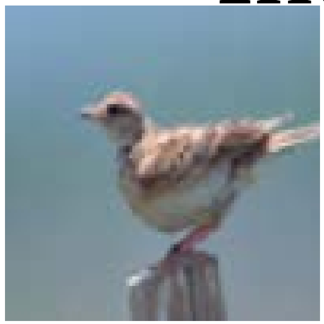
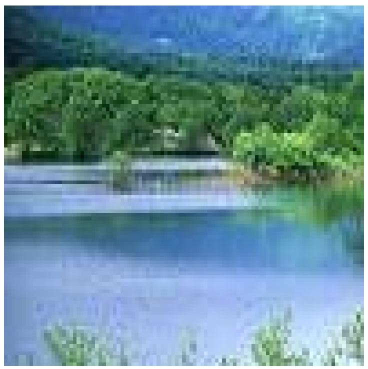
Object Recognition

Hongmei Yan

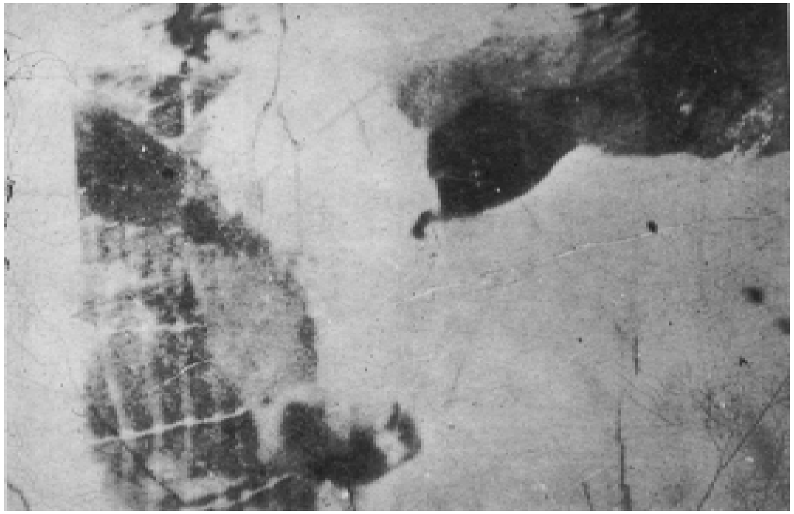
hmyan@uestc.edu.cn

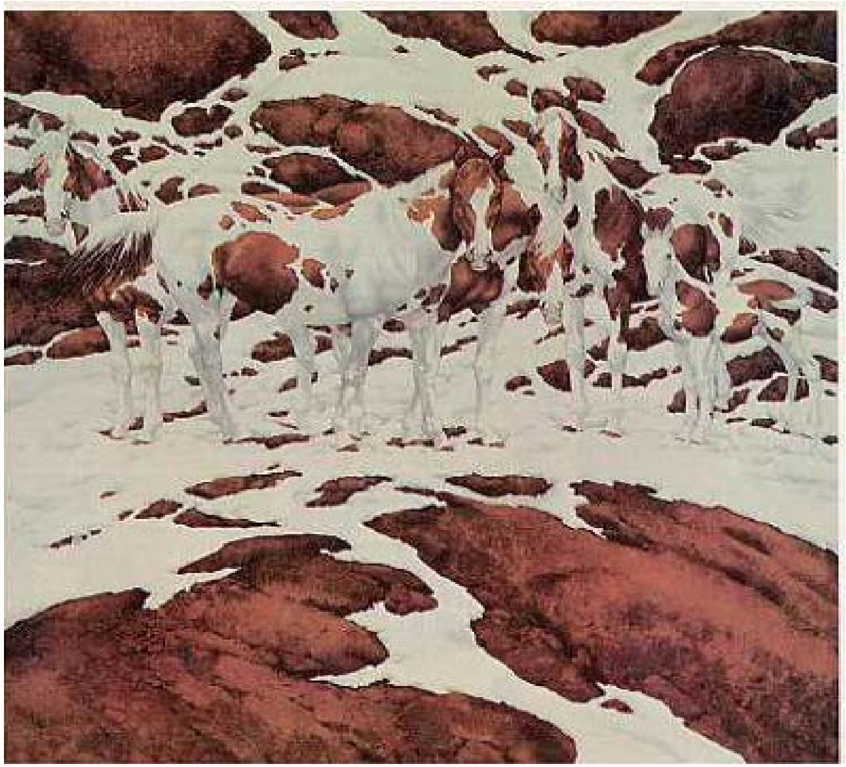
2018.10

Introduction

- Tens of thousands of times every day we identify or recognize objects around us.
- What is this?
  - fast recognition
  - Less than 100ms
- Is it an object or a background?
- how do we decide where one object ends and another begins?





- It is hard to recognize objects when objects are similar with background.

- **How do we manage to assign diverse stimuli to the same category when object recognition?**

- Chair with different visual properties (color, size, shape, material, et al)

Contents

- Perceptual Organization
- Theories of object recognition
- Brain processes involved in object recognition

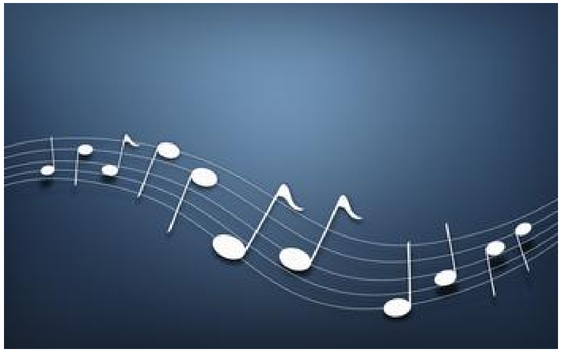
1. Perceptual Organization

- Visual environment is often complex and confusing, with many objects overlapping others and so hiding parts of them from view.

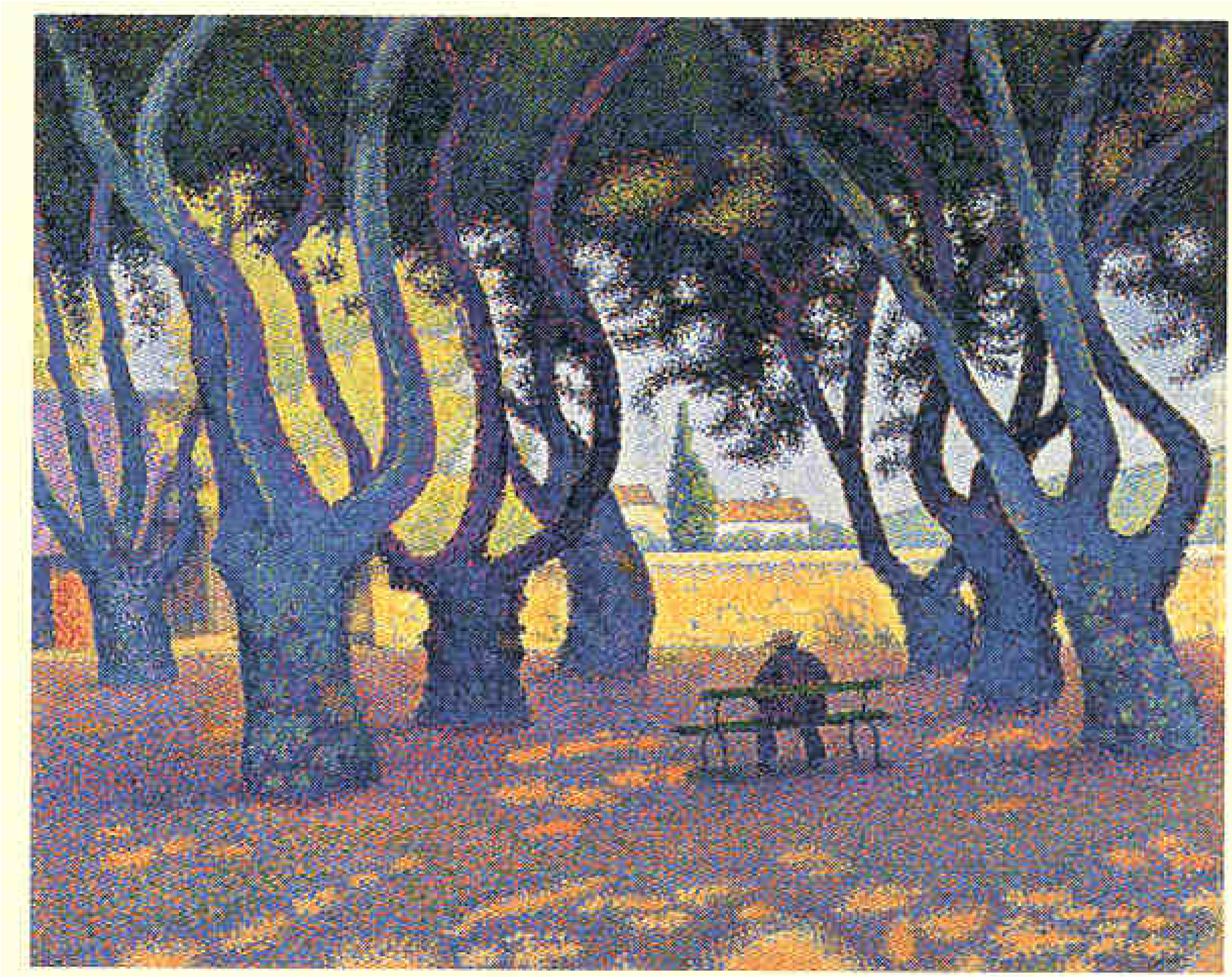
- The goal of understanding "perceptual organization" is to understand **how we put together the basic features to see a coherent, organized world of things and surfaces**

- The first systematic attempt to study perceptual segregation (perceptual organization) was made by the Gestaltists.
- They were German psychologists.
- Their fundamental principle was the **law of Prägnanz**, according to which we typically perceive the simplest possible organization.

Gestalt



- The basic motto of the Gestaltists was that the whole is greater than the sum of its parts.
- Examples:
  - A melody is composed of notes, but much more than notes (emotion, strength, power…) ;
  - the pointillist painters



Gestalt Psychology Principles

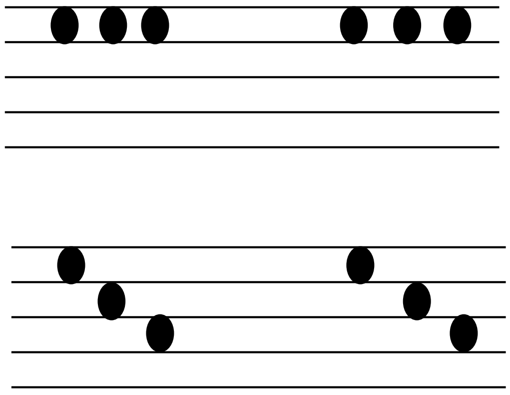
- *Proximity(接近性)*
- *Similarity(相似性)*
- *Common Fate (Common Direction)*
- *Orientation & symmetry*
- Good Continuation(连续性)
- Common region (Belongingness)
- Closure(封闭性)

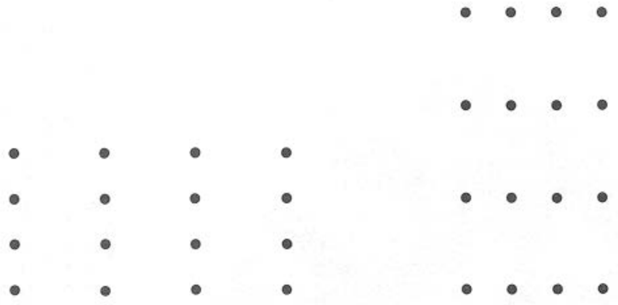
Bottom Up: Not Learned (primitive)

Top Down: Plastic, Learned (schema-driven)

(1)Proximity

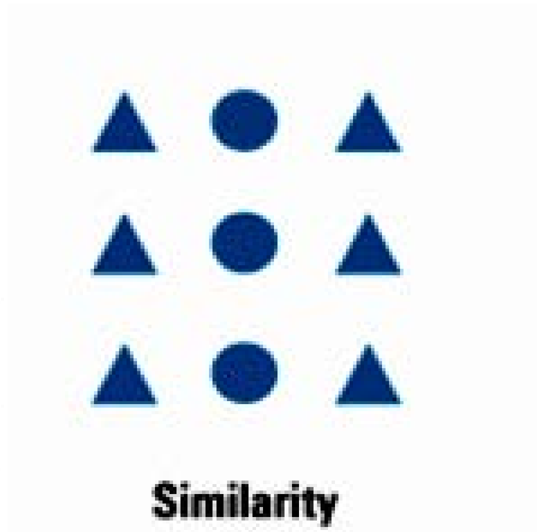
- In vision when elements in an image are **close** together they are perceived to be together and separate from others that are further away, even though they are similar

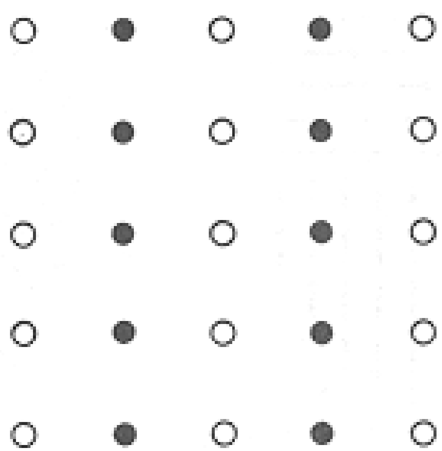


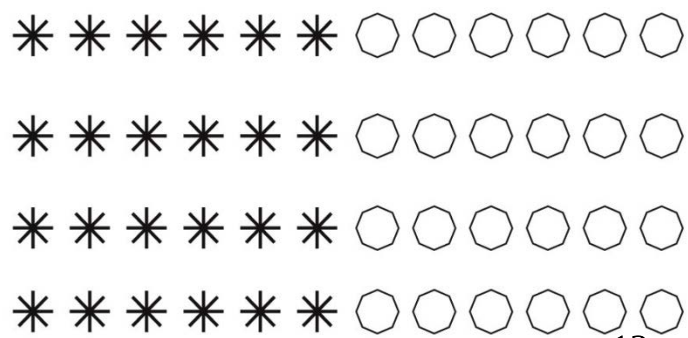


(2)Similarity

- Similar elements tend to be grouped together.

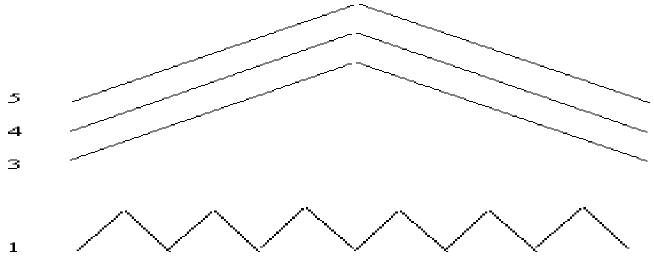


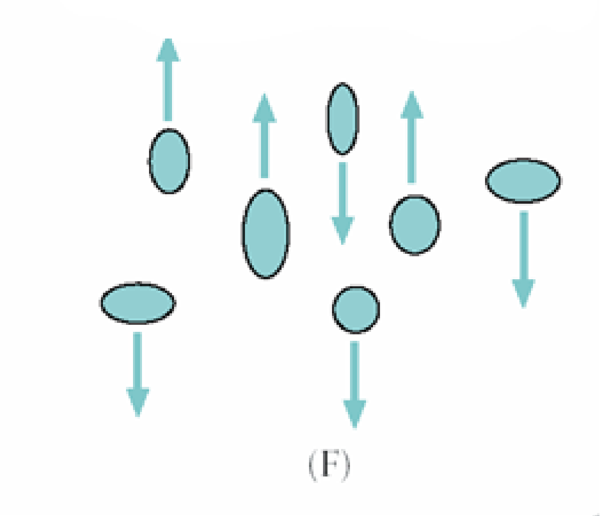




(3)Common Fate

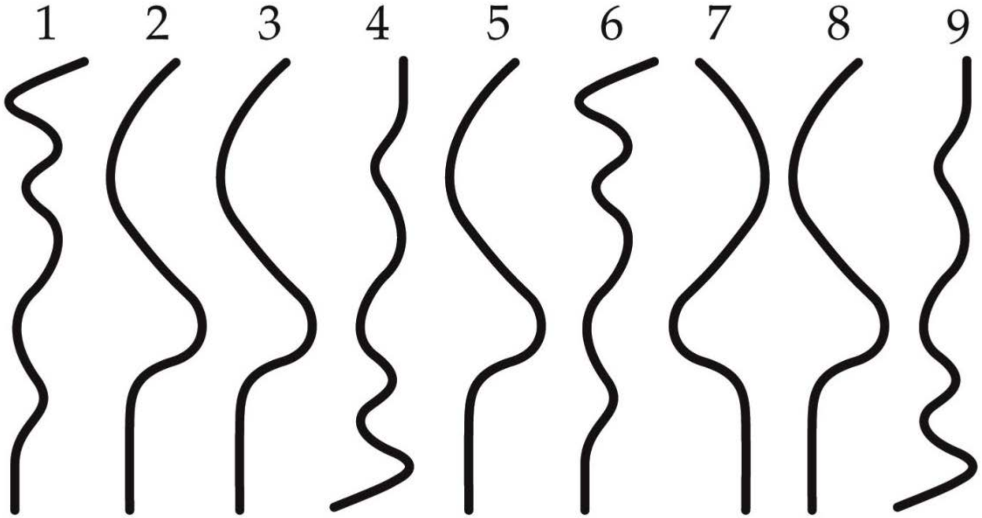
- Things tend to start and finish together
- Things that move together tend to group





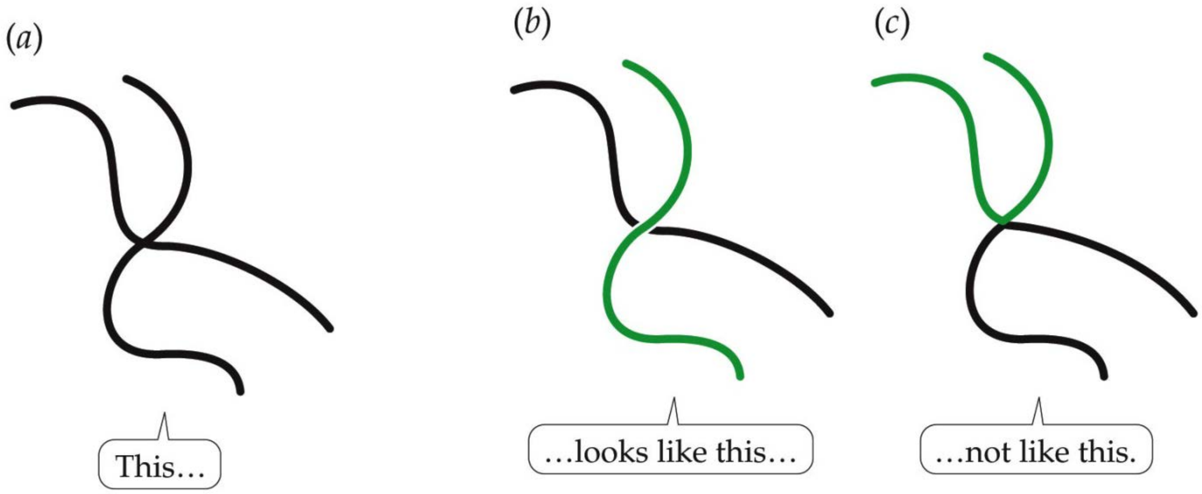
(4)Orientation & symmetry

- Objects oriented with horizontal and vertical axes, or ones that are symmetric, are more often perceived as figures



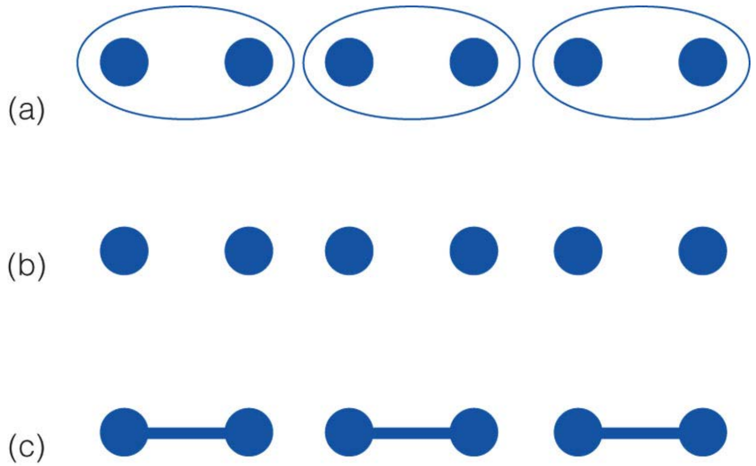
(5)Good Continuation

- Perceptual mechanisms tend to preserve smooth continuity in favor of abrupt edges
- Lines are seen as following the smoothest path



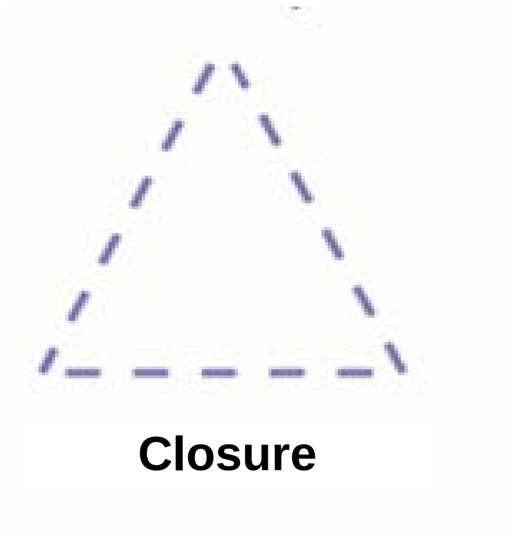
(6)Common region (Belongingness)

- Common region - elements in the same region tend to be grouped together
- Uniform connectedness - connected region of visual properties are perceived as single unit



(7)Closure

- Of several possible perceptual organizations, ones yielding "closed" figures are more likely than those yielding "open" ones.



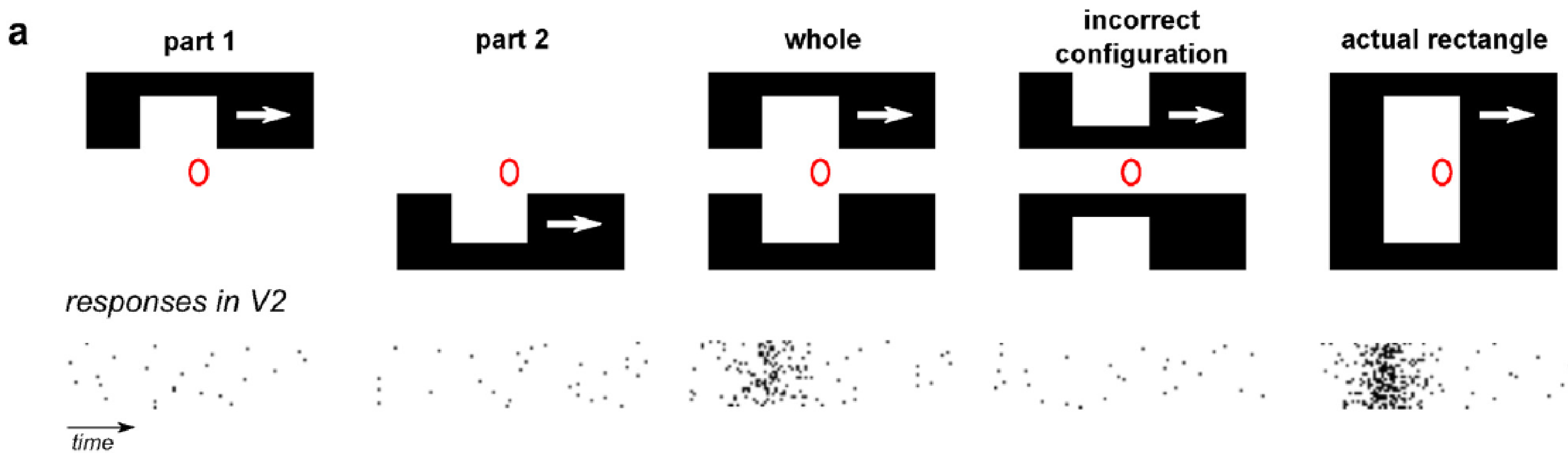
Triangle, not segments

The law of Pragnanz (Gestalt):

- Simple -- every stimulus is seen as simply as possible

- Meaningfulness or familiarity - things form groups if they appear familiar or meaningful

- the best, simplest, and most stable shape" (Koffka, 1935)



- **Neural responses in monkey visual areas V1 and V2. Neurons did not respond because the parts were outside the receptive field of the neuron. However, when the two parts were put together in the correct configuration, a robust neural firing in V2 was observed**
- **Information processing in the human brain is highly nonlinear.**

Perceptual Segregation

- The Gestaltists emphasized **figure–ground segregation in perceptual** organization.
- Figure-ground segregation - determining what part of environment is the figure so that it “stands out” from the background
- The principle of figure/ground is one of the most basic laws of perception

- Properties of figure and ground
  - **The figure is more “thinglike” and more memorable than ground**
  - **The figure is seen in front of the ground**
  - **The ground is more uniform and extends behind figure**
  - **The contour separating figure from ground belongs to the figure**

- The eye differentiates an object from its surrounding area. A form, or shape is naturally perceived as figure (object), while the surrounding area is perceived as ground (background).
- **Balancing figure and ground can make the perceived image more clear. Using unusual figure/ground relationships can add interest and subtlety to an image**.

Perceptual Segregation

- Reversible figure-ground
- Faces-goblet illusion

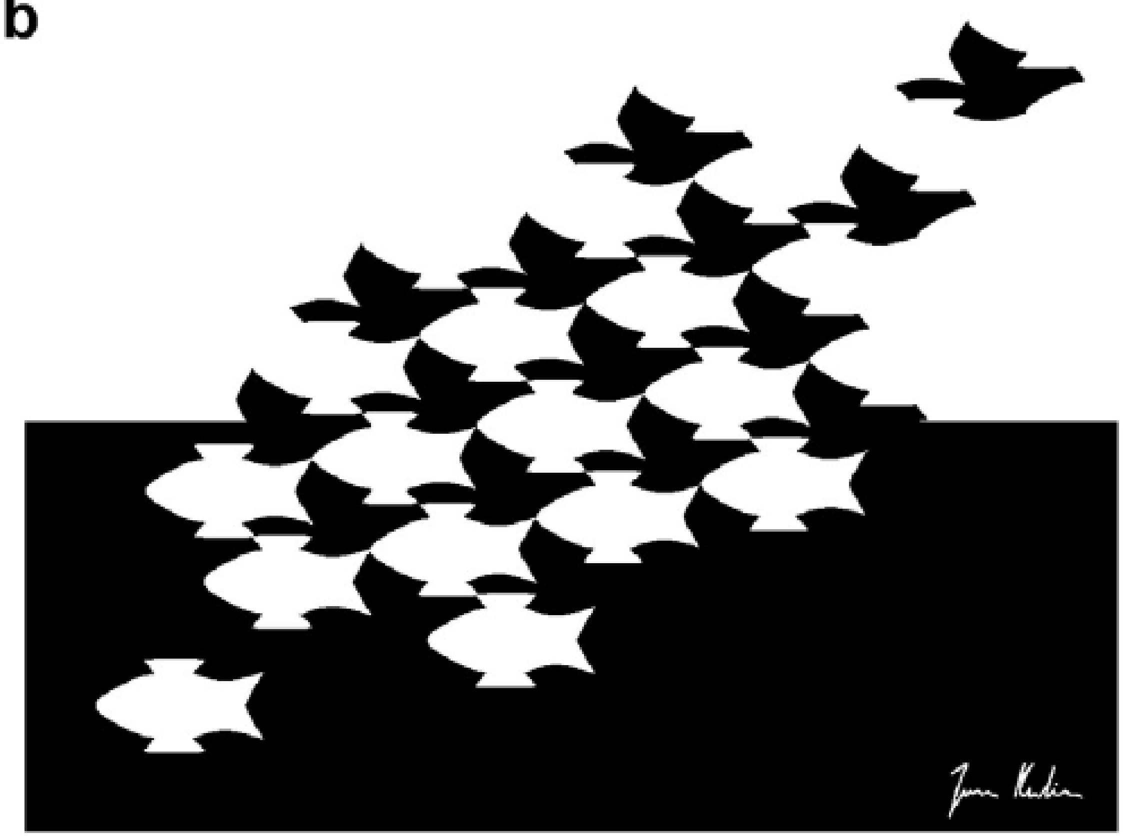
<u>Figure-ground segregation</u> - determining what part of environment is the figure so that it “stands out” from the background

reversible figure-ground objects：

Gestalt approach: perception is parsing (segmenting) image into figure and ground, which equals form perception

What is Figure = the thing, the important object, easier to remember

What is ground = not important, harder to remember b



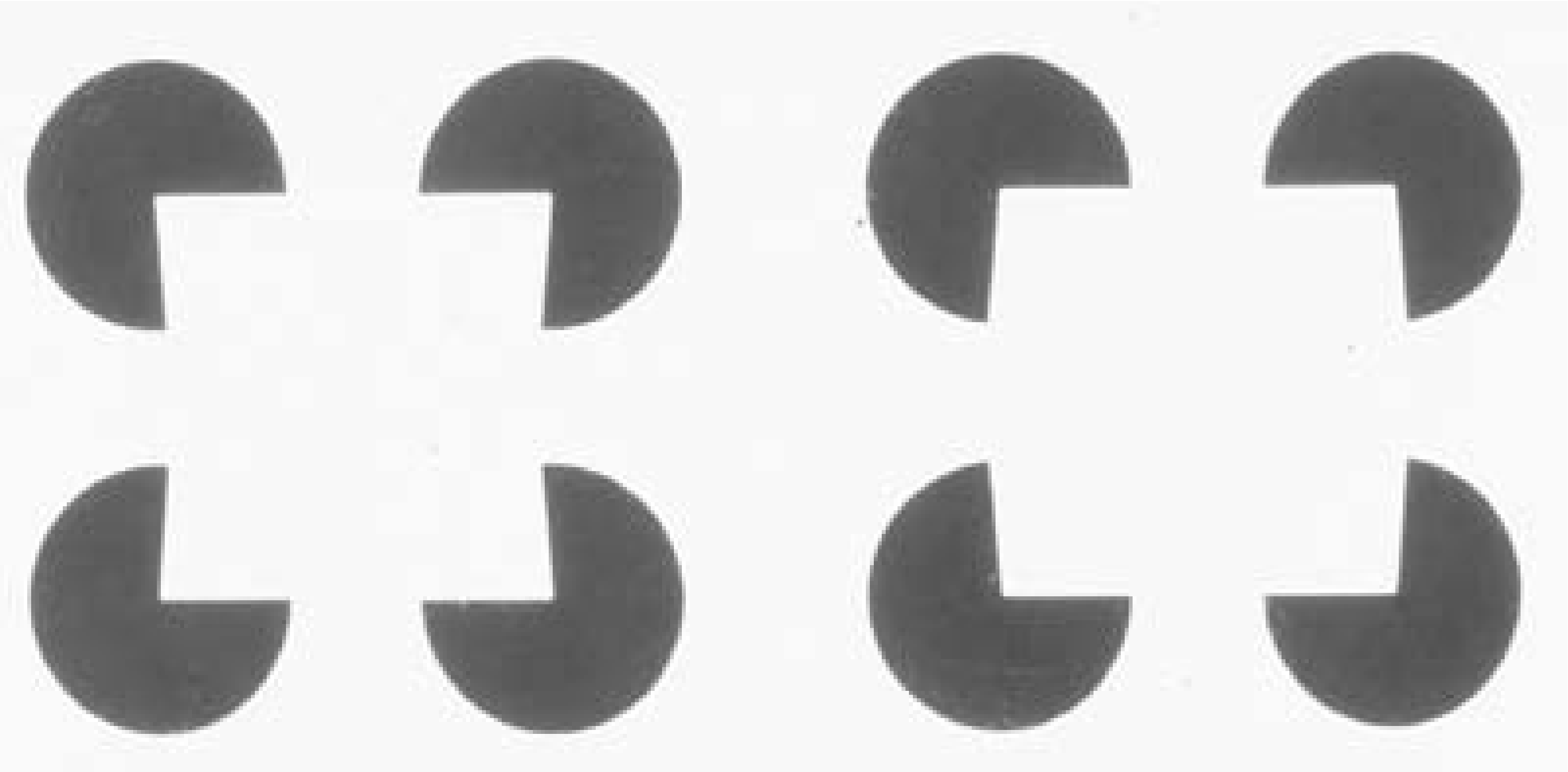
- An ambiguous figure-ground display. Although many people would see fish in the water and birds in the sky, in fact both animals are present in the water and in the sky but the color of background defines what is seen as a figure.

Ambiguous pictures 两可图

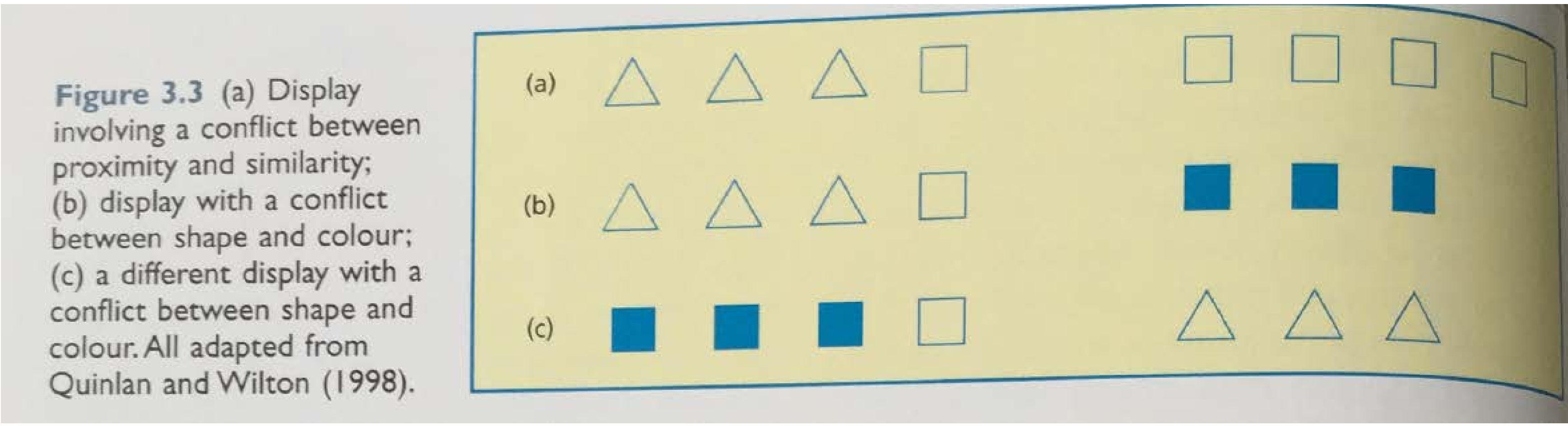
Illusions

- Sometimes, your visual system must group/segment an object, not based on edges in the retinal image, but by inferring parts of the missing boundaries

Illusory contour. Note that there is no physical contour for most of the perceived "edge"



What happens when different laws of organization are in conflict?



- Mainly group or cluster on the basis of proximity
- Additional processes are used if within-cluster mismatch
- Proximity, similarity, colour…

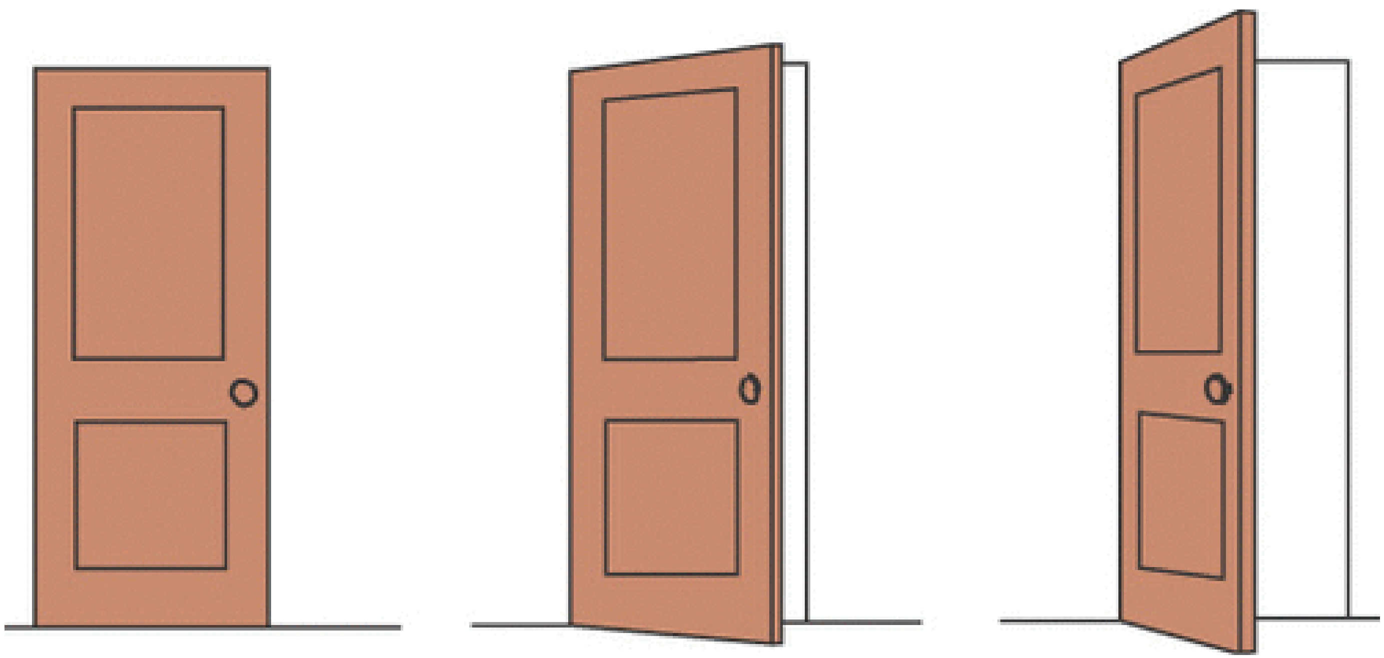
PERCEPTUAL CONSTANCY

- How we recognize objects without being deceived by changes in their:

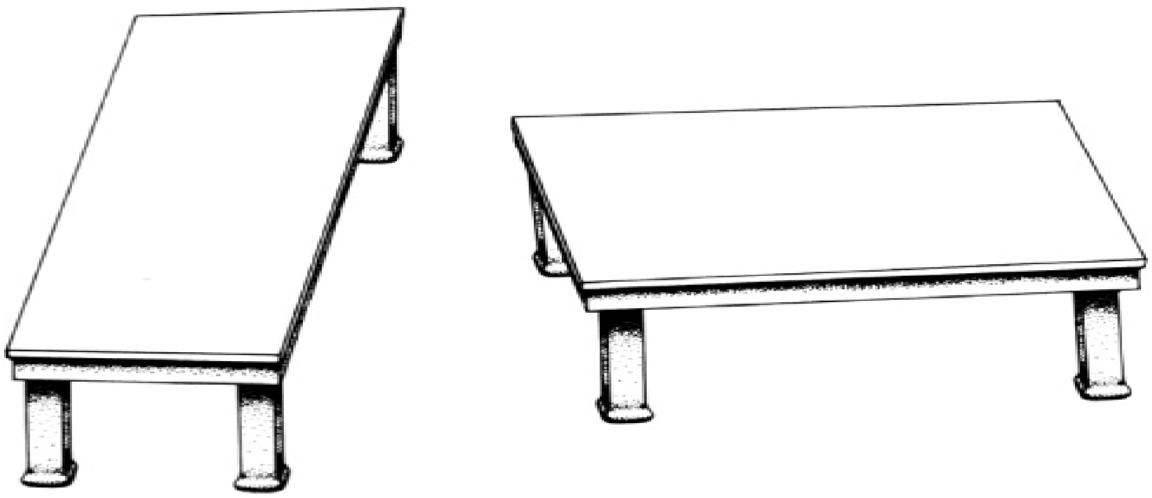
  - Shape
  - Size
  - Brightness
  - Color

Shape Constancy

Shape Constancy is the tendency to perceive an object as having the same shape regardless of its orientation or the angle from which we view it.

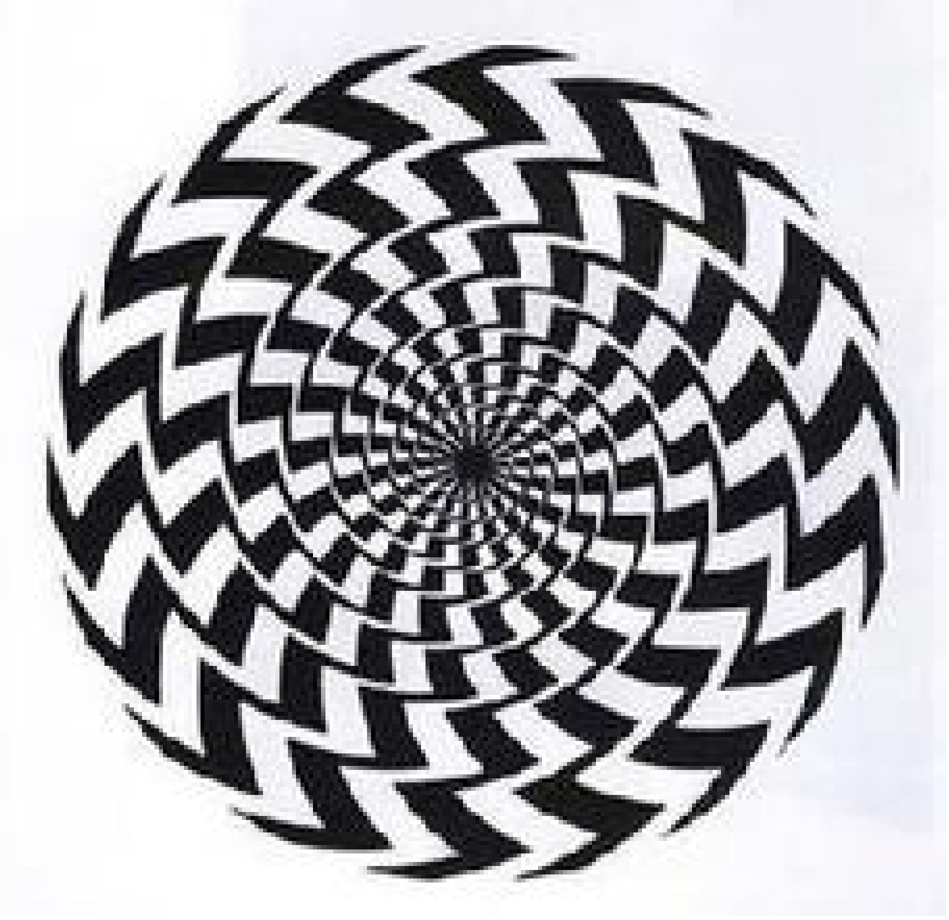


Shape Constancy



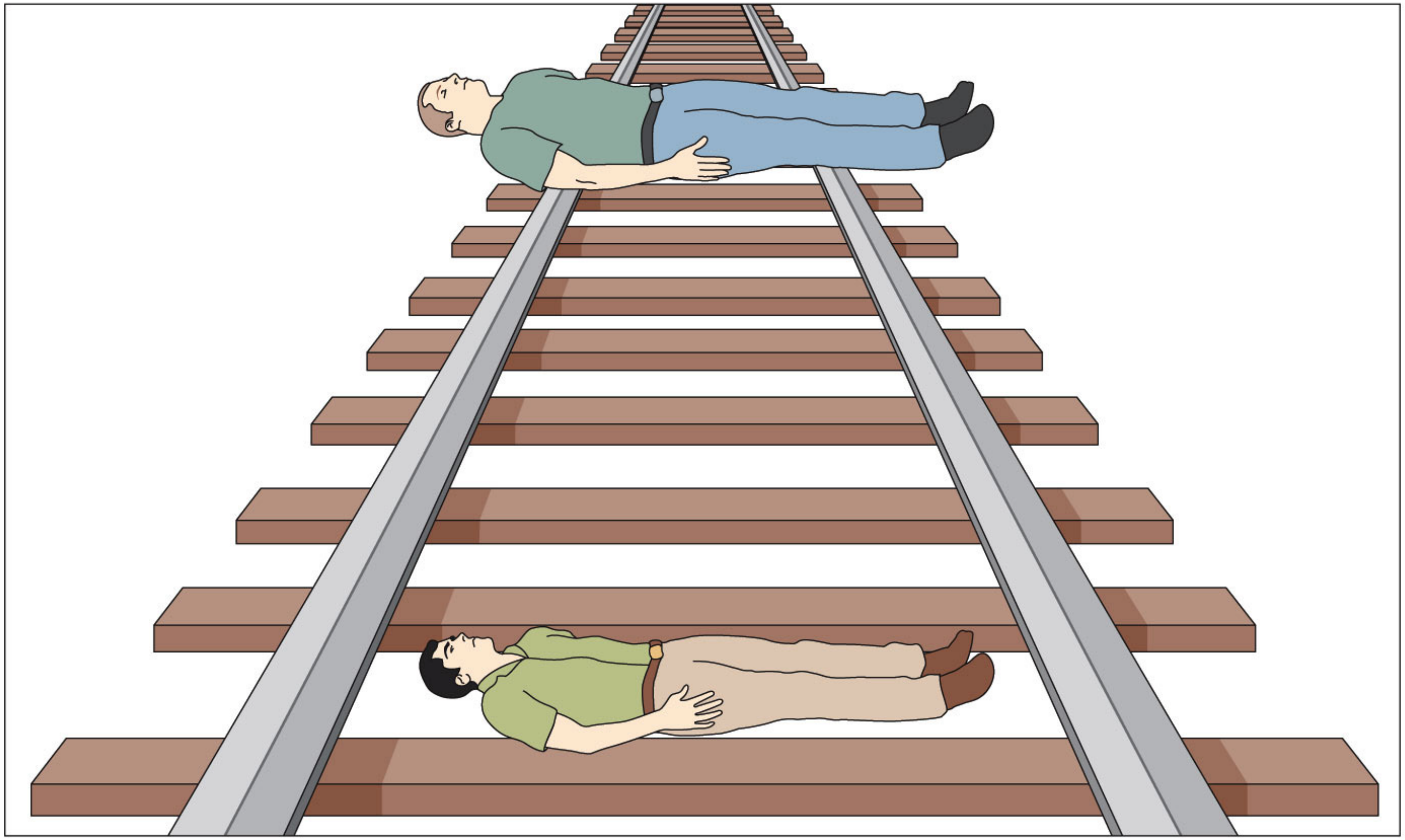
- Our brain compensates for the distortion of the shape by taking into account visual cues about distance and depth to keep our perception of the frame constant.
- How compensate for the distortion ?

- But we can be fooled



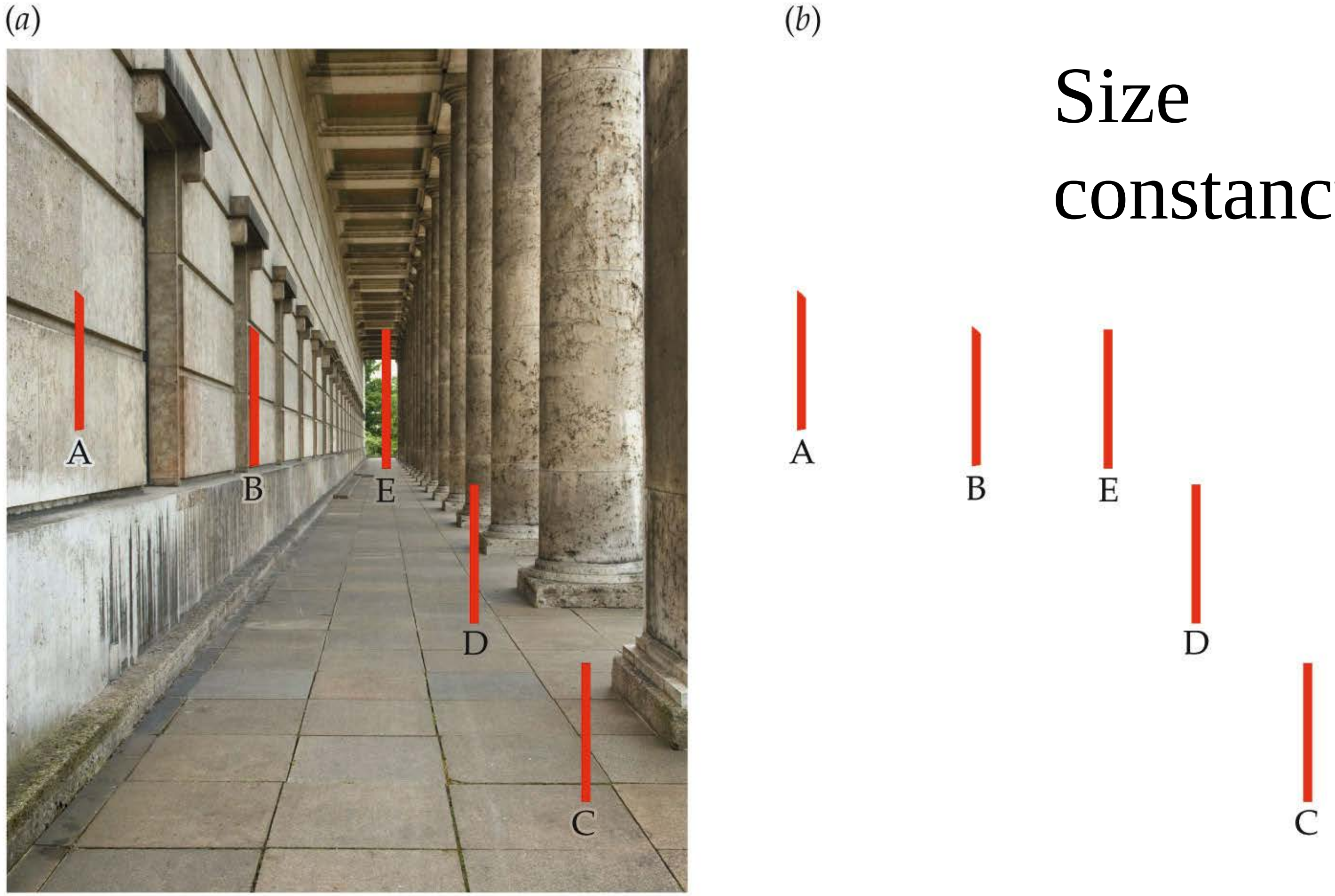
The line is
Spirality
Or Circle?

Size Constancy



The two figures lying across these train tracks are the same size in the image

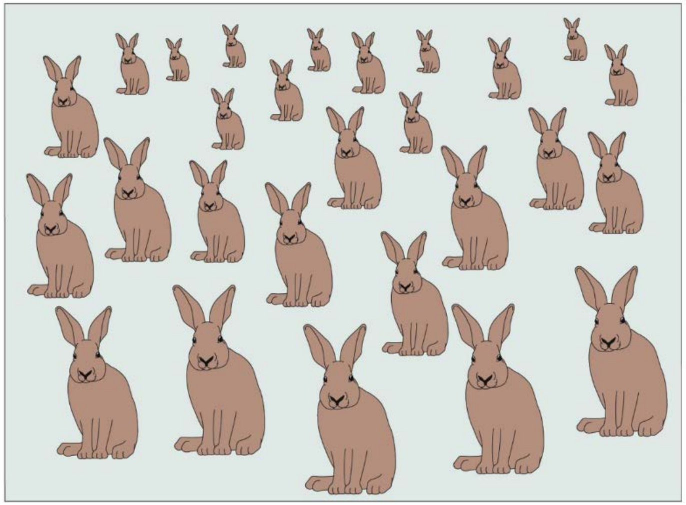
All of the red lines in this illustration are the same length



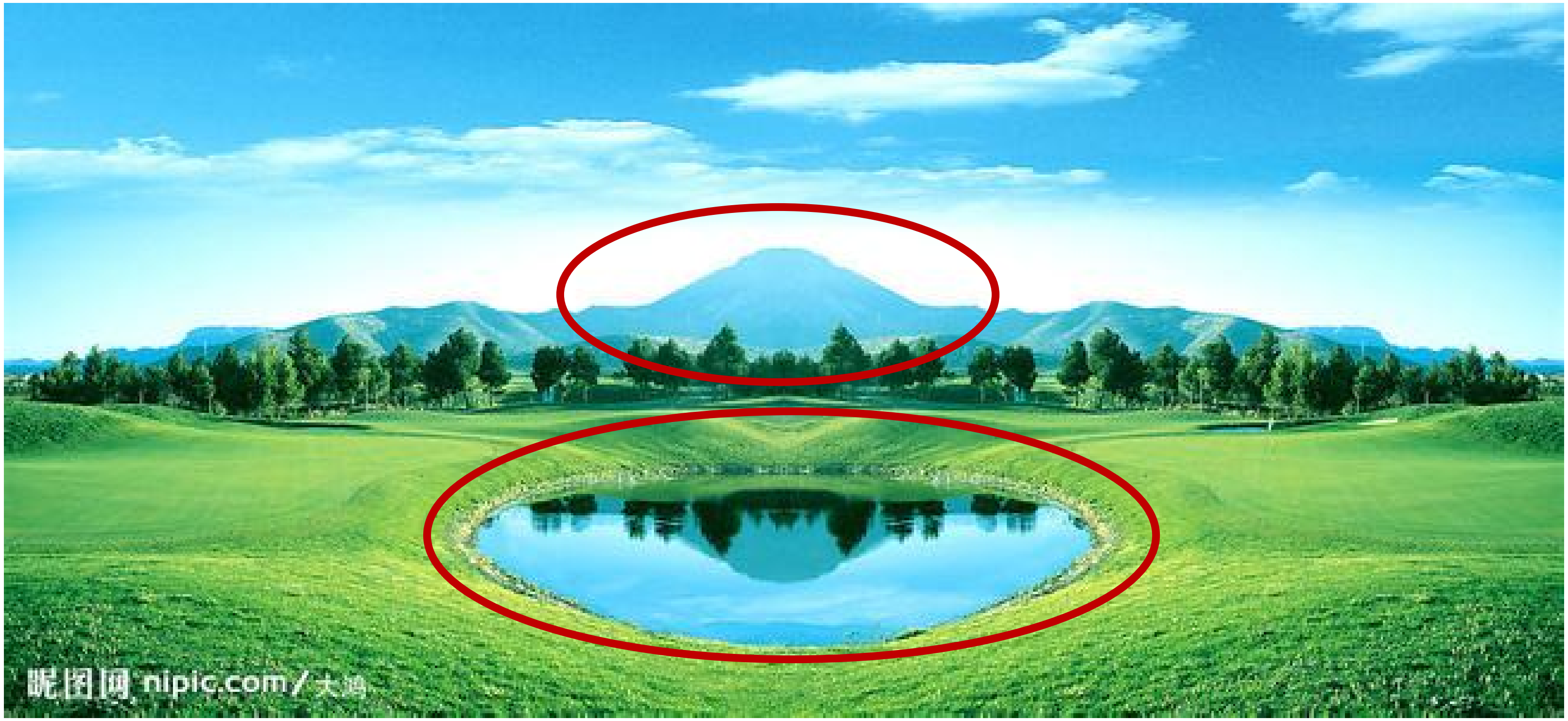
Size constancy

Size constancy

- Size constancy is the tendency for any given object to appear the same size whether its size in the retinal image is large or small.
- Size perception and size constancy depend mainly on perceived distance.
- Several other factors, such as horizon, scene complexity, familiar size, and purposeful interactions, also contribute to size judgements.



Size Constancy



Neural mechanism evidence?

The representation of perceived angular size in human primary visual cortex, 2006, Nat Neurosci

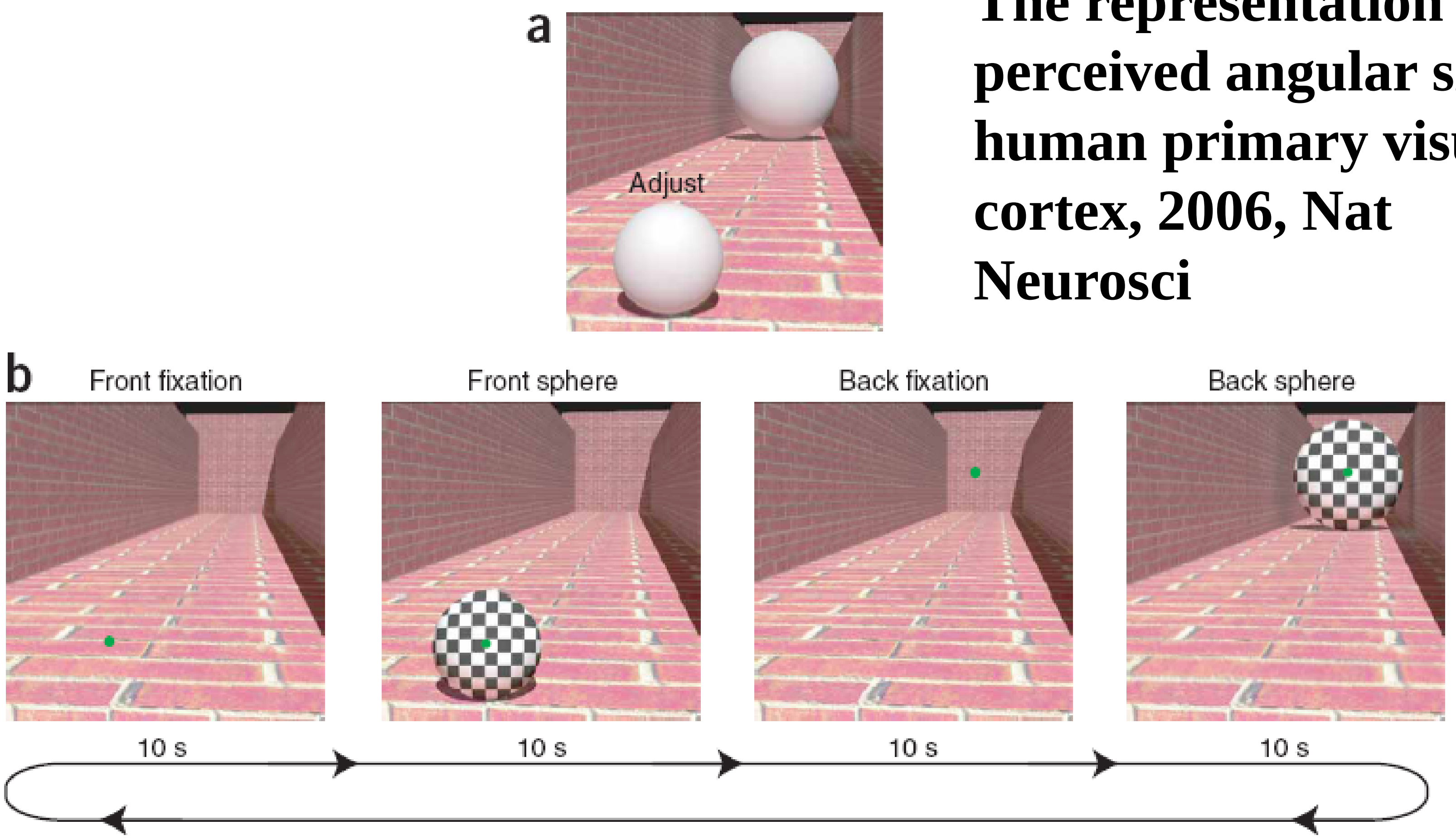
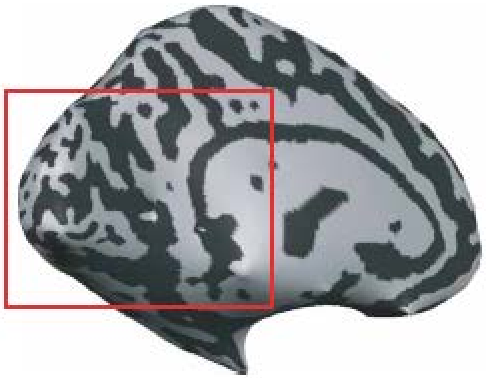


Figure 1 Stimulus for the behavioral and fMRI experiments. (**a**) In the behavioral experiment, subjects were asked to adjust the front sphere to match the angular size of the back sphere so that the two images of the spheres would overlap perfectly if they were moved to the same location on the screen. A larger-scale demonstration of the stimulus and size illusion is available at http://faculty.washington.edu/somurray/sizedemo.html. (**b**) A schematic of the experimental design used in the fMRI experiment. Subjects maintained fixation on a small green dot. The spheres were rendered with counterphase-flickering checkerboard patterns. Each condition was presented in succession for 10 s and then repeated five times in each scan.

a



b

c

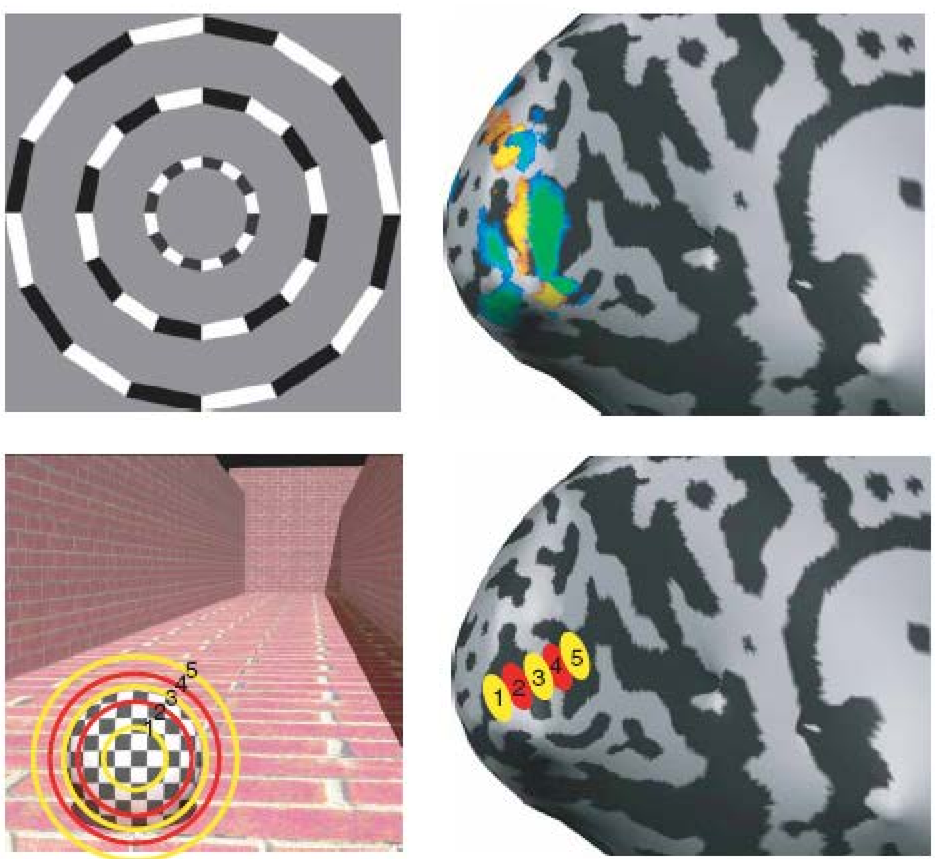


Figure 2 ROIs. fMRI data were analyzed in V1 in five adjacent ROIs of equal size extending along the eccentricity dimension. (**a**–**c**) In **a**, the area shown in greater detail in **b** and **c**. Counterphase-flickering annuli (**b**, left) were used at varying eccentricities to define the ROIs. The cortical area (**b**, right) activated by three such annuli that correspond to positions 1, 3 and 5 in **c**. The blue/green colormap represents responses to the inner annulus (leftward activation on the inflated cortical representation) and outer annulus (rightward activation on the inflated cortical representation). The red/yellow colormap represents responses to the middle annulus. (**c**) The five ROI positions and their approximate mapping onto the stimulus.

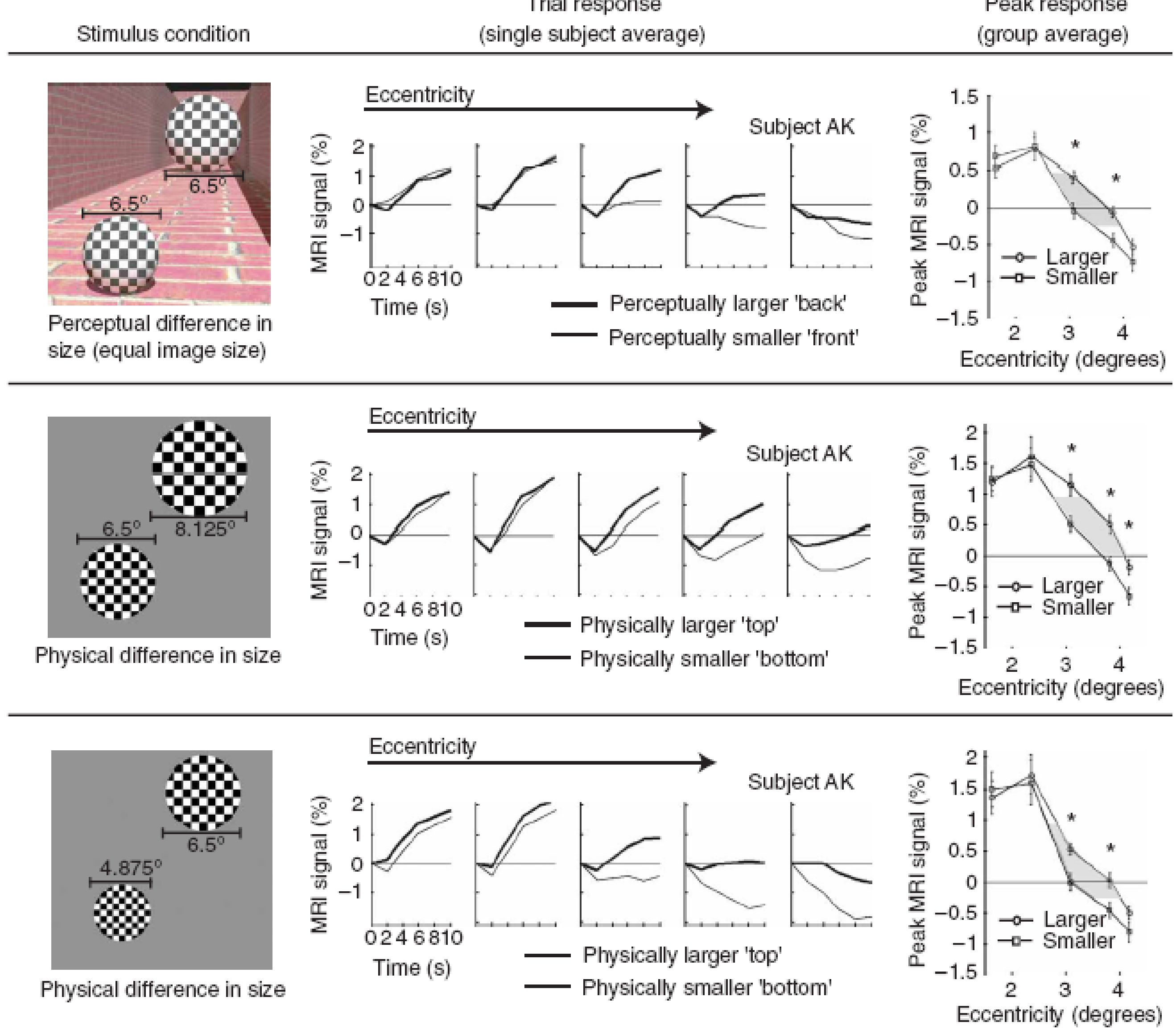


Figure 3 Imaging results. For each stimulus condition, event-related averages in V1 were calculated for each of five subjects (example subject, middle column), where time = 0 is the start of a checkerboard presentation. Peak measurements were averaged across hemispheres and subjects (right column). For both a perceptual difference in angular size (top row) and an actual difference in angular size (middle and bottom rows), activity for the larger stimulus extended in eccentricity beyond that for the smaller stimulus. The shaded region between the peak response curves (right column) shows the range over which differences in the curves were calculated (see main text and Methods). *$P < 0.05$.

DISCUSSION

We have demonstrated a relationship between the spatial extent of activation in human primary visual cortex and an object's perceived angular size. Previously it was assumed that retinotopic maps are fixed and solely reflect a feedforward mapping between the retina and cortex, such that when two images occupy the same portion of the visual field, they will activate the same cortical area. When the information to be represented is 2D (as in almost all previous retinotopic mapping studies), this characterization seems to be valid. But adding a third dimension to the information (such as exists in almost all naturally occurring stimuli) complicates this characterization. The ultimate goal of the visual system is clearly not to precisely measure the size of an image projected onto the retina. A more behaviorally critical property of an object is its size relative to the environment, which helps determine its identity and how one should interact with the object. Estimating an object's behaviorally relevant size requires that its retinal projection be scaled by its distance from the observer or from other objects in the environment. The results presented here indicate that the scaling process affects the retinotopic representation in V1, one of the earliest stages of the human visual system. On the basis of the current results, we suggest that the topographic representation in V1 should be thought of as dynamic and dependent on the 3D content in the scene. It remains to

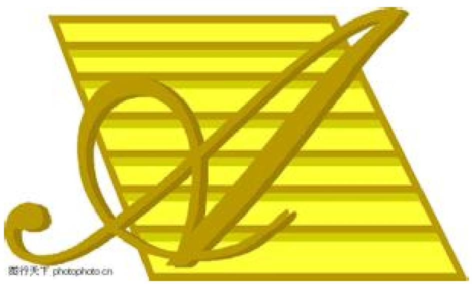
Color & Light Constancy

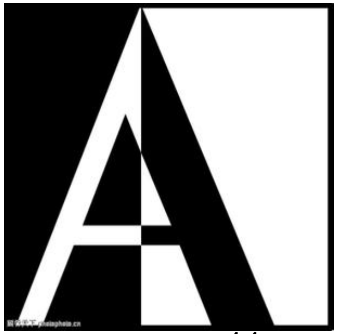
- Chromatic adaptation

- Cells in V4 , V1,V2,V3 may play roles in colour constancy.

2. Theories of object recognition

- We spend much of our time (e.g., when reading) engaged in pattern recognition
- A key issue here is the *flexibility of the human perceptual system.*
- For example, we can recognize the letter "A" rapidly and accurately across large variations in orientation, typeface, size, and writing style.



Why is pattern recognition so successful?

Theories of object recognition

- **Template-matching**
- **Feature analysis**
- **Recognition-by-components theory**

(1)Template-matching

- According to template theories, we have templates (forms or patterns stored in long-term memory) corresponding to each of the visual patterns we know.
- A pattern is recognized on the basis of which template provides the closest match to the stimulus input.
- Simple
- However, it isn't very realistic in view of the enormous variations in visual stimuli allegedly matching the same template.


AVATAR
昵图网 www.nipic.com
BY: qq951357
NO:20100604075008066019

Improvement to the basic template theory

- assume that the visual stimulus undergoes a normalization process. This process produces an internal representation of the visual stimulus in a standard position (e.g., upright), size, and so on *before* the search for a matching template begins.

- assume that there is more than one template for each letter and digit. This would permit accurate matching of stimulus and template across a wider range of stimuli, but at the cost of making the theory more complex.

Matching-Models

Weaknesses:

- **Memory limit**
- **3D world**
- **New objects**

Power:

- Simple
- Image processing about object recognition

(2) Feature analysis

- According to feature theories, a pattern consists of a set of specific features or attributes
- Instead of looking for a match to a standard shape, feature analysis emphasizes that what makes an 'A' is the letter's unique set of defining features
- For example, key features of the letter “A” are two straight lines and a connected cross-bar.
- This kind of theoretical approach has the advantage that visual stimuli varying greatly in size, orientation, and minor details can be identified as instances of the same pattern.

- The feature-theory approach has been supported by studies of visual search in which a target letter has to be identified as rapidly as possible.

- Neisser (1964) compared the time taken to detect the letter “Z” when the distractor letters consisted of straight lines (e.g., W, V) or contained rounded features (e.g., O, G) (see Figure 2.4

detect the letter "Z" as rapidly as possible

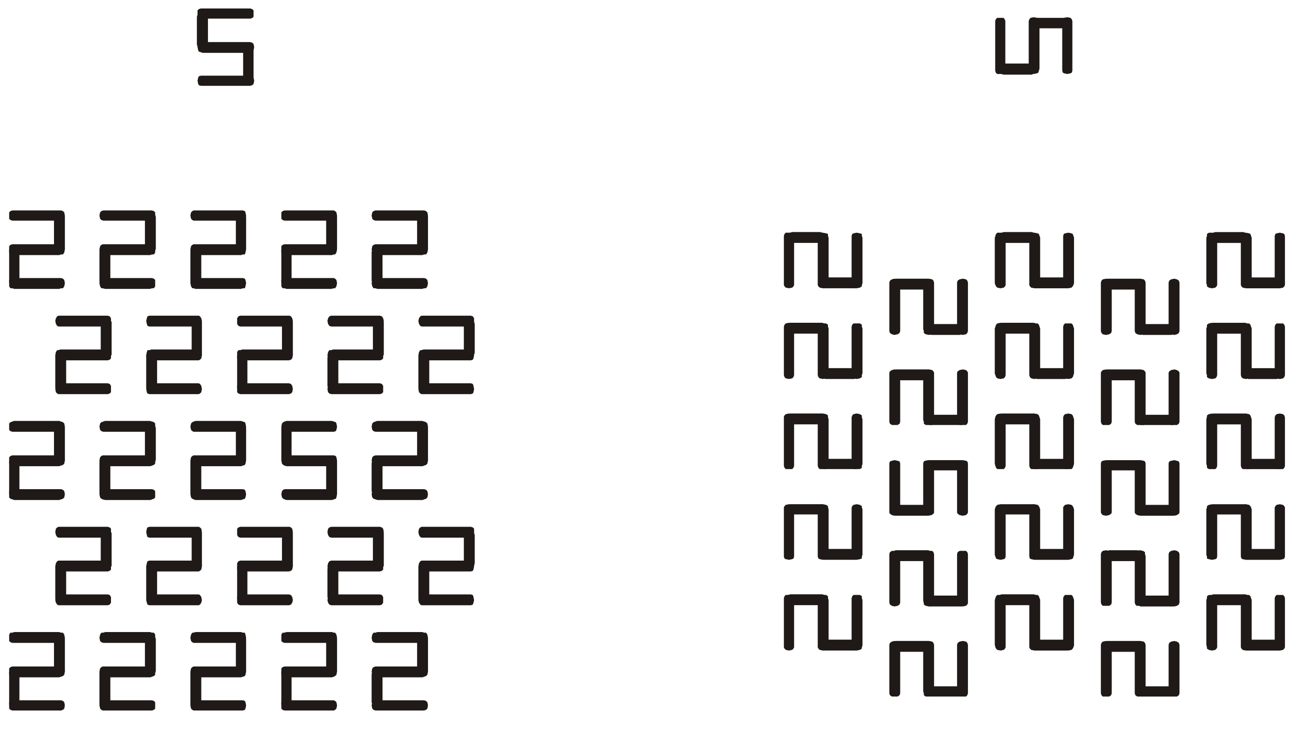
LIST 1
IMVXEW
WVMEIX
VXWIEM
MIEWVX
WEIMXV
IWVXEM
IXEZVW
VWEMXI
MIVEWX
WXEIMV

LIST 2
ODUGQR
GRODUQ
DUROQG
RGOUDQ
RQGOUD
UGQDRO
GUQZOR
ODGRUQ
DRUQGO
UQGORD

LIST 1	*LIST 2*
IMVXEW	ODUGQR
WVMEIX	GRODUQ
VXWIEM	DUROQG
MIEWVX	RGOUDQ
WEIMXV	RQGOUD
IWVXEM	UGQDRO
IXEZVW	GUQZOR
VWEMXI	ODGRUQ
MIVEWX	DRUQGO
WXEIMV	UQGORD

Figure 2.4 Illustrative lists to study letter search: The distractors in List 2 share fewer features with the target letter Z than do the distractors in List 1.

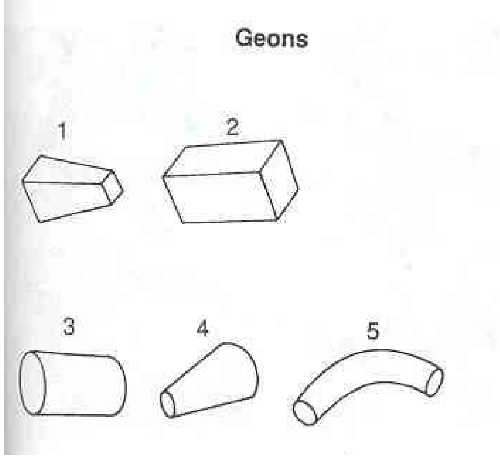
Performance was faster in the list2 because the distractors shared fewer features with the target letter Z.



feature-theory cannot explain some phenomenons.

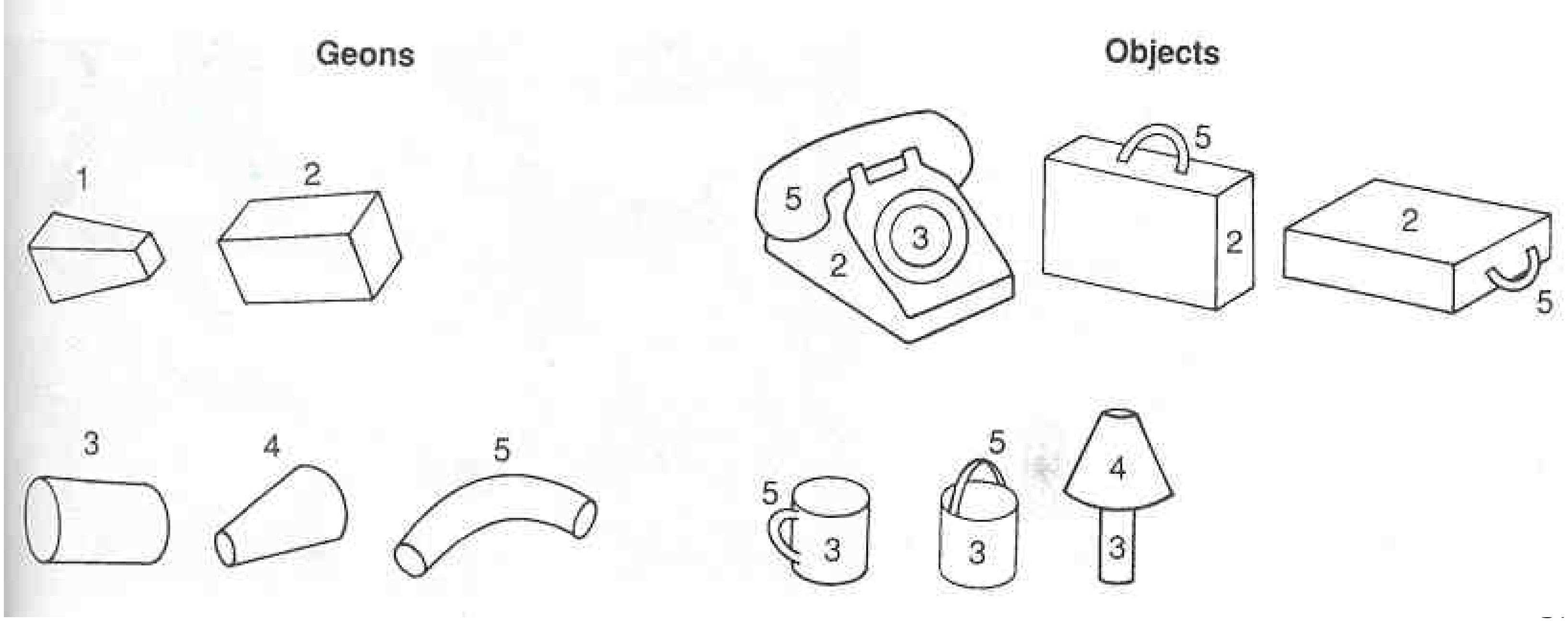
(3) **recognition-by-components theory--** Building Blocks of Recognition

- *What processes are involved in object recognition?*
- An influential answer was provided by Irving Biederman (1987) in his recognition-by-components theory.
- He argued that objects consist of basic shapes or components known as **geons (geometric** ions).
- Examples of geons are blocks, cylinders, spheres, arcs, and wedges.

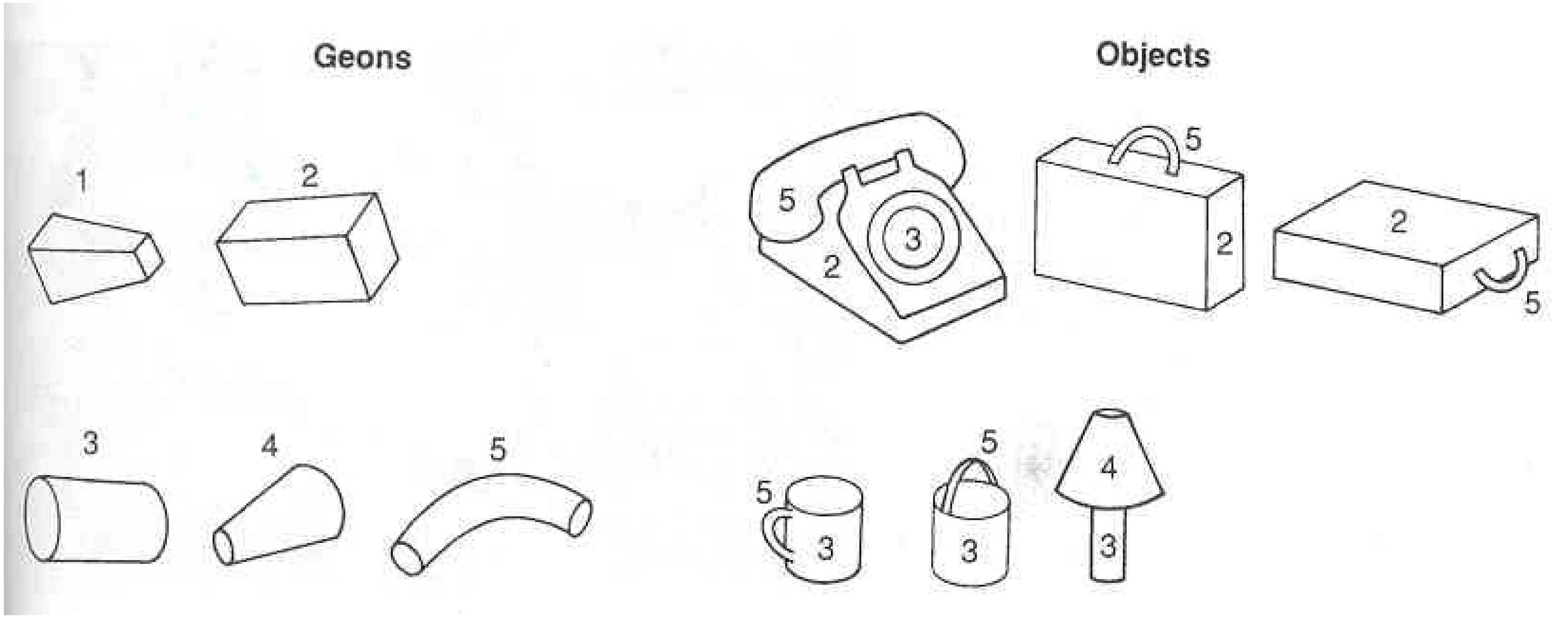


Geons (Biederman)

- According to geon theory, complex objects are made up of arrangements of basic component parts ('geons‘)



- ***How many geons are there?***
- *According to Biederman (1987), there are about* 36 different geons. That may sound suspiciously few to provide descriptions of all the objects we can recognize and identify.
- However, we can identity enormous numbers of spoken English words even though there are only about 44 phonemes (basic sounds) in the English language. This is because they can be arranged in almost limitless combinations.
- The same is true of geons – the reason for the richness of the object descriptions provided by geons stems from the different possible spatial relationships among them.
- For example, a cup can be described by an arc connected to the side of a cylinder



Geons
Objects
1
2
3
4
5
5
3
2
5
2
2
5
5
3
5
3
4
3

◆Geon-based information about common objects is stored in long-term memory.

◆As a result, object recognition depends crucially on the identification of geons.

◆Of major importance, an object's geons can be identified from numerous viewpoints. Thus, object recognition is equally rapid and easy regardless of the angle from which an object is viewed.

◆ In other words, it is viewpoint-invariant.

DOES VIEWPOINT AFFECT OBJECT RECOGNITION?

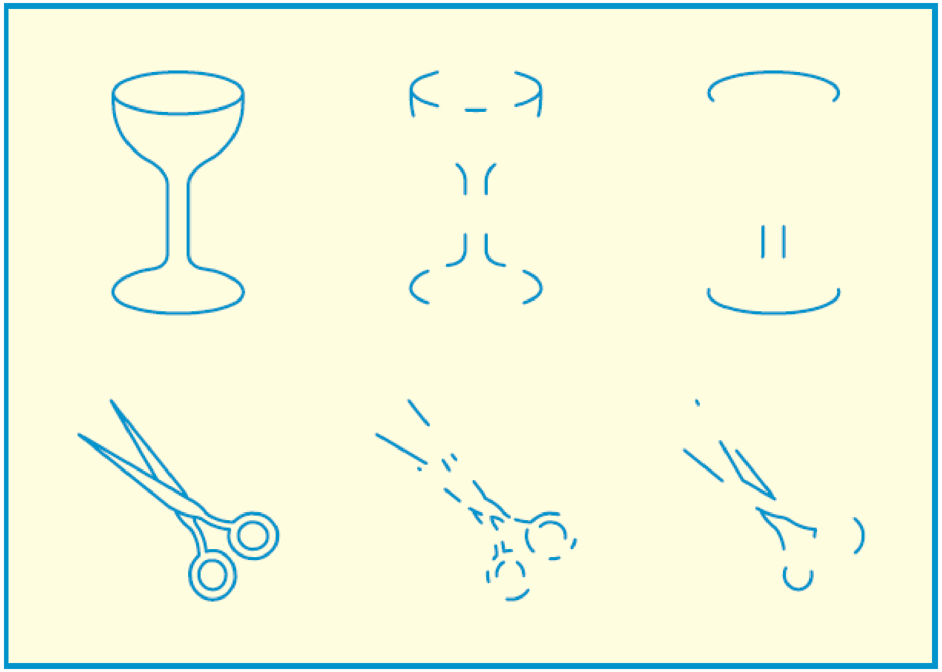
- Friedman et al. (2005) argue that object recognition is generally faster and easier when objects are seen from certain angles. Such theorists favor the view that object recognition is viewpoint-dependent.
- Performance was close to viewpoint-invariant when the object recognition task was easy. However, it was viewpoint-dependent when the task was difficult (e.g., no feedback provided).
- Face recognition is typically strongly viewpoint dependent




◆*How do we recognize objects when only some of the relevant visual information is available?*

◆According to Biederman (1987), the concavities (hollows) in an object's contour provide especially useful information.

◆Object recognition was much harder to achieve when parts of the contour providing information about concavities were omitted than when other parts of the contour were deleted.



◆Recognition-by-components theory strongly emphasizes bottom-up processes in object recognition.

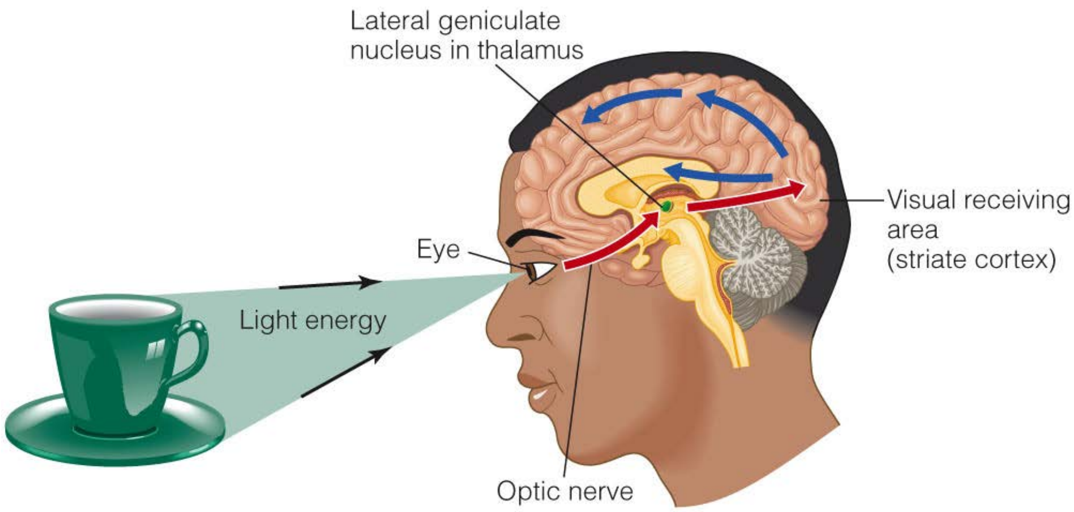
◆However, top-down processes depending on factors such as expectation and knowledge are often important, especially when object recognition is difficult.

◆For example, Viggiano et al. (2008) found that observers relied more on top-down processes when animal photographs were blurred than when they weren’t blurred.

Theories of object recognition

- **Template-matching**
- **Feature analysis**
- **recognition-by-components theory**
- **Other theory?**

3 brain processes involved in object recognition

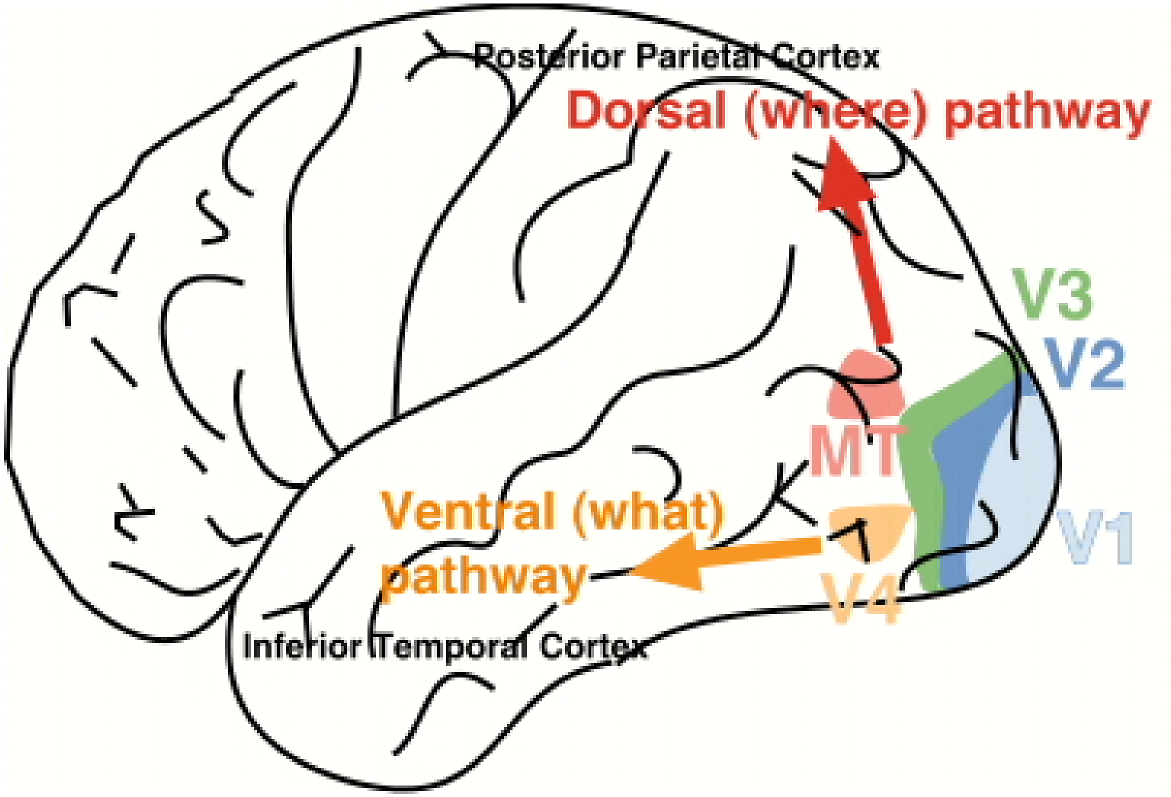


Side view of the visual system, showing the three major sites: the eye, the lateral geniculate nucleus, and the visual cortex.

Pathway from Retina to Cortex

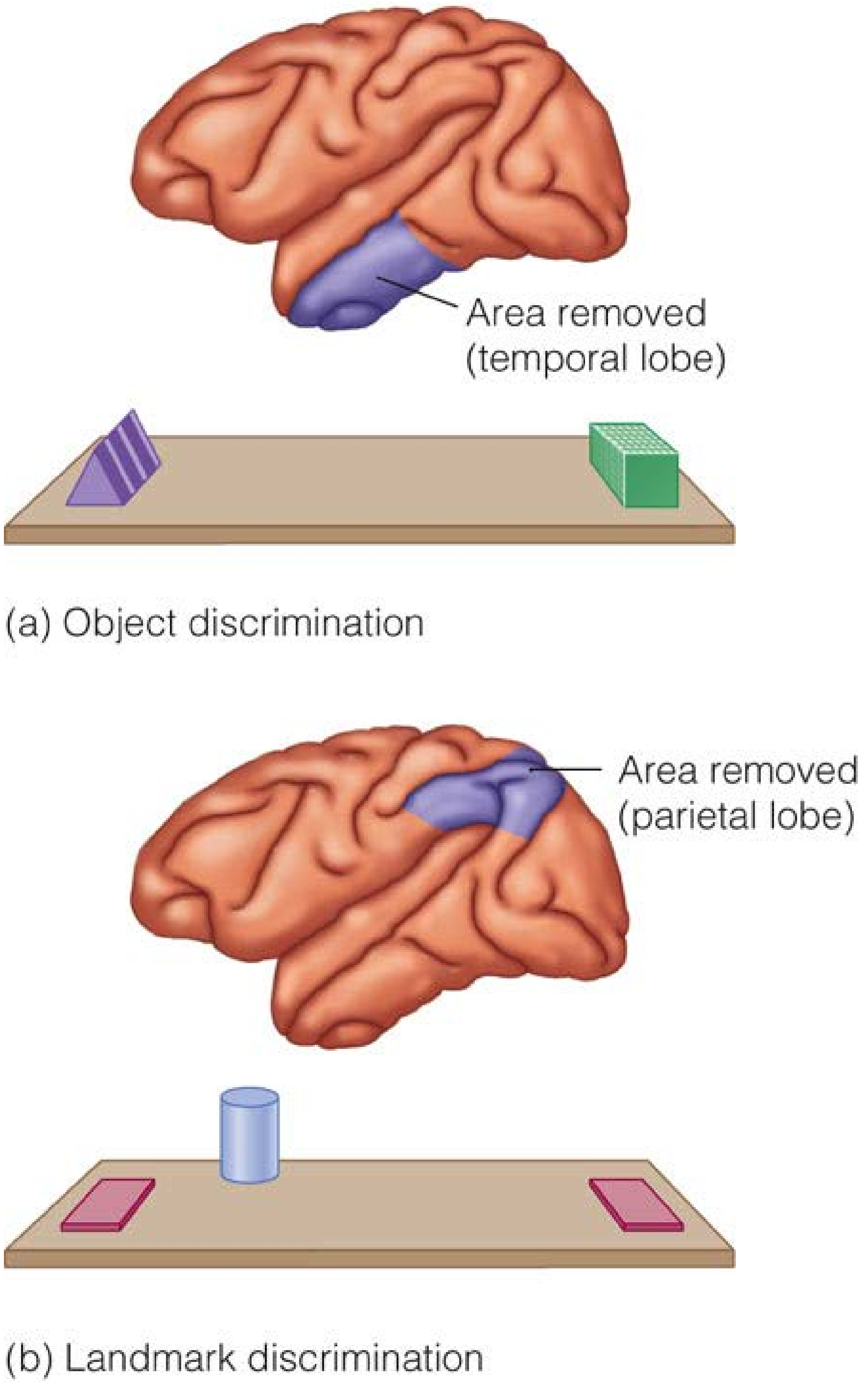
- Signals from the retina travel through the optic nerve to the
  - Lateral geniculate nucleus (LGN)
  - Primary visual cortex (the striate cortex)
  - And then through two pathways to the temporal lobe and the parietal lobe

What and Where Pathways



What and Where Pathways

- Ungerleider and Mishkin (1983)
  - Object discrimination
    - Monkey is shown an object
    - Then presented with two choice task
    - Reward given for detecting the target
  - Landmark discrimination problem
    - Monkey is trained to pick the food well next to a cylinder



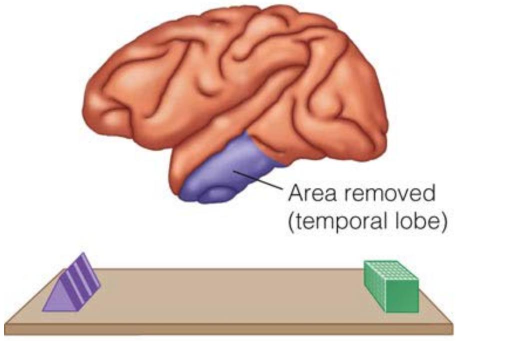
What and Where Pathways - continued

- Ungerleider and Mishkin (cont.)
  - Using ablation/lesion, part of the parietal lobe was removed from half the monkeys and part of the temporal lobe was removed from the other half
  - Retesting the monkeys showed that:
    - Removal of temporal lobe tissue resulted in problems with the object discrimination task - What pathway

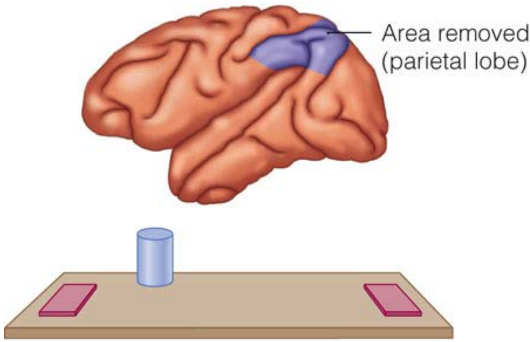
    (I see it, but don't know what it is)

    - Removal of parietal lobe tissue resulted in problems with the landmark discrimination task - Where pathway

    (I know what it is, but I don't know where it is)



(a) Object discrimination



(b) Landmark discrimination

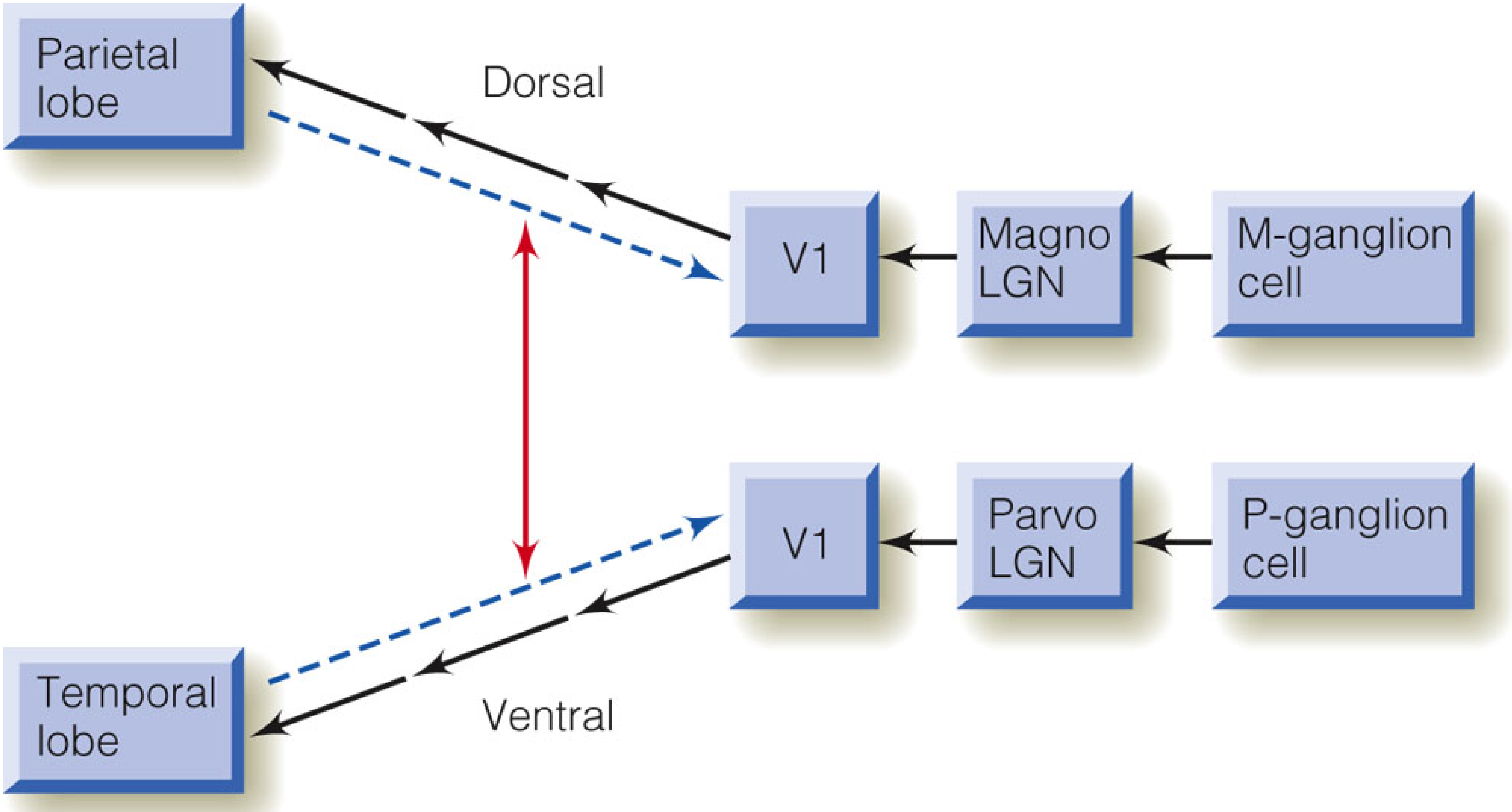
What and Where Pathways - continued

- Both pathways originate in retina
  - Ventral pathway begins in small or medium ganglion cells
    - Called P-cells
    - Axons synapse in layers 3, 4, 5, & 6 of LGN
    - Called parvocellular layers

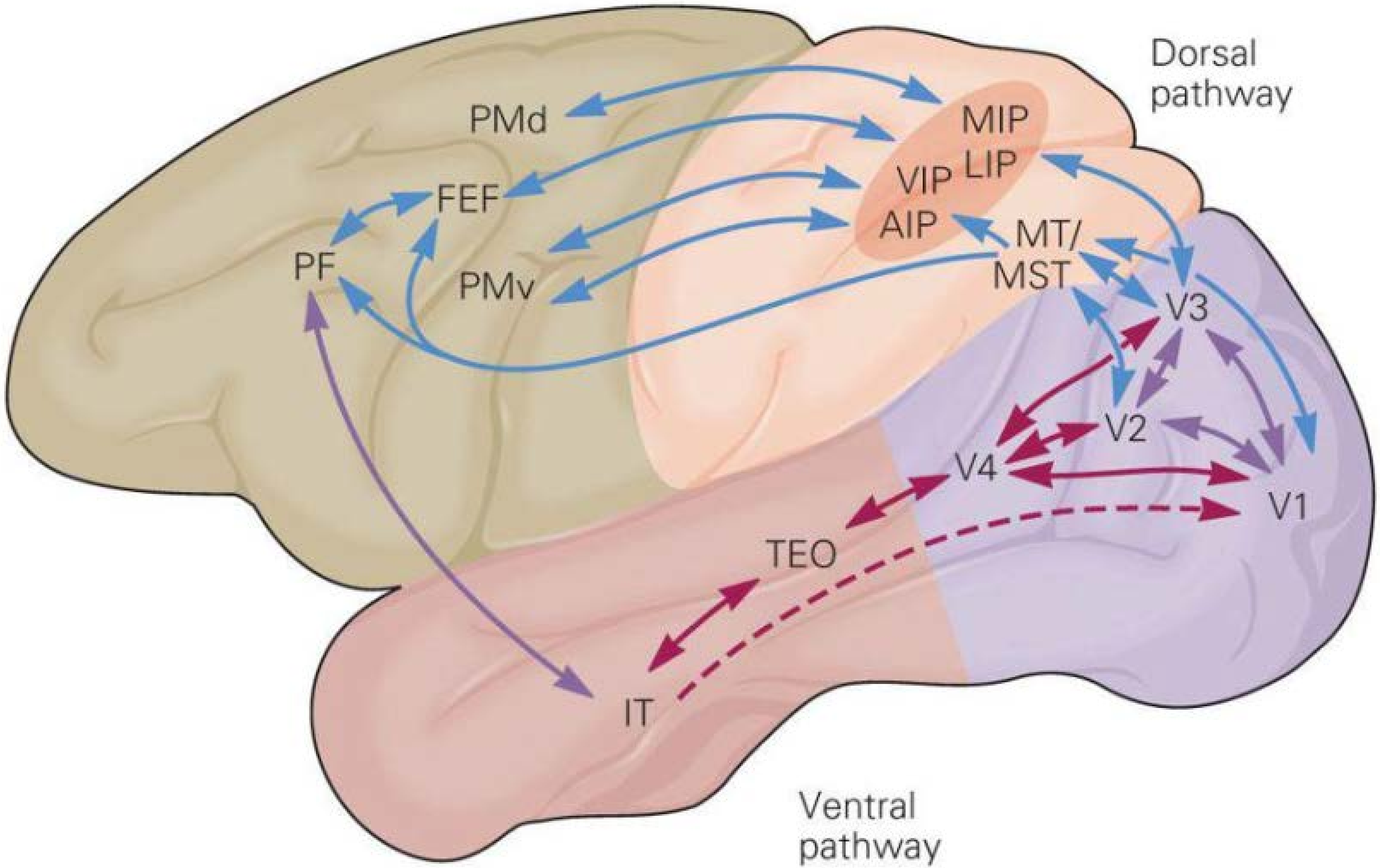
What and Where Pathways - continued

- Dorsal pathway begins in large ganglion cells
    - Called M-cells
    - Axons synapse in layers 1 & 2 of LGN
    - Called magnocellular layers

- Ablation(部分切除) research with monkeys shows:
  - Parvo channels send color, texture, shape and depth information
  - Magno channels send motion information



The dorsal and ventral streams in the cortex originate with the magno and parvo ganglion cells and the magno and parvo layers of the LGN. The red arrow represents connections between the streams. The dashed blue arrows represent feedback - signals that flow “backward.”



Dorsal
pathway
PMd
MIP
LIP
FEF
VIP
AIP
MT/
MST
PF
PMv
V3
V2
V4
V1
TEO
IT
Ventral
pathway

What and Where Pathways - patient

- Where pathway lesion

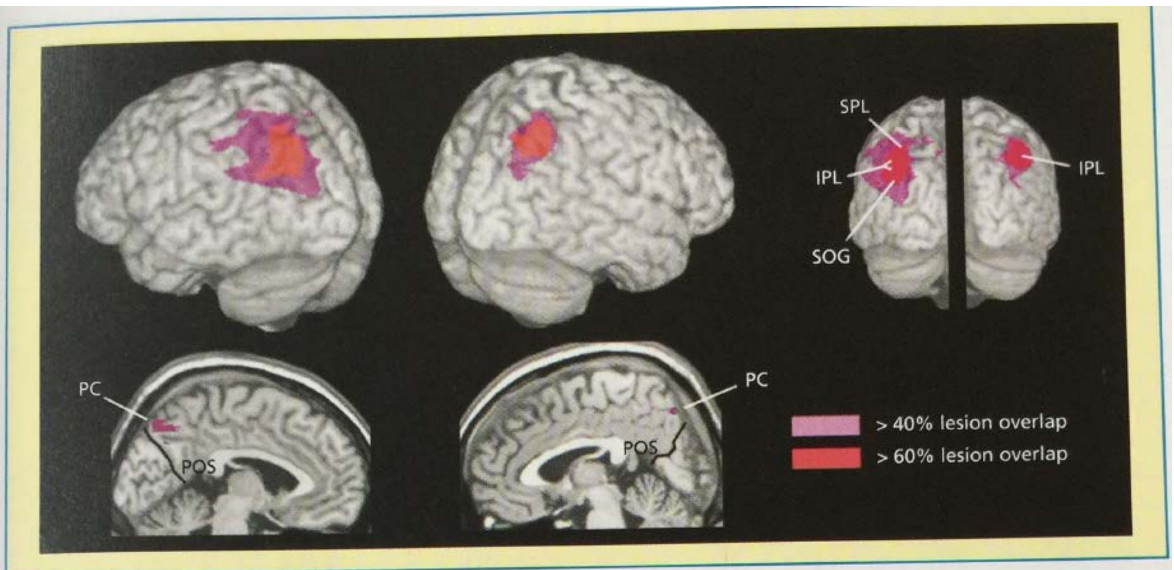


Figure 2.8 Percentage of overlapping lesions (areas of brain damage) in patients with optic ataxia (SPL = superior parietal lobule; IPL = inferior parietal lobule; SOG = superior occipital gyrus; Pc = precuneus; POS = parieto-occipital sulcus). From Karnath and Perenin (2005), by permission of Oxford University Press.

Akinetopsia is a condition in which stationary objects are generally perceived fairly normally but motion perception is often deficient. Free-flowing liquids, for example, can appear to be frozen, which can make a simple task, such as pouring a glass of water, very difficult.

What and Where Pathways - patient

- What pathway lesion- Agnosia

- DV: Agnosia 失认症

Summary

- Perceptual Organization
  - Gestalt Psychology Principles
  - Perceptual Segregation
  - Perceptual Constancy
- Theories of object recognition
  - Template-matching
  - Feature analysis
  - recognition-by-components theory
- Brain processes involved in object recognition
  - What Pathway
  - Where Pathway