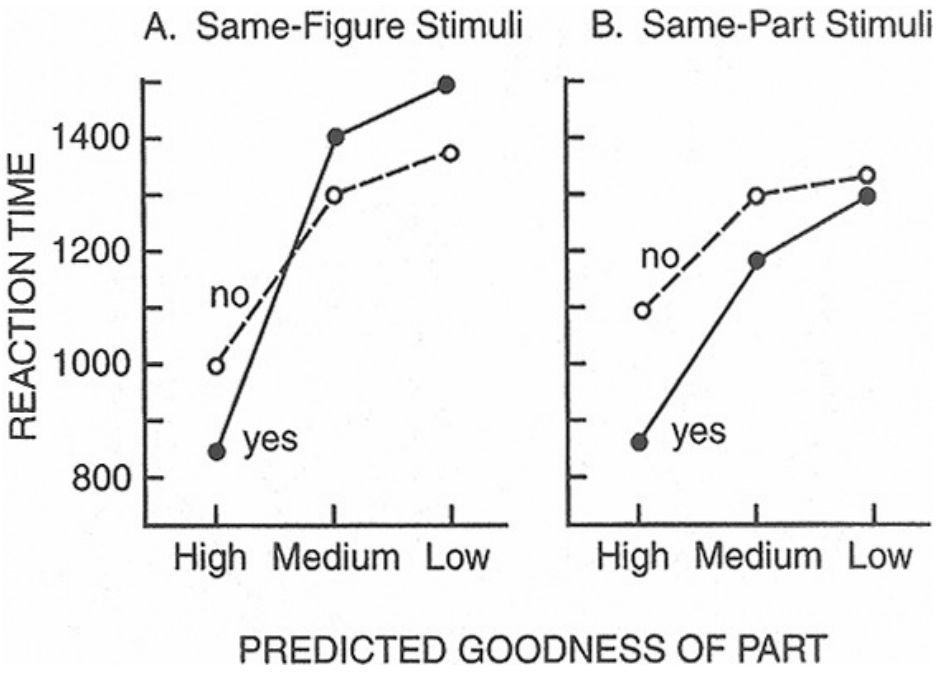
Phenomenon Evidence





Perceptual Experiments





The perception of parts is strongly context dependent.

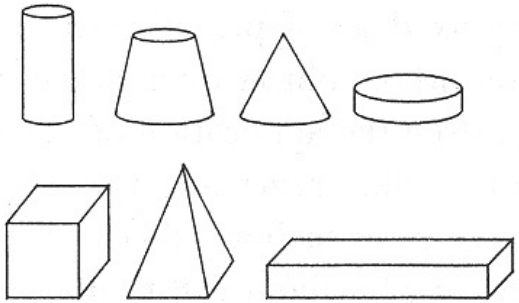
6.2 Part Segmentation

- How does the visual system determine what the parts are?
- How might the parts of an object be determined from the structural characteristics of the object itself?
  - Shape Primitives
  - Boundary Rules

6.2.1 Shape Primitives

- Basic proposal is that all perceptions of object shape can be decomposed into configurations of a relatively small set of atomic shapes.
  - 1-D: the primitive elements are line segments, edge segments, blobs, and terminators.
  - 2-D: the primitive surfaces
  - 3-D: volumetric – 3-D volumes

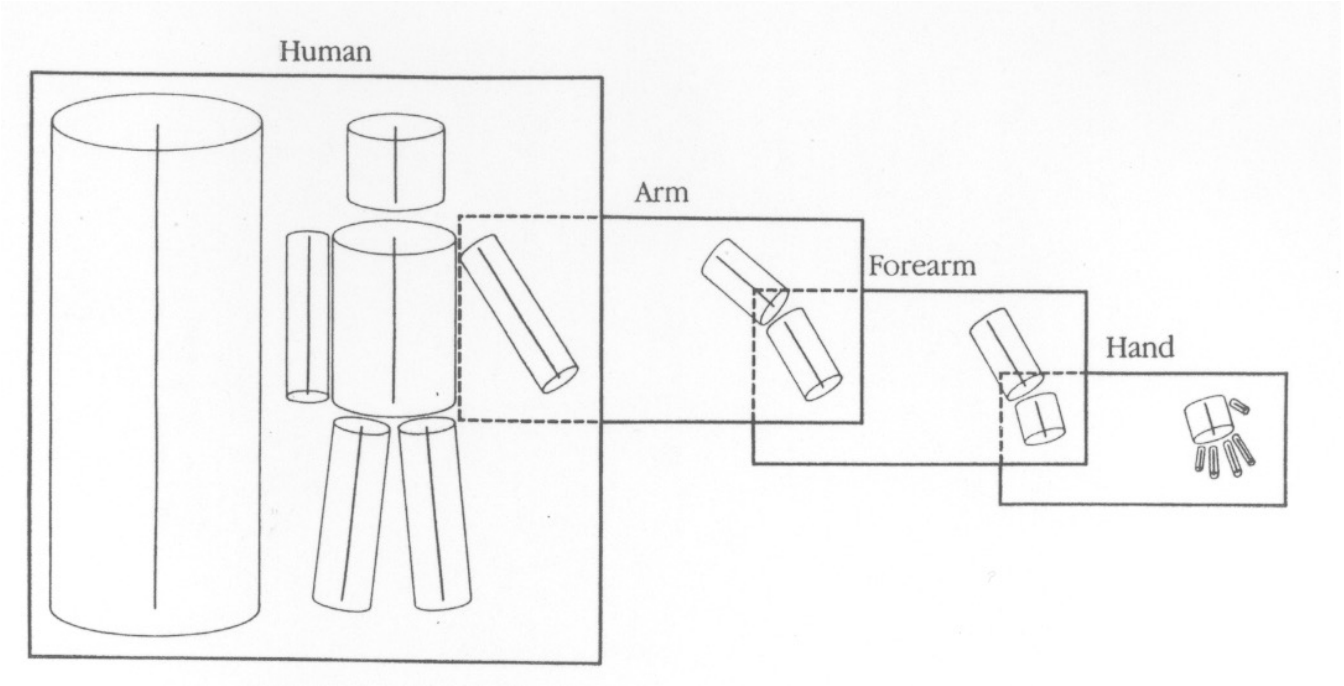
- Complex shapes can be analyzed into primitives consisting of generalized cylinders.



A

B

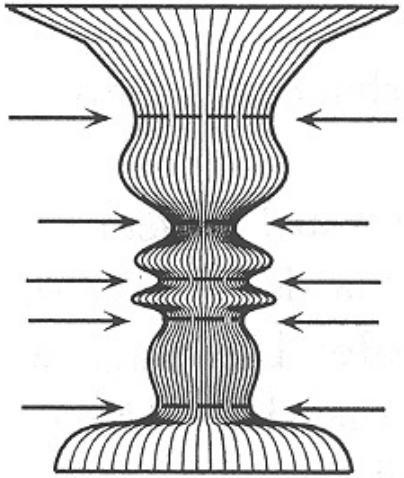
- Shapes can be represented as constructions of appropriately sized and shaped cylinders.



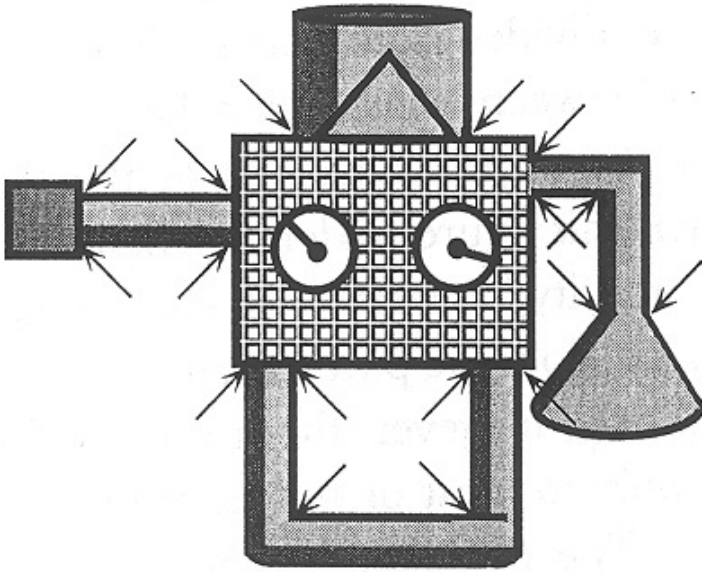
- Several potential problems:
  - The existence of contextual effects
    - There may be a preference ordering among primitives such that the visual system tends to detect certain primitives before others.
  - Often perceive those parts as having subparts.
    - assuming that there are multiple scales of description at which the primitives can be detected.
    - Using grouping principles to unite different subsets of primitives into higher level parts
  - The most difficult problem is ***what set of primitives will do justice*** to the huge diversity and immense subtlety of shapes that must be described?

6.2.2 Boundary Rules

- To define a set of general rules that specify where the boundaries lie between parts.
  - In the shape primitives approach, parts are primary and boundaries are by-products of finding parts.
  - In the boundary rules approach, boundaries are primary and parts are by-products.
- **The concave discontinuity rule**: The visual system divides objects into parts where they have abrupt changes in surface orientation.



A



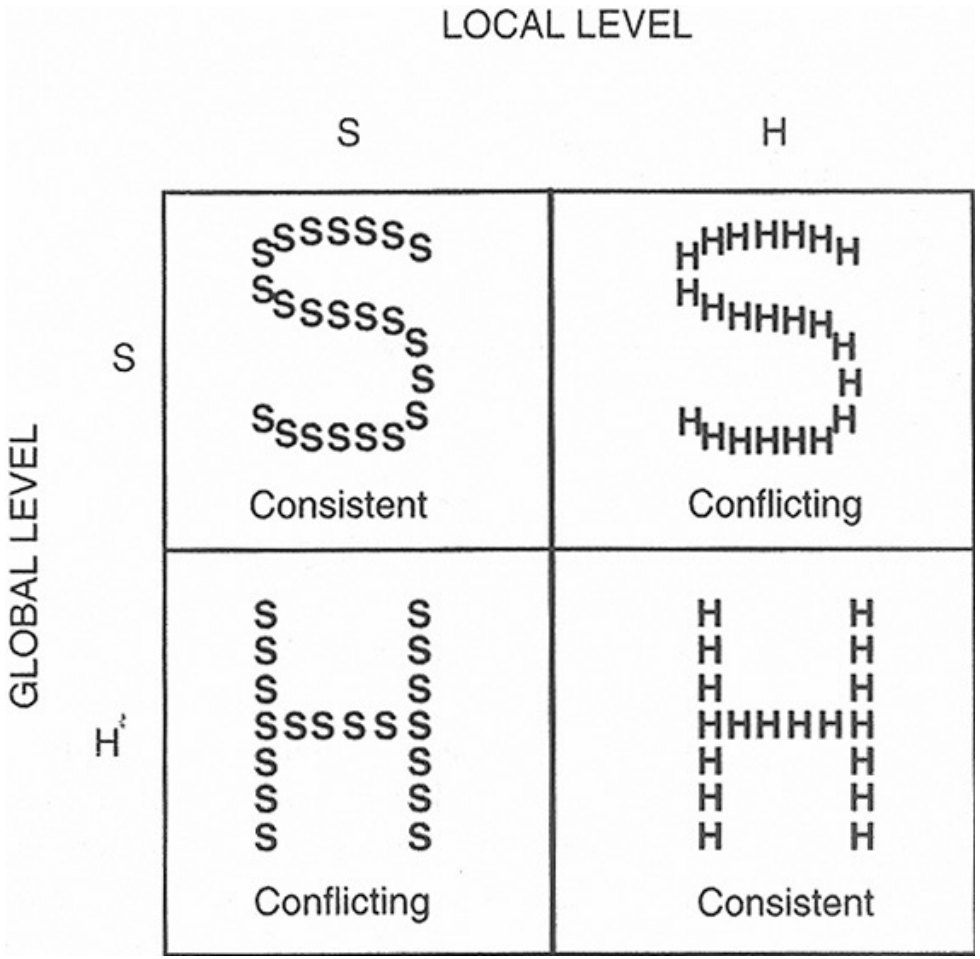
B

- A difficult problem is it merely identifies the points at which cuts would be made to divide an object into parts; it does not say which pairs of these points should be the endpoints of the cuts.

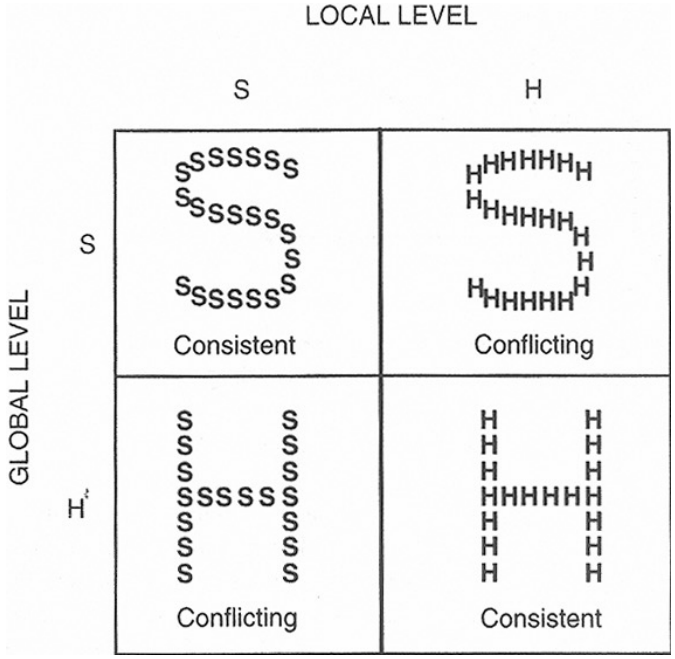
6.3 Global and Local Processing

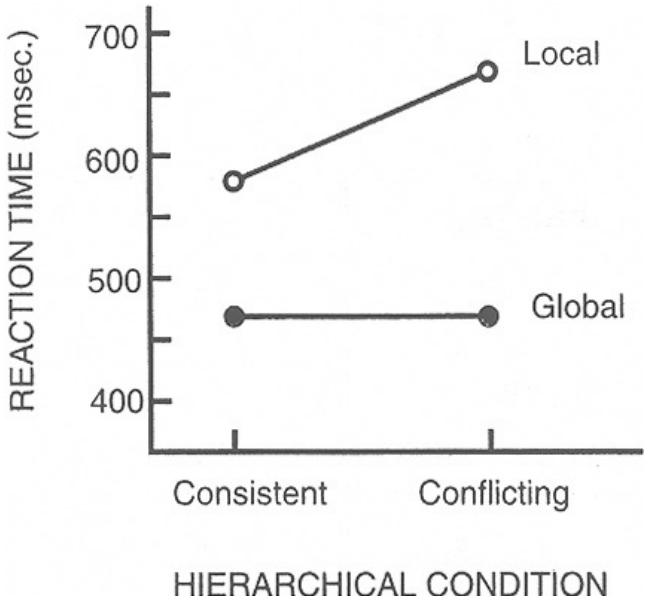
- Given that objects are perceived as being structured into parts, subparts, and so on, the question naturally arises of which level has priority in perceptual processing.
  - Within ***a physiological framework***, the answer almost certainly is that local parts are processed before global wholes.
  - Anatomical evidence shows that *massive backward projections* exist from higher to lower centers of the visual system.
  - ***Conscious perception*** might even run in the opposite direction from the order in which neural processing seems to occur, that is, from larger wholes to smaller parts.

6.3.1 Global Precedence



- If the global level is perceived first, three predictions follow:
  - **Global advantage**: Responses to global letters should be faster than those to local letters.
  - **Global-to-local interference:** Inconsistent global letters should slow responses when subjects attend to the local level, because the local level is perceived only after the global one.
  - **Lack of local-to-global interference:** Inconsistent local letters should not slow responses when subjects attend to the global level because the global level is perceived first.





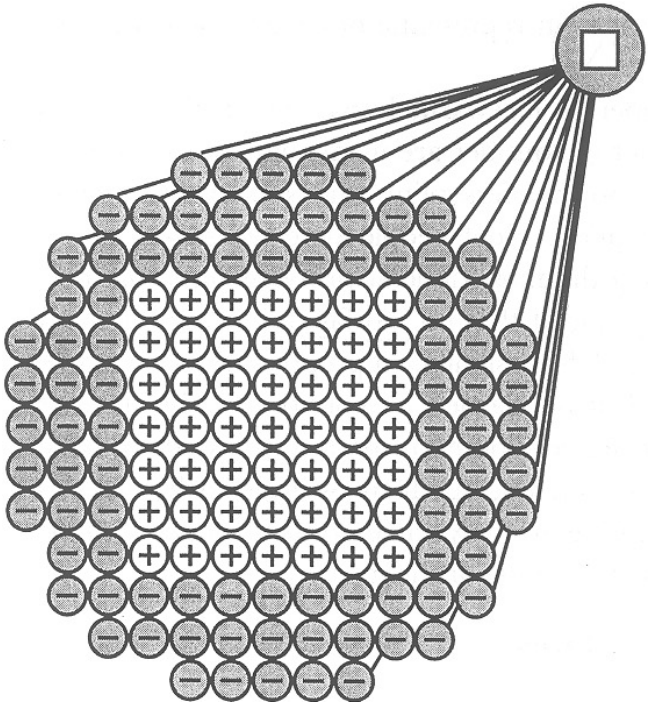
- The data thus appear to indicate that perceptual processes proceed from global processing toward more and more fine-grained analysis.
- The visual system followed a **“middle-out” order** of processing, beginning with structure at intermediate sizes and working “outward” toward both smaller and larger levels.
  - Speed of naming local versus global forms depended on their retinal sizes.
- Other experiments suggest that global and local levels of information may be processed **simultaneously** rather than sequentially.
  - Parallel processing in different size channels, some being processed slightly faster than others.

7 Shape Representation

- Of all the properties we perceive about objects, ***shape*** is probably the most important.
  - Shape allows a perceiver to predict more facts about an object than any other property.
- How the shape of objects and their parts might actually be ***represented*** within the human visual system
- How such representations might be ***compared*** for similarity.

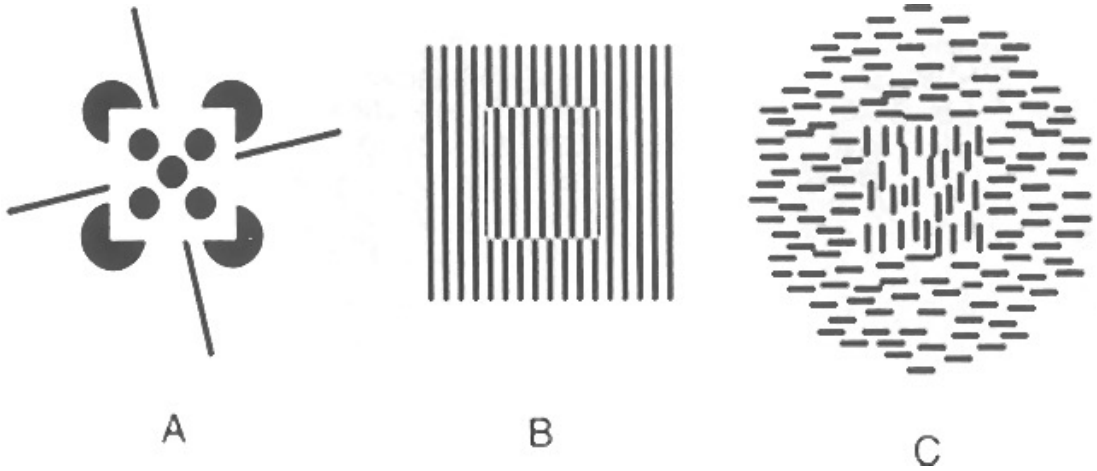
7.1 Templates

- Shape is specified by the concatenation of receptor cells on which the image of a particular object would fall.
  - Standard template: binary
  - Template: gray-scale
  - “grandmother detector”



- Strengths
  - We know with a fair degree of certainty, then, that the visual system makes use of templates to represent very simple “shapes,” such as lines and edges or local patches of sinusoidal gratings.
  - The question is whether there are any principled reasons why this approach cannot be extended to encode ***more complex shapes of real objects***, such as square, face.

- Weaknesses
  - The problem of multiple sensory channels



  - The problem of spatial transformations: translations, rotations, dilations, reflections, and their various combinations. Solutions: ***replication*** or ***normalization***
  - The problem of part structure. Solutions: ***hierarchical templates***.
  - The problem of three-dimensionality
  - Lighting, color variances

7.2 Fourier Spectra

- If the early stages of visual analysis perform a Fourierlike analysis of the image into its spatial frequency components, a good bet for representing shape and other higher-level properties of visual perception would be to describe them in terms of their spatial frequency content.
- The global Fourier analysis of an image consists of two spectra:
  - The power spectrum specifies the amplitude of the component sinusoidal gratings
  - The phase spectrum specifies the phase (spatial position) of the component gratings

- Strengths
  - Local spatial frequency theory.
  - Its formal mathematical status.
- Weaknesses
  - A global Fourier analysis represents an entire, uninterpreted image rather than individual objects.
  - It also fail to solve the problems of part structure and three-dimensionality.

7.3 Features and Dimensions

- For several decades, the most popular class of shape representation was **feature lists**: a symbolic description consisting of a simple set of attributes.
- According to this view, an object's perceived shape is defined by the set of its spatial features, and the degree of similarity between two shapes can be measured by the degree of correspondence between the two feature sets.
- Two types of features:
  - ***Global properties***: such as symmetry, closedness, and connectedness,
  - ***Local parts***: such as containing a straight line, a curved line, or an acute angle.

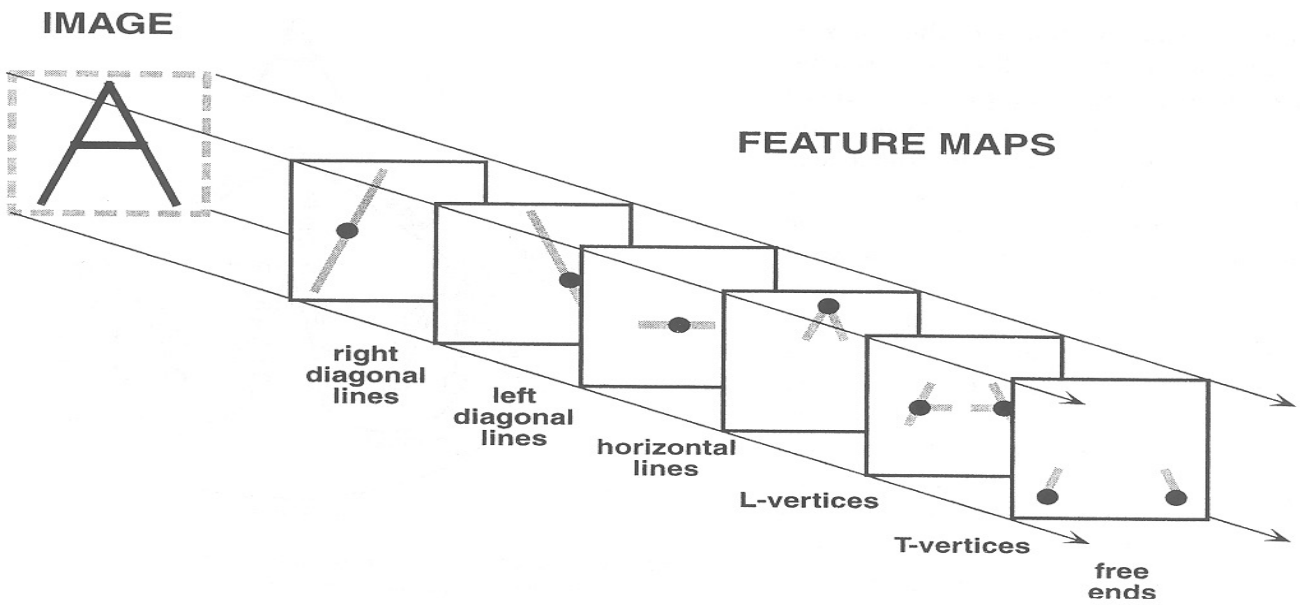
Object recognition by features

from Lowe, 2003

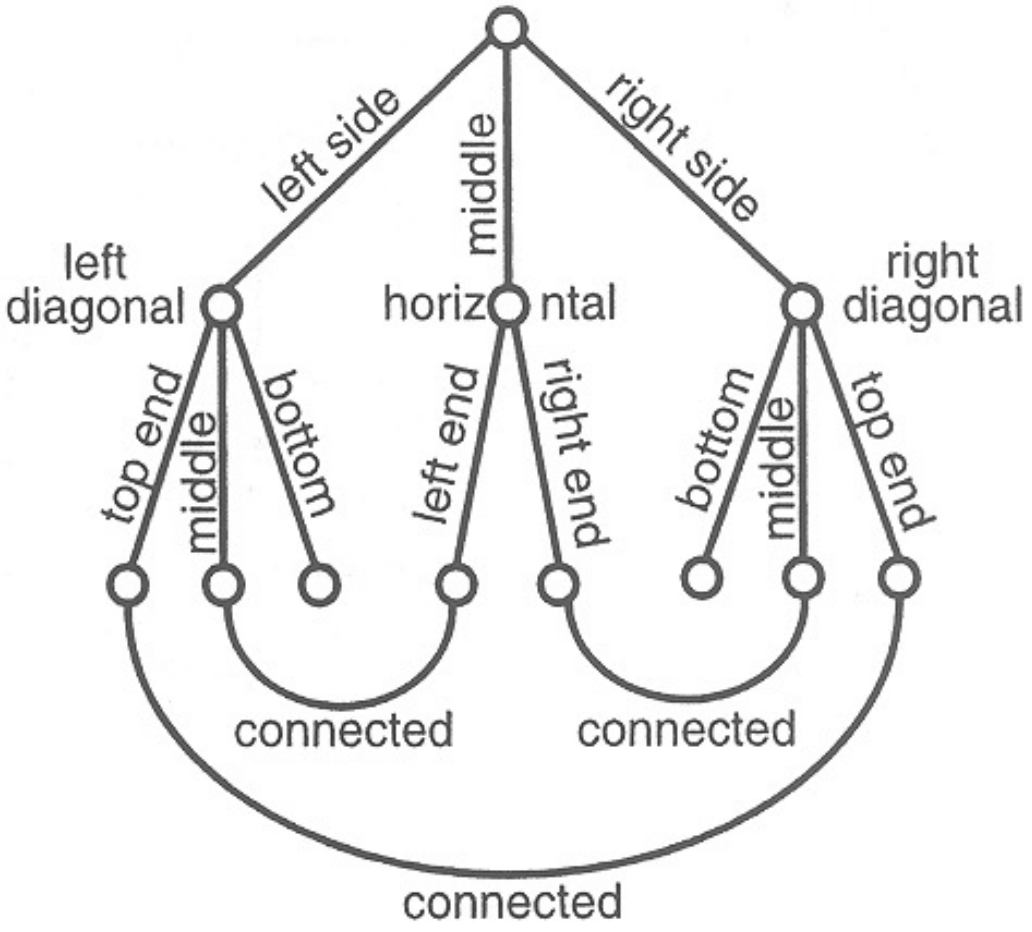
- Strengths:
  - Features also seem to be able to solve the problem of part structure simply by including the different parts of an object in its feature list.
  - Features also seem capable of solving the problems resulting from three-dimensionality, at least in principle.
- Weaknesses
  - It is still the very difficulty problem of specifying ***what the proper features for a shape representation might be.***
  - It is often unclear how to determine computationally ***whether a given object has the features that are proposed to make up its shape representation***.

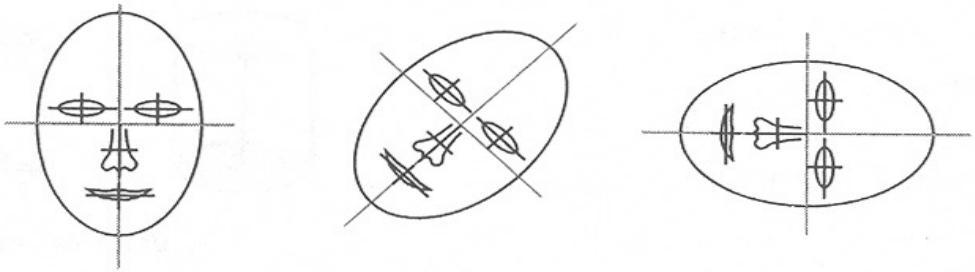
7.4 Structural Descriptions

- Representations that contain ***explicit information*** about parts and about relations between parts.
- They are usually depicted as **networks** in which nodes represent the whole object and its various parts and labeled links (or arcs) between nodes represent the relations between parts.

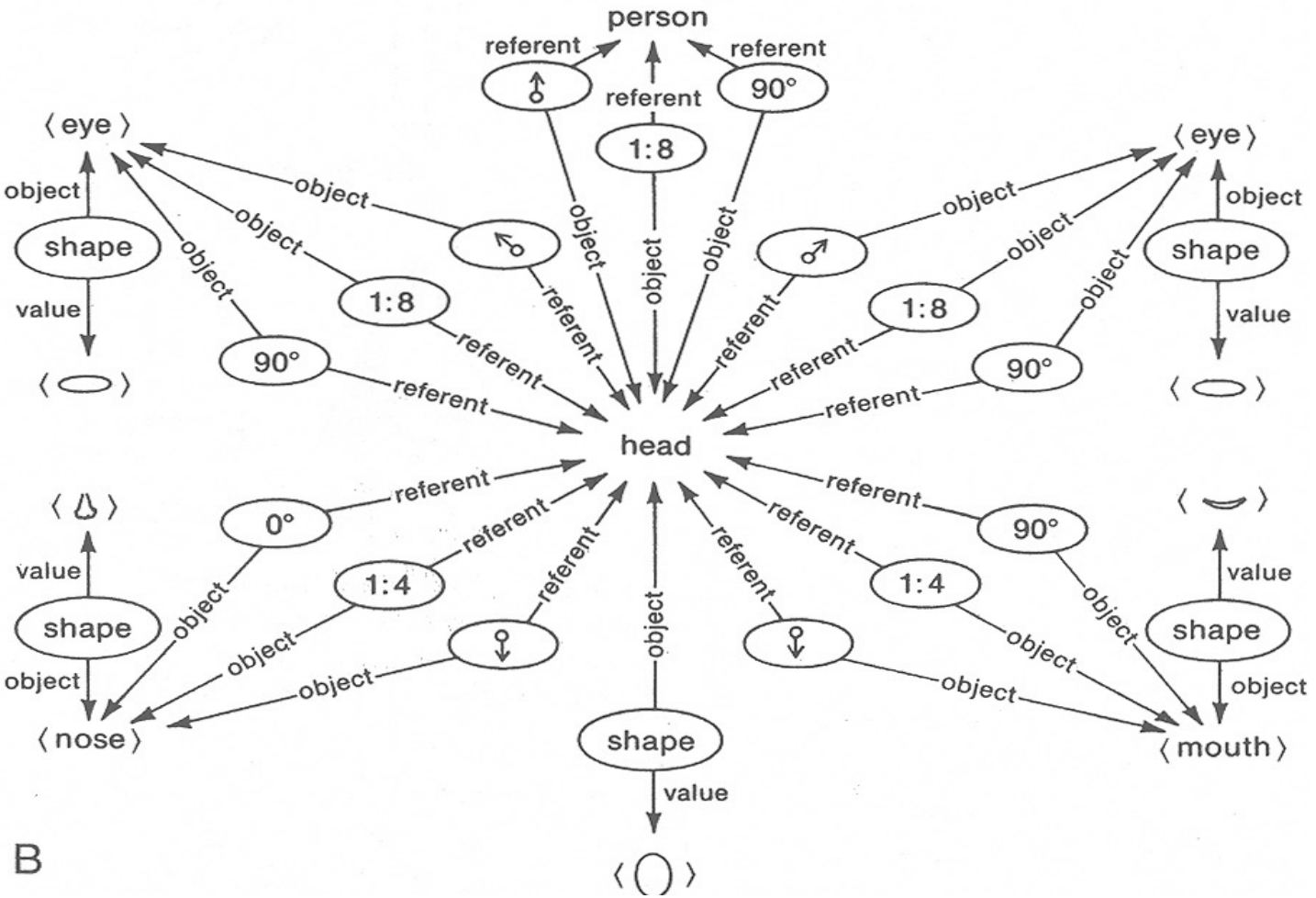


A

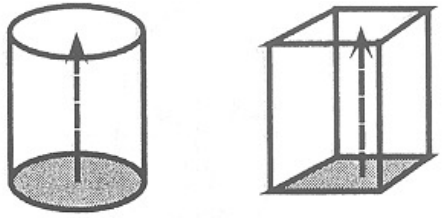

left side
middle
right side
left diagonal
horizontal
right diagonal
top end
middle
bottom
left end
right end
bottom
middle
top end
connected
connected
connected



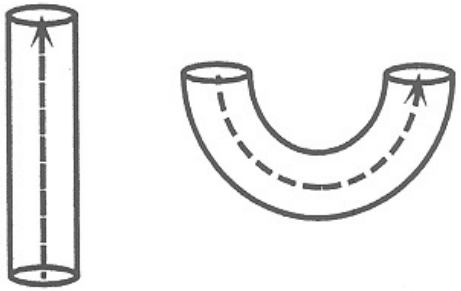
A


person
referent
referent
referent
1:8
90°
⟨ eye ⟩
object
shape
value
⟨ ⟩
object
object
object
1:8
90°
referent
referent
referent
head
⟨ eye ⟩
object
shape
value
⟨ ⟩
object
object
object
1:8
90°
referent
referent
referent
⟨ ⟩
value
shape
object
⟨ nose ⟩
0°
1:4
referent
referent
referent
object
object
object
shape
value
⟨ ⟩
90°
1:4
referent
referent
referent
object
object
object
⟨ ⟩
value
shape
object
⟨ mouth ⟩

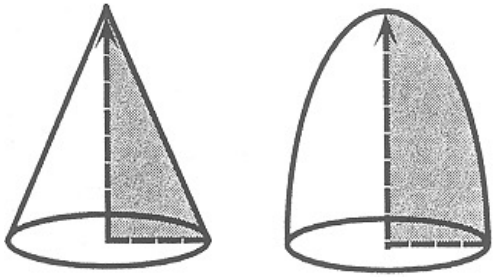
B



A. Variable Base



B. Variable Axis

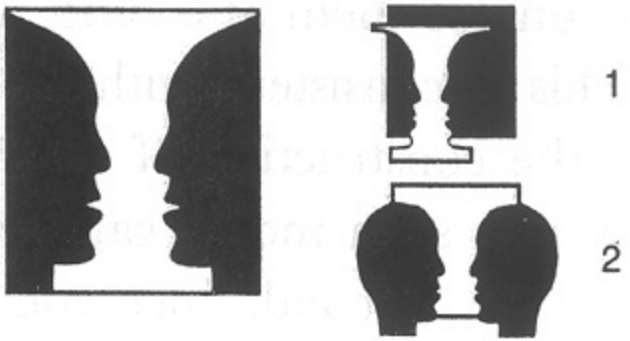


C. Variable Sweeping Rule

- A set of volumetric shape primitives

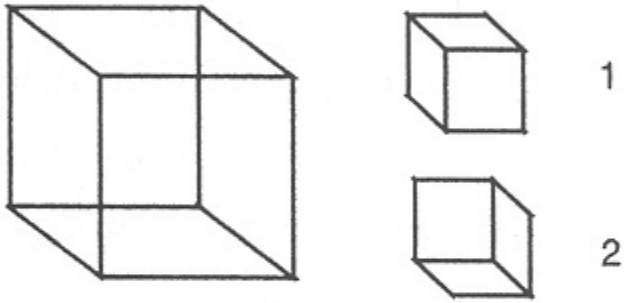
- Strengths
  - They are invariant over different sensory channels.
  - They can account for the effects of spatial transformations on shape perception.
  - They deal explicitly with the **problem of part structure** by having quasi-independent representations of parts and spatial relations among them.
  - And they are able to **represent 3-D shape** by using volumetric primitives and 3-D spatial relations in representing 3-D objects
- Weaknesses
  - The representations become quite **complicated**, so matching two graph structures constitutes a difficult problem by itself.
  - The shape primitives and their relations must be **computable from real images.**

3D Structural Descriptions



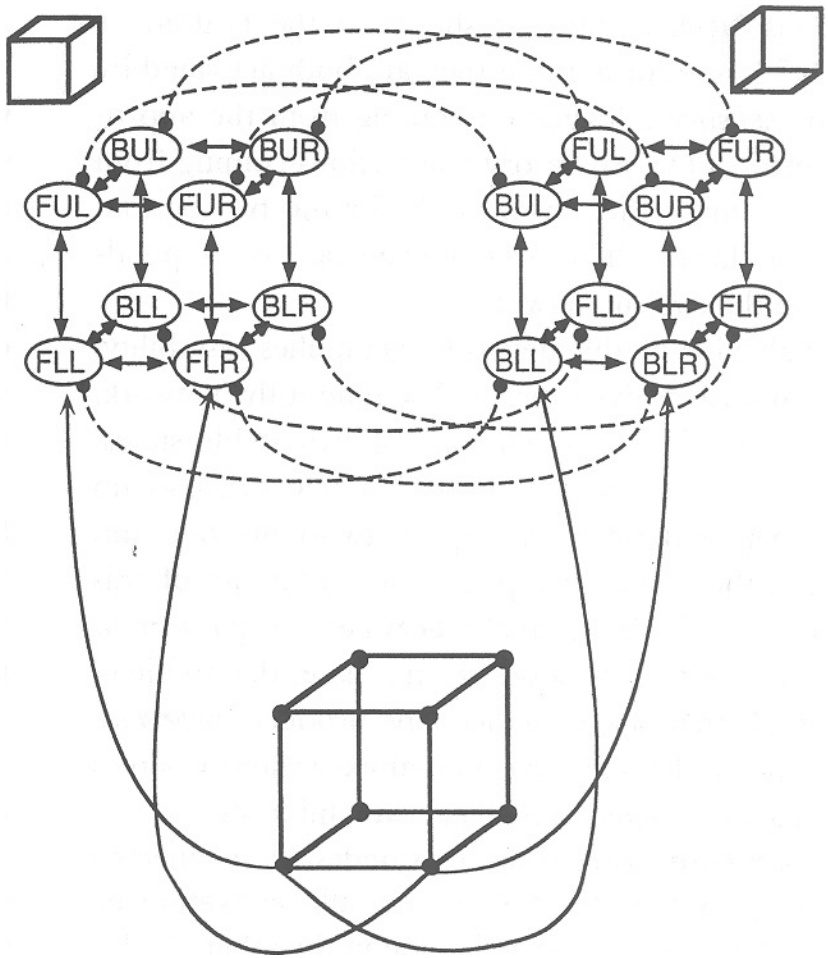
A. Vase/Faces

C. Duck/Rabbit



B. Necker Cube

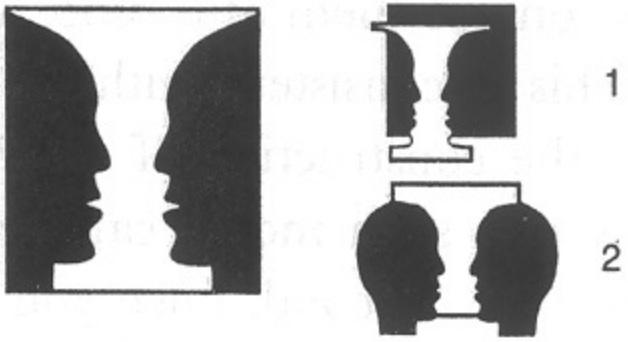
Connectionist Network Models



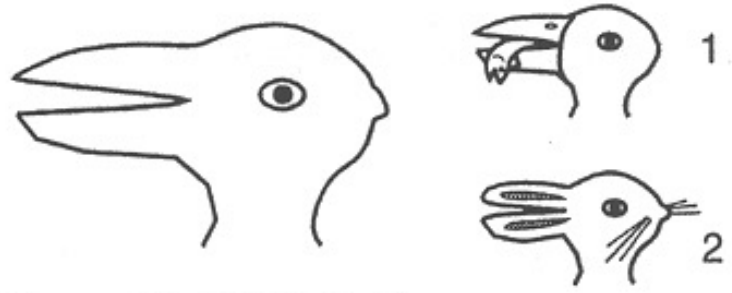
Connectionist network model of perceiving a Necker cube.

- The network representations of the two interpretation are two subnetworks embedded within the larger interconnected network.
- The behavior of this network arises from its ***architecture***.
  - **Cooperation** arises when two nodes are connected by mutually excitatory links. When one node of a given interpretation is activated, this activation spreads first to its nearest neighbors and eventually to all the nodes within its subnetwork.
  - **Competition** arises when two nodes are connected by mutually inhibitory links. As a result, activation in one tends to decrease activation in the other.

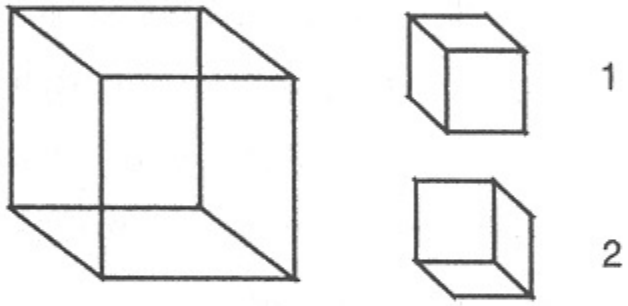
Multistable perceptions



A. Vase/Faces



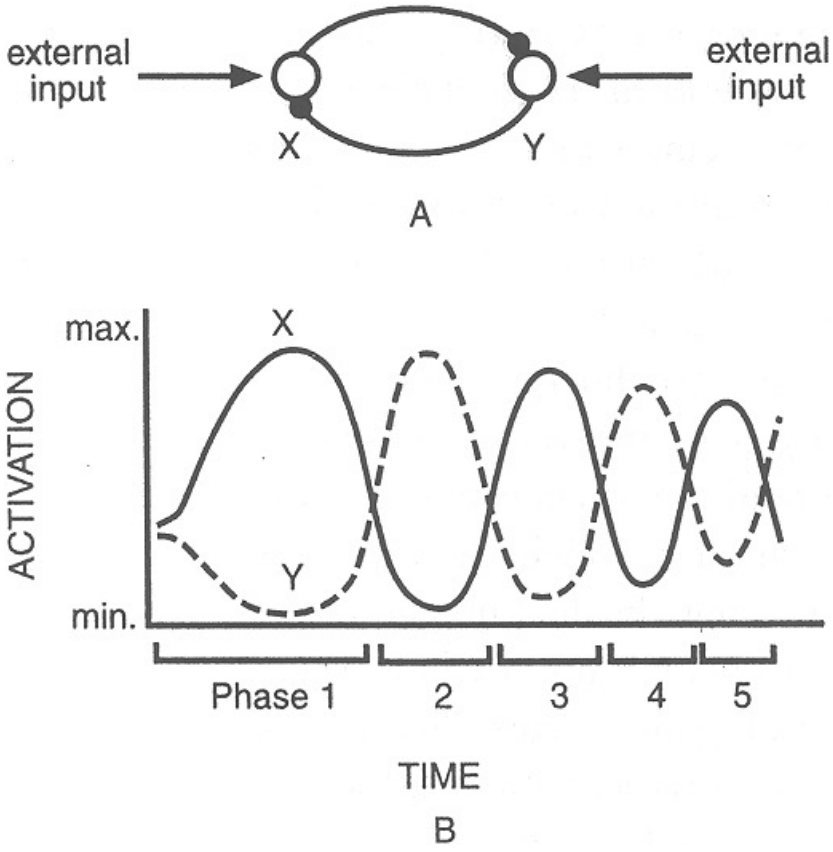
C. Duck/Rabbit



B. Necker Cube

Perceptions alternate among two or three different interpretations, once they have been seen.

- Neurons tend to fire less vigorously after a period of prolonged stimulation.
  - The biochemical resources that the cell needs to continue firing begin to become depleted over a matter of seconds, thus lowering its firing rate to the same stimulus.
- Consider what happens to the two mutually inhibitory units in the simplified network.



- Evidences:
  - Data from experiments on the perception of ambiguous figures support several predictions derived from neural fatigue theories.
  - Another prediction of the neural fatigue hypothesis that has been supported by experimental evidence is that ***the rate of alternation*** between the two interpretations should accelerate over time.
- The critical ingredients:
  - ***mutual excitation*** within each of the subnetworks
  - ***mutual inhibition*** between the subnetworks
  - ***neural fatigue*** that decreases the activation of highly active units over time.

8 Perceiving Category

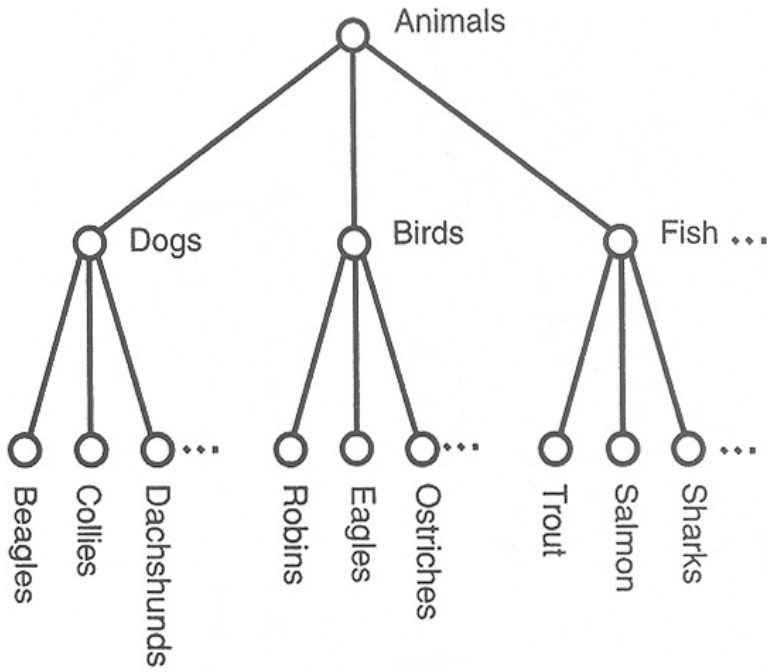
- Four Components of Categorization.
  - **Object representations**: The relevant characteristics of the to-be-categorized object must be perceived and represented within the visual system.
  - **Category representation**: Each of the set of possible categories must be represented in memory in a way that is accessible to the visual system.

– **Comparison Processes**

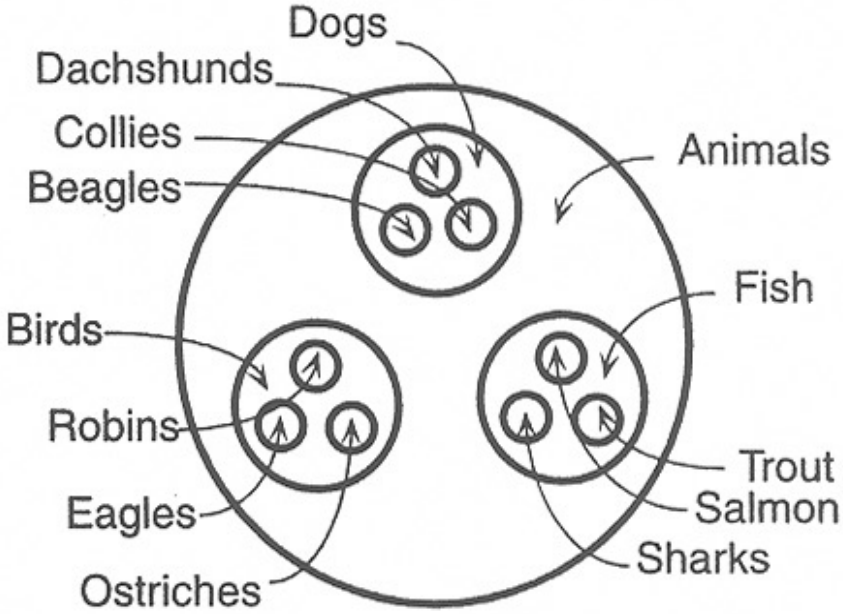
- The comparison process requires matching the object representation to the category representation.
- Major issues:
  - ***Comparing representations across categories***: Whether the process of matching a given object representation to the set of all category representations takes place ***serially or in parallel.***
  - ***Comparing elements within a representation***: Assuming that each object representation consists of multiple elements (features, dimensions, parts, or whatever), whether these elements are matched to a given category ***serially or in parallel***.

- **Decision Processes**
  - There are two important problems:
    - **Novelty**: The decision rule should be able to recognize that an object is novel so that a new category can be established for it rather than incorrectly assigning it to a known category.
    - **Uniqueness**: Each object is a member of just one category.
  - Three rules:
    - **Threshold rules**
    - **Maximum rules**
    - **Maximum-over-threshold rule**

8.1 Categorical Hierarchies



A. Hierarchical Tree Representation

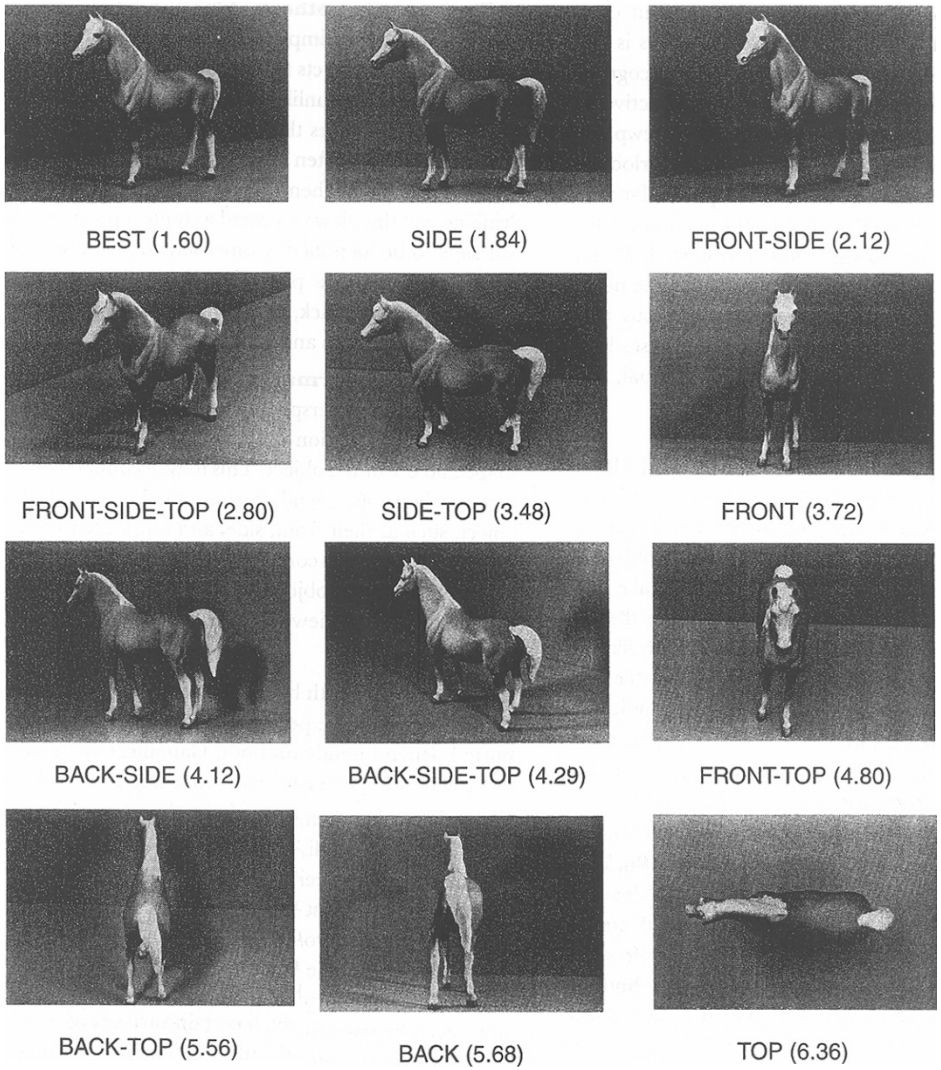


B. Venn Diagram Representation

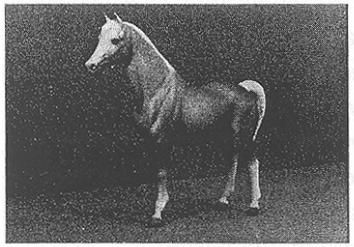
Basic-Level Categories

- When a given object is categorized, ***at which vertical level of the hierarchy*** does this process occur?
- The answer appears to be that most people recognize objects first ***at an intermediate level*** in the categorical hierarchy.
- Three criteria of basic-level categories
  - Similar shape: dogs
  - Similar motor interactions: piano vs. guitar
  - Common attributes:

8.2 Perspective Viewing Conditions

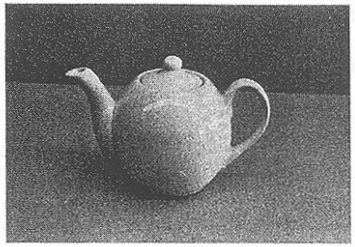


- Canonical Perspective
  - the best, most easily identified view for each object.



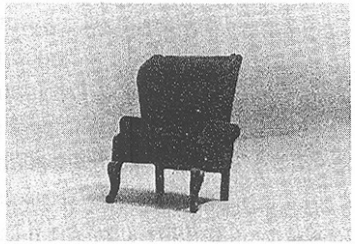
HORSE

PIANO

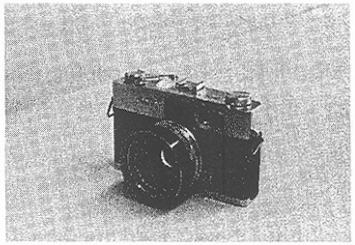


TEAPOT

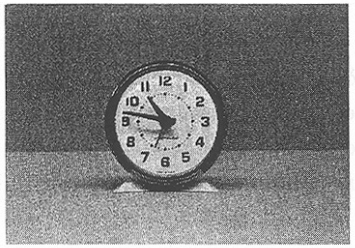
CAR



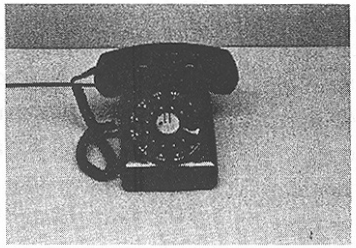
CHAIR



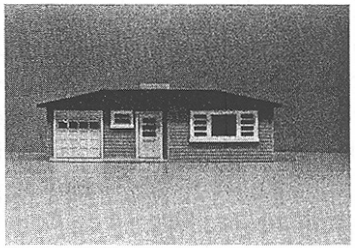
CAMERA



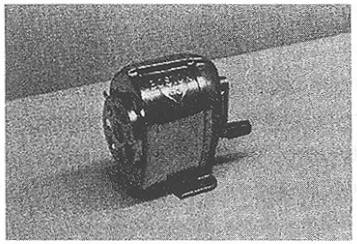
CLOCK



TELEPHONE

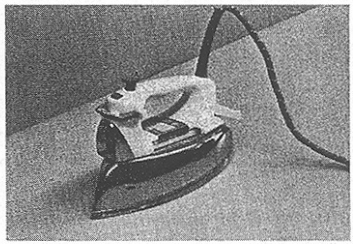


HOUSE



PENCIL SHARPENER

SHOE



IRON

Orientation Effects

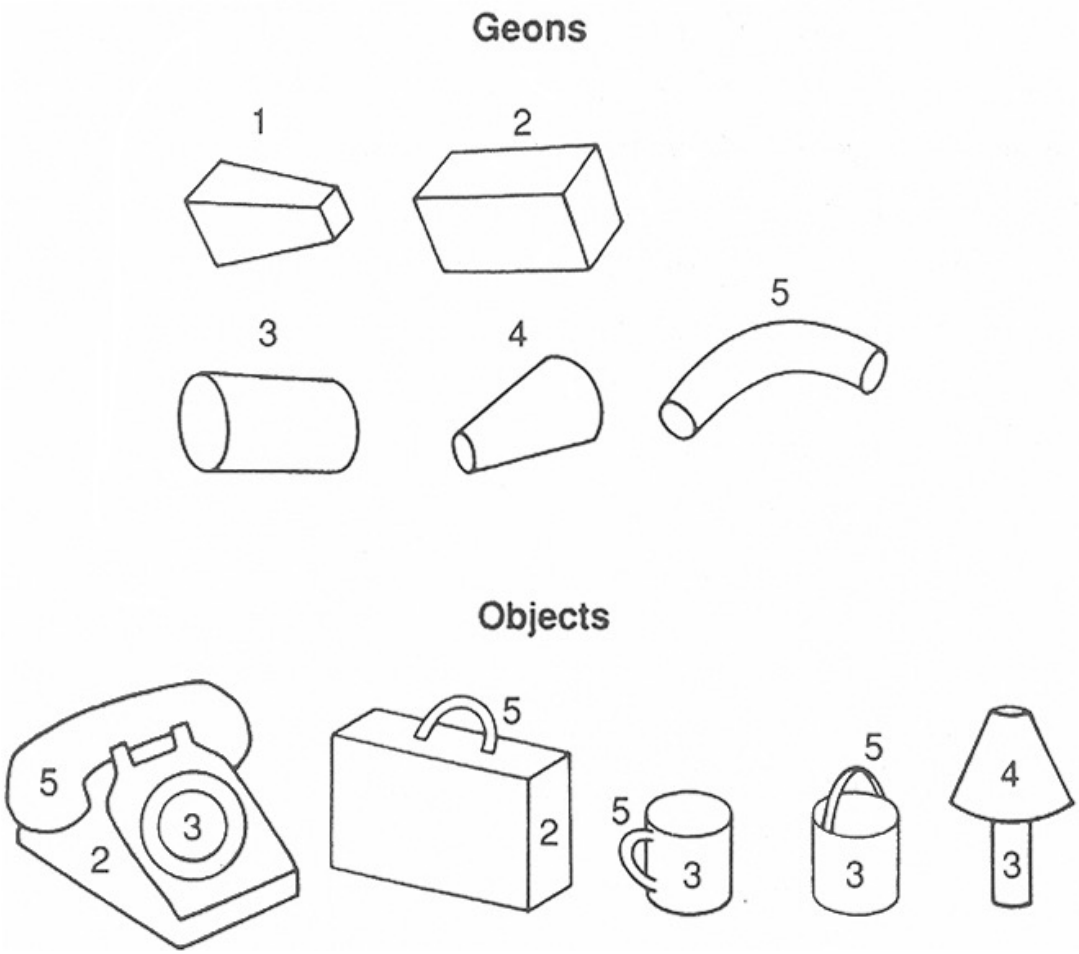
- Although object categorization initially seemed like an orientation- and view-invariant process.
- People may actually store ***multiple representations*** of the same object at different orientations rather than either a single representation at one canonical orientation or an orientation-invariant representation.

8.3 Object Categorization

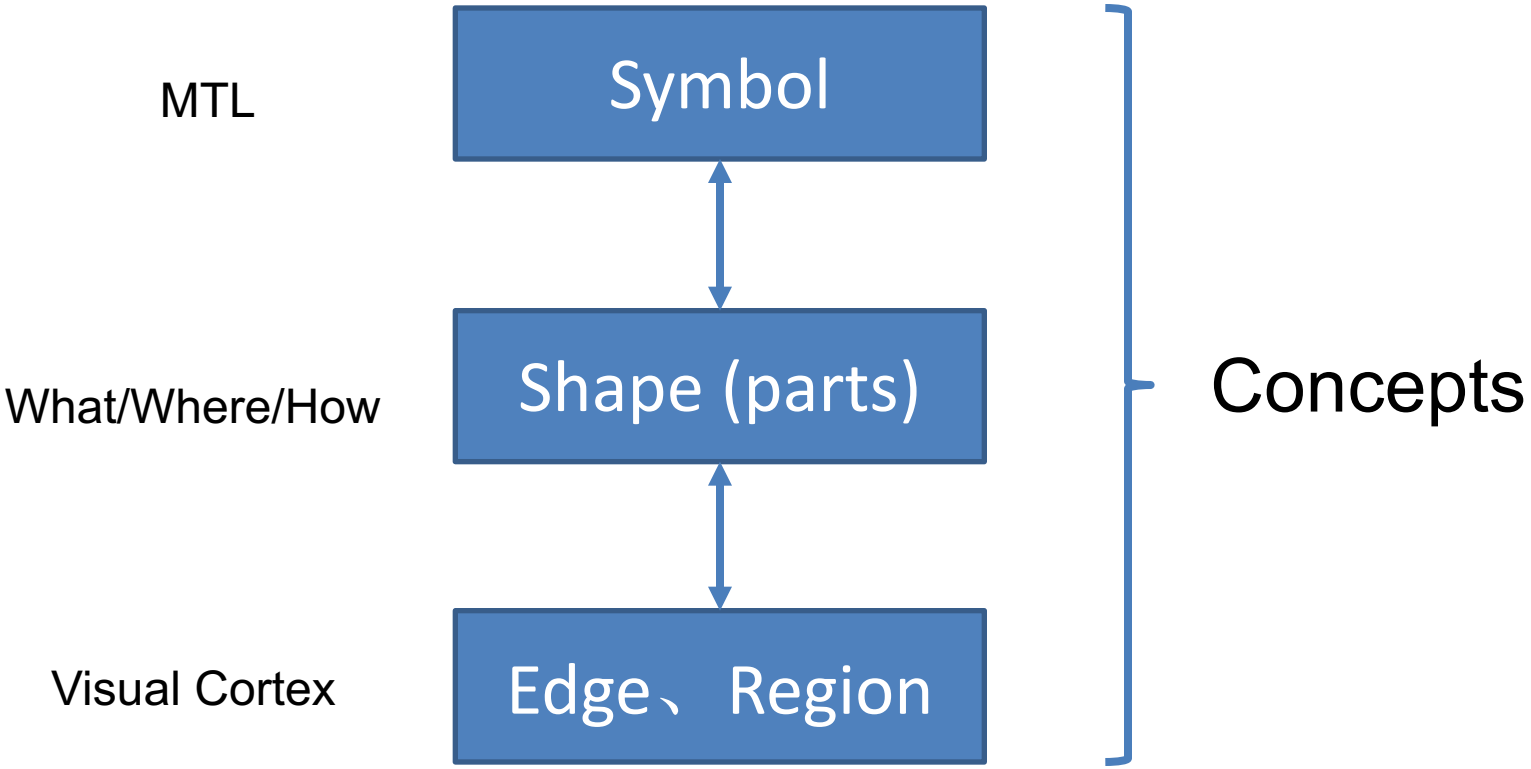
- Perhaps the single most influential theoretical approach to object categorization over the past several decades is an extension of ideas behind ***structural description theories*** of shape representation.

Recognition by Components Theory

- **Geons**: objects can be specified as spatial arrangements of primitive volumetric components.
- **Assumption**: both the stored representations of categories and the representation of a currently attended object are ***volumetric structural descriptions***.



Examples of geons and their presence in objects


MTL
Symbol
What/Where/How
Shape (parts)
Visual Cortex
Edge、 Region
Concepts

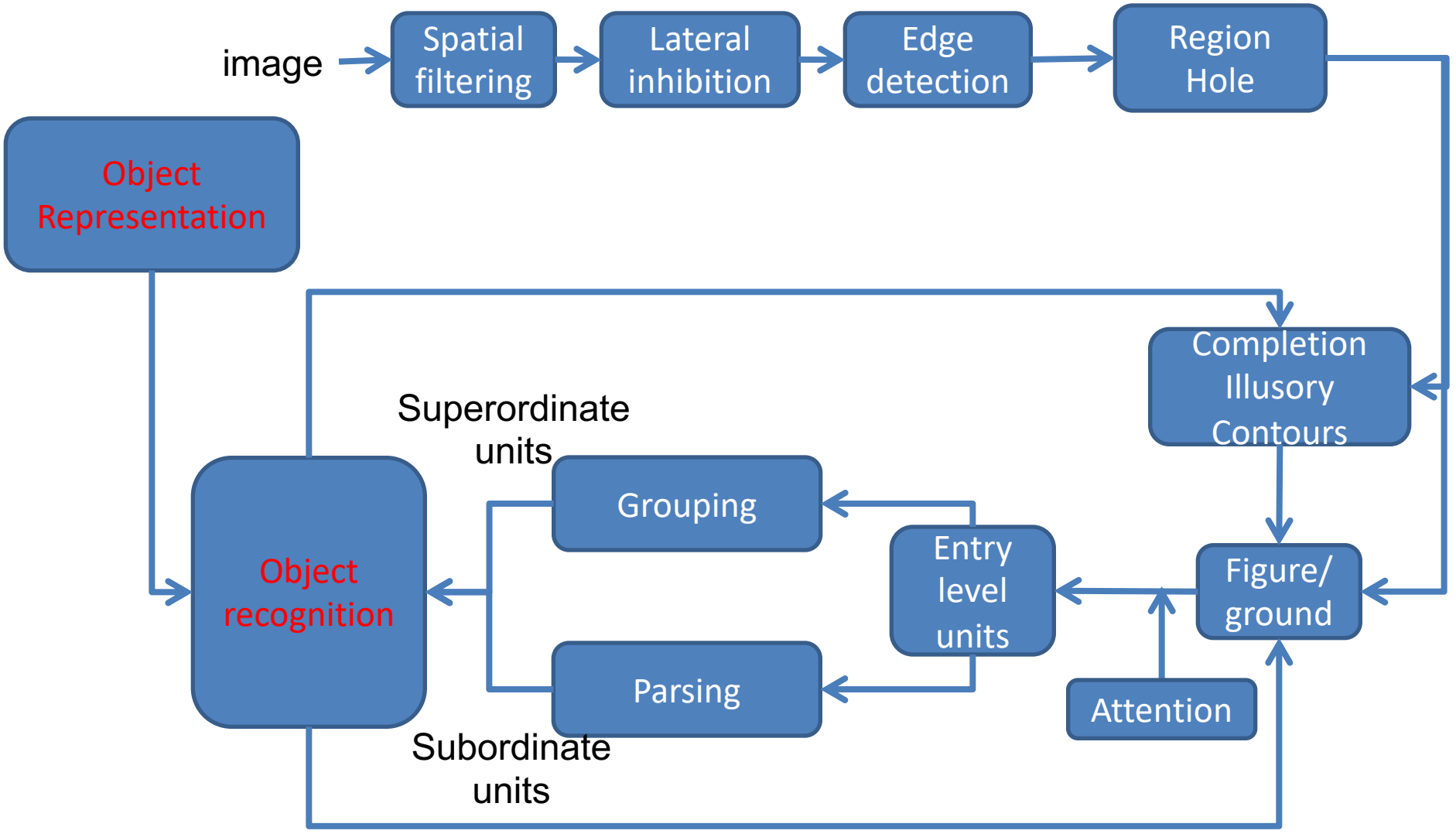
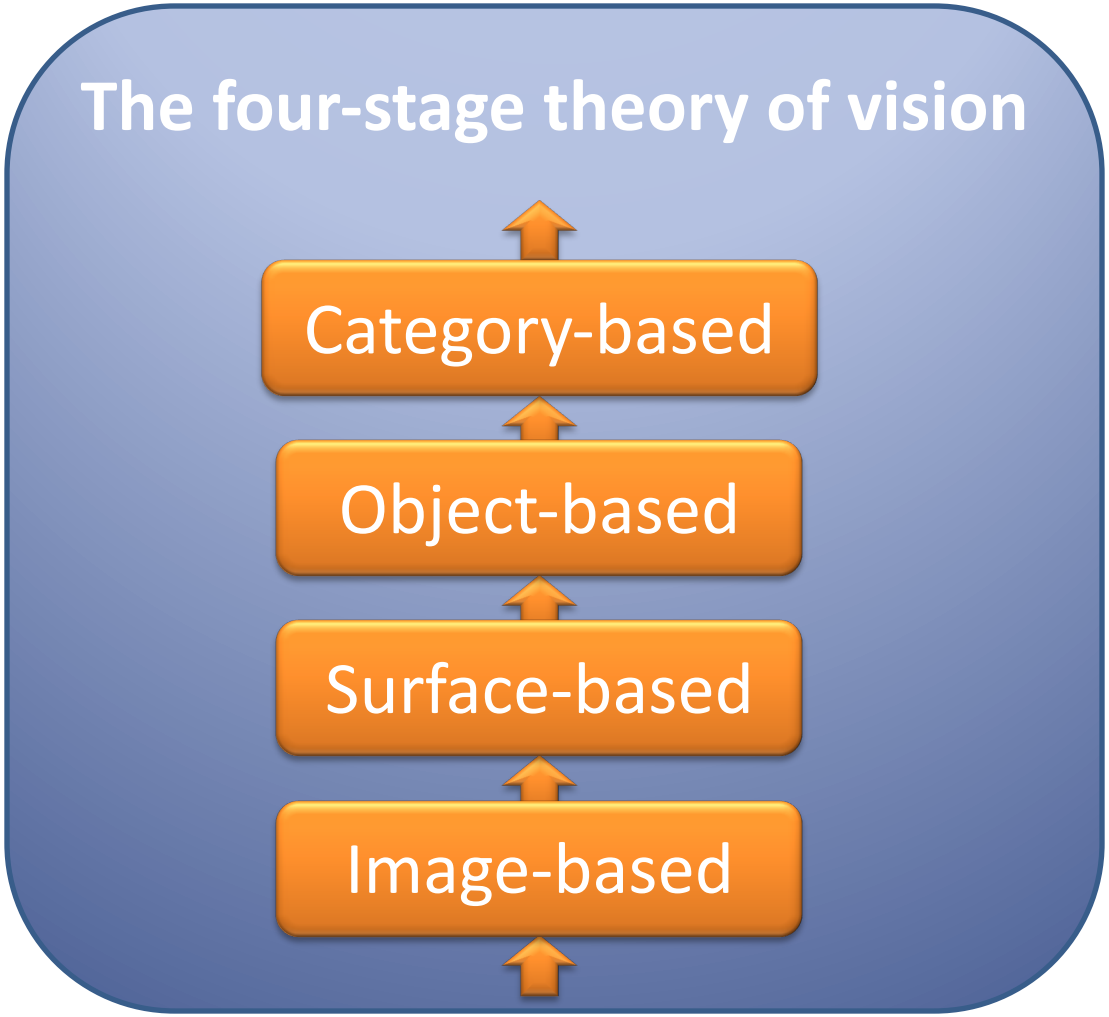


image
Spatial filtering
Lateral inhibition
Edge detection
Region Hole
Object Representation
Completion Illusory Contours
Superordinate units
Grouping
Object recognition
Entry level units
Figure/ ground
Parsing
Attention
Subordinate units



Perceptual Categorization

Shape representation
Part and whole
Perceptual organization
Figure/ground organization

Depth perception

Edge detection

Question?