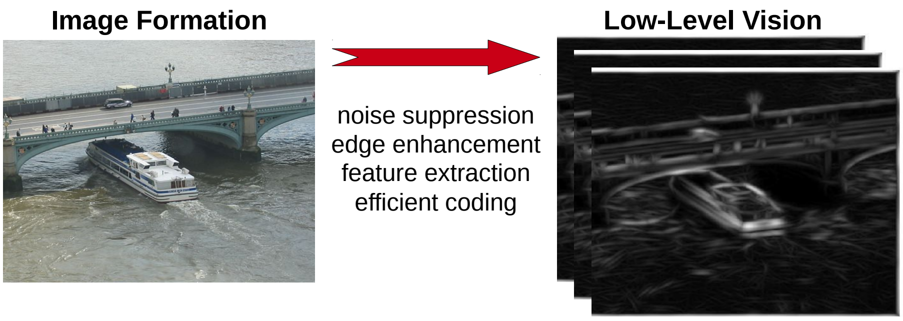
Computer Vision (7CCSMCVI / 6CCS3COV)

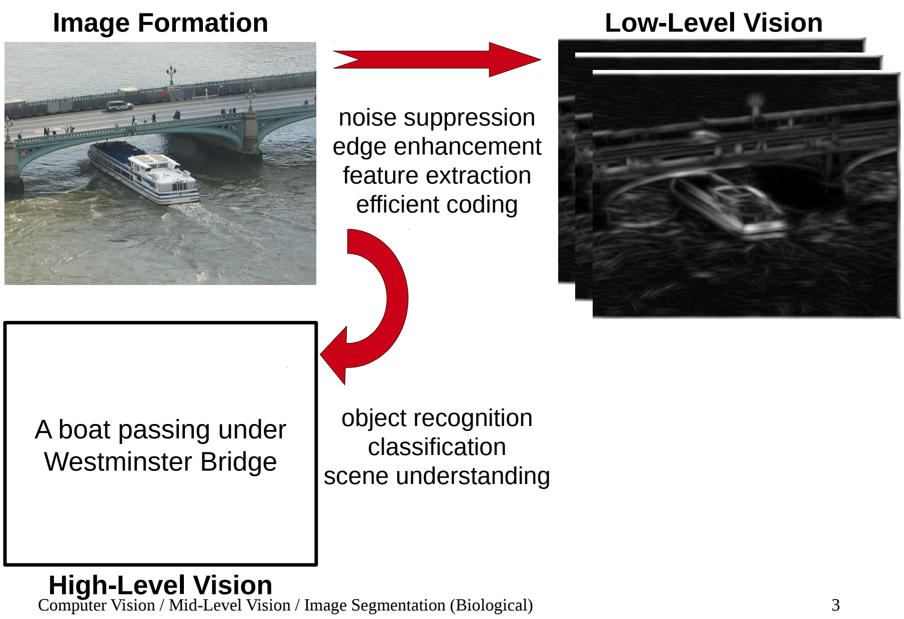
Recap

- **Image formation**
- **Low-level vision**
  - Artificial
    - Convolution:
      smoothing, differencing, edge detection
    - Multi-scale image representations:
      Gaussian and Laplacian image pyramids
  - Biological
    - V1
      - cRFs selective for colour, **orientation**, direction of motion, spatial frequency, eye of origin, binocular disparity, position
      - Gabor functions provide good model of orientation selectivity (convolution of an image with a Gabor performs edge detection)
      - ncRF allow neurons to be sensitive to contextual influences
- **Mid-level vision** ←Today
- **High-level vision**

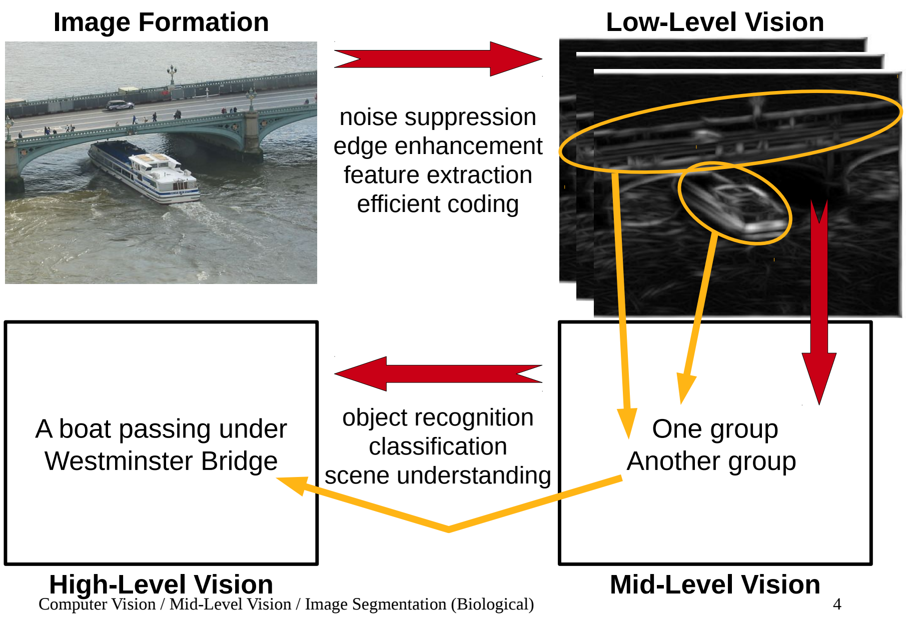
Recap in pictures



Recap in pictures



Mid-Level Vision



Mid-Level Vision

The role of Mid-Level vision is to:

- group together those elements* of an image that "belong together", and to
- segment (differentiate) these elements* from all others.

*elements (or tokens) can be whatever we need to group (pixel intensities, pixel colours, edges, features, etc.)

Grouping and segmentation are "two sides of the same coin": doing one implies doing the other

Terms are used interchangeably, and there are also a number of other names of the same process...

Grouping and Segmentation: Names

- Image Parsing
- Perceptual Organisation
- Binding
- Perceptual Grouping
- Figure-Ground segmentation (restricted case):
  - segment one object (the figure) from everything else (the ground / background)

Grouping and Segmentation: Approaches

Top-down (coming from internal knowledge, prior experience)

- elements belong together because they lie on the same object

Bottom-up (coming from the image properties)

- elements belong together because they are similar or locally coherent

These approaches are NOT mutually exclusive

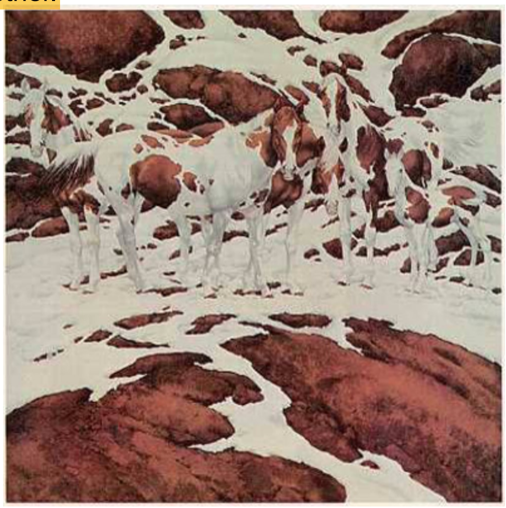
Both bottom-up and top-down influences are required for general purpose image segmentation

In humans we can observe (with the right choice of image) the influences of individual bottom-up and top-down cues to perceptual grouping.

Top-down Influences: Knowledge

Meaningfulness or Familiarity:

Elements that, when grouped, form a familiar object tend to be grouped together.

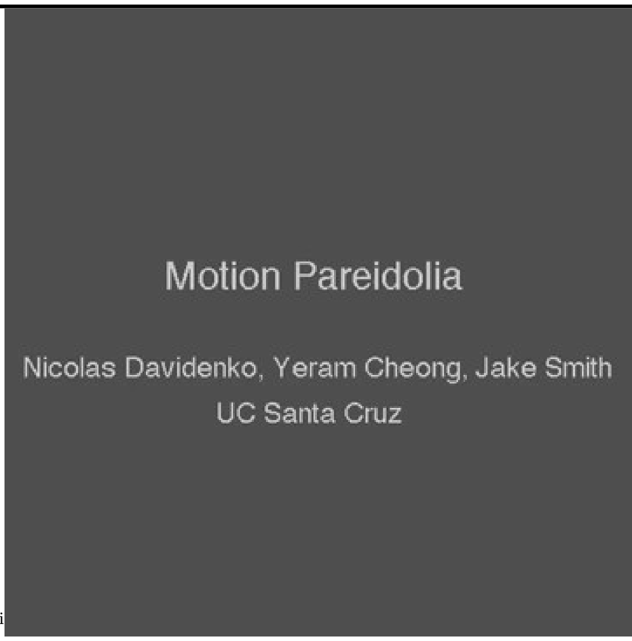


Top-down Influences: Knowledge

Meaningfulness or Familiarity:

Elements that, when grouped, form a familiar object tend to be grouped together.

Top-down Influences: Expectation



"grouping together"

Bottom-up Influences: Gestalt Laws

Many bottom-up factors influence how elements are grouped.

These influences are called the "Gestalt Laws" or the "Gestalt Grouping Principals"

Gestalt Laws are not really laws (that must be obeyed) but heuristics ("rules of thumb" that are often obeyed)

Original Gestalt Laws (proposed in early 20th century) include:

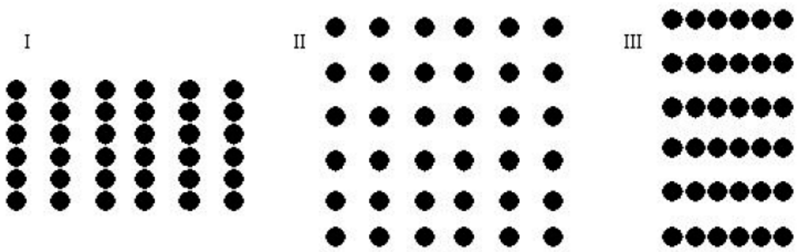
- Proximity
- Similarity
- Closure
- Continuity (or Good Continuation)
- Common Fate
- Symmetry

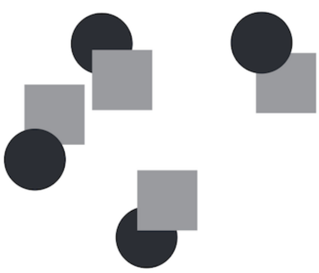
More Recent Laws (hence, not *true* Gestalt Laws) include:

- Common Region
- Connectivity

Gestalt Laws: Proximity

Elements that are near to each other tend to be grouped.

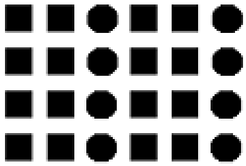




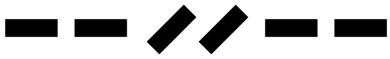
Gestalt Laws: Similarity

Elements that are similar tend to be grouped together.

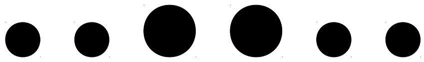
Similar shape



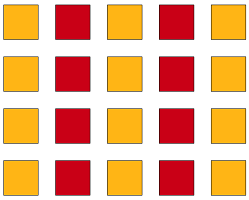
Similar orientation



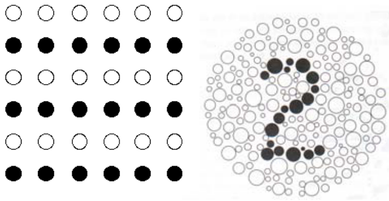
Similar size



Similar colour

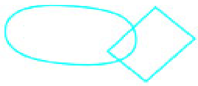


Similar luminance

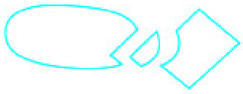


Gestalt Laws: Closure

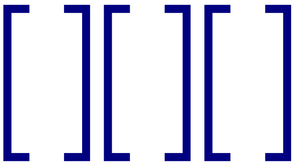
Elements that, when grouped, result in closed boundaries tend to be seen as belonging together.

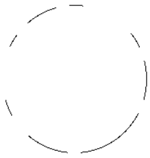


These are perceived as a square and triangle (left), and an ellipse and a rectangle (right), not as a combination of strange shapes, e.g.:



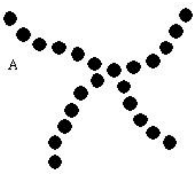
Potential closure is sufficient to result in grouping:

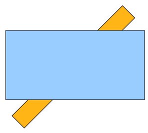


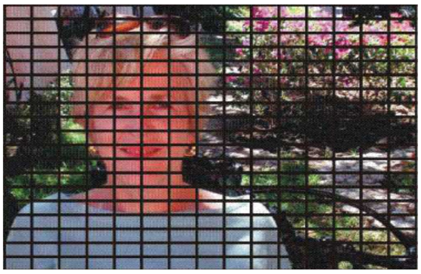


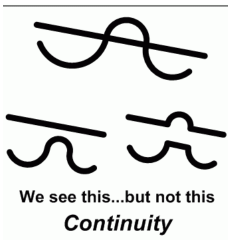
Gestalt Laws: Continuity

Elements that, when grouped, result in straight or smoothly curving lines, or smooth surfaces, tend to be seen as belonging together





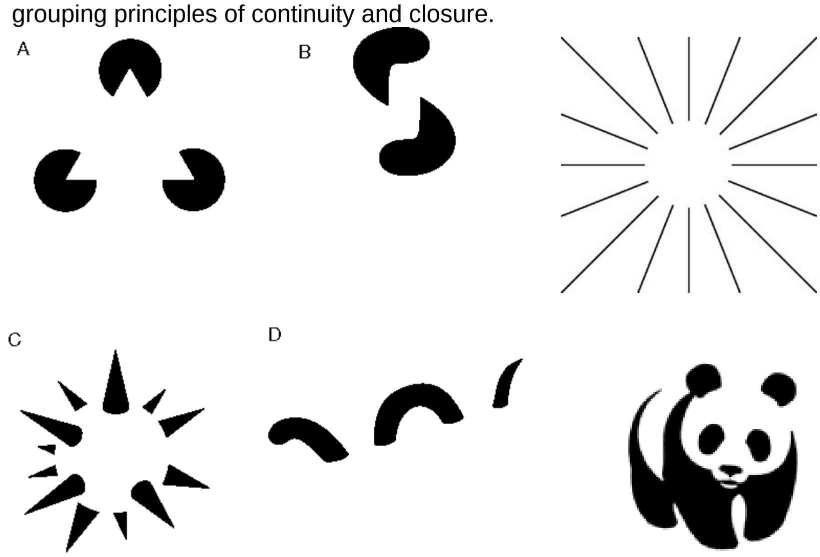




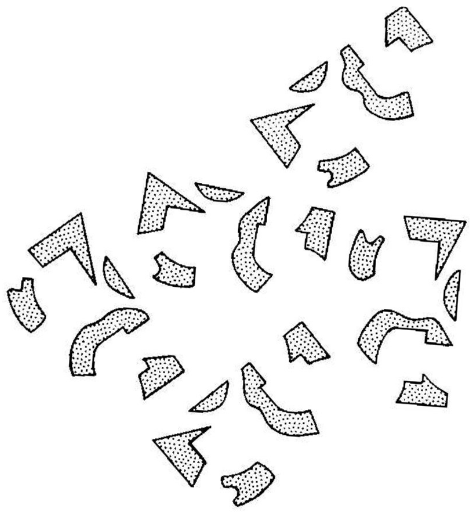
We even "invent" illusory contours (groups of white elements) to conform to our expectations about continuity.

Gestalt Laws: Continuity and Closure

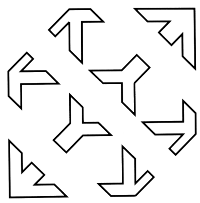
More examples of illusory contours and surfaces resulting from the Gestalt grouping principles of continuity and closure.



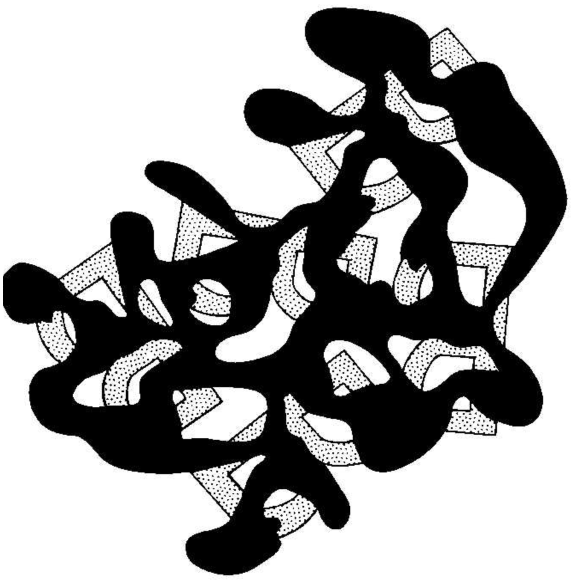
Gestalt Laws: Continuity and Closure



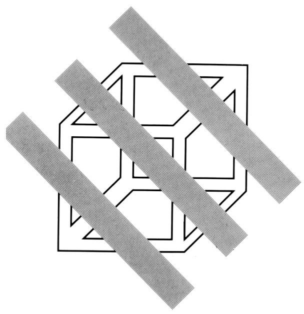
Seen as a collection of unrelated blobs.



Gestalt Laws: Continuity and Closure

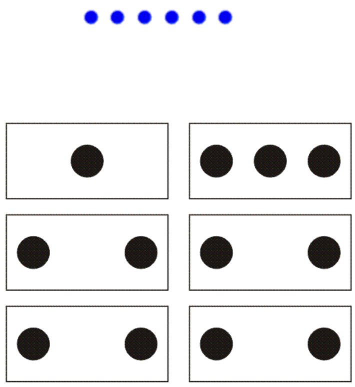


The presence of an occluding surface (the elements of which can be grouped together and segmented from the rest of the image) enables us to group the blobs into meaningful shapes.

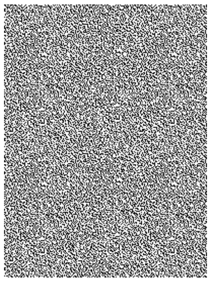


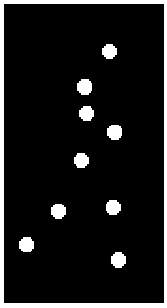
Gestalt Laws: Common Fate

Elements that have coherent motion tend to be grouped together.



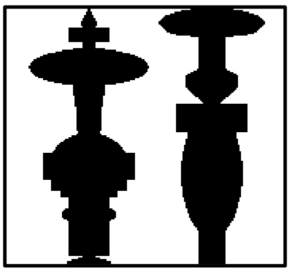
Kolers & Pomerantz, 1971

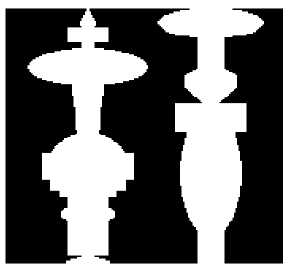




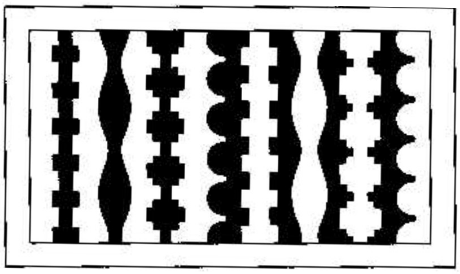
Gestalt Laws: Symmetry

Elements that form symmetric groups tend to be grouped together.





Black areas seen as figure on this side



White areas seen as figure on this side

Gestalt Laws: Symmetry

Elements that form symmetric groups tend to be grouped together.

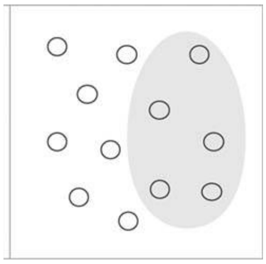
[ ]{ }[ ]

괄호로 그룹을 정한다

c.f.:

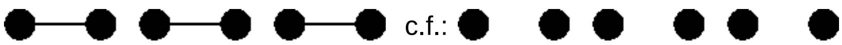
Gestalt Laws: Common Region

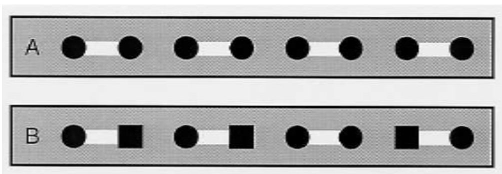
Elements that lie inside the same closed region tend to be grouped together.

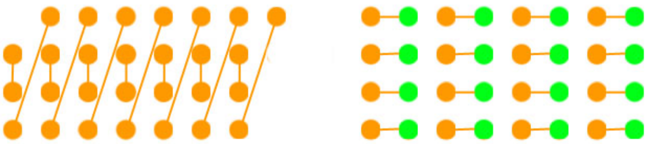


Gestalt Laws: Connectivity

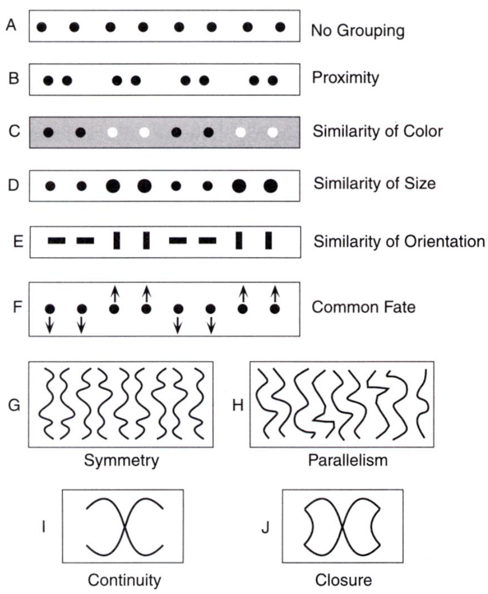
Elements that are connected by another element tend to be grouped together.

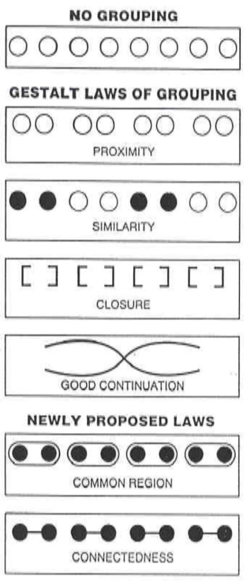






Gestalt Laws: Summary





Gestalt Laws: Summary

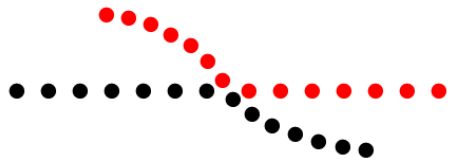
Many other bottom-up cues have been shown to effect segmentation!

Not all are clearly distinct, e.g.:

- Law of proximity = Law of Similarity for location
- Law of Common Fate = Law of Similarity for motion

No clear rules about what happens when multiple Gestalt principles are in conflict, e.g.:

- Law of proximity vs Law of common region (see earlier)
- Law of proximity vs Law of connectivity (see earlier)
- Law of proximity vs Law of symmetry (see earlier)
- Law of similarity vs Law of continuity:

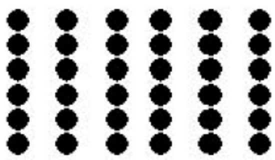


Gestalt Laws: Simplicity

The over-riding Gestalt principle (that encompasses all the others) is:

Simplicity (or Pragnanz): Every stimulus pattern is seen in such a way that the resulting structure is as simple as possible

i.e. We organize pattern into groups that make it easier to communicate, reason, and remember things about the image.



e.g. It is easier to remember this image as six columns of dots than as 36 individual dots

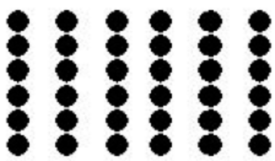
e.g. It is easier to remember this image a triangle and a square than as 7 individual lines

Alternative to Simplicity

Likelihood Principle (Helmholtz, 1910):

The preferred perceptual organization of a sensory pattern reflects the most likely object or event.

i.e What we see is inferred from both the sensory input data and our prior experience.



e.g. past experience has told us that neighbouring image elements have a high probability of being part of the same object. Hence, we infer that the nearby elements belong together.

e.g. past experience has made us familiar with regular shapes, hence we see this image as a square and a triangle.

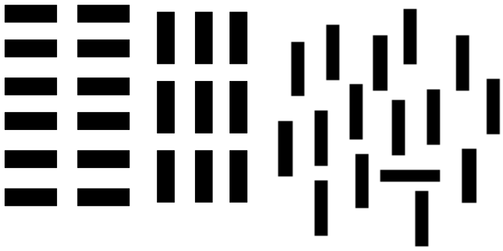
Implementing grouping principles

Both the Gestalt Law of Simplicity and Helmholtz's Likelihood Principle are abstract explanations: they say nothing about how these computations are implemented in the brain.

V1 Lateral Connections

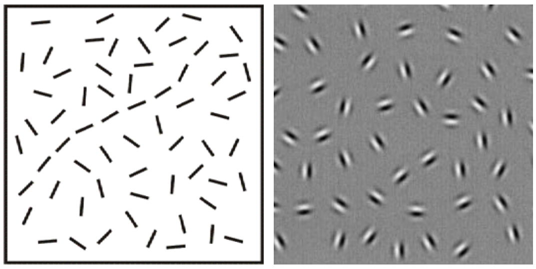
Lateral inhibitory connections, give rise to texture segmentation and pop-out

= Gestalt Law of similarity



Lateral excitatory connections, give rise to contour integration

= Gestalt Law of continuity



Object Segmentation: border ownership

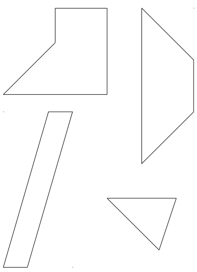
In most tasks we want to group elements into objects.

An object consists of one or more surfaces plus a boundary.

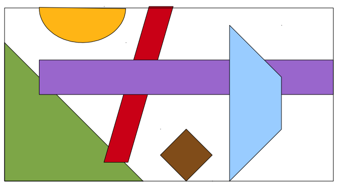
The boundary is “owned” by the object.

The background appears borderless (and hence shapeless).

Demonstration:



where do these shapes appear in this image?



Object Segmentation: border ownership

Where is the arrow?

Difficult to spot as letters are seen as foreground.

Birds or fish?

Can't see both at the same time, as border can only be owned by one or the other interpretation.

Object Segmentation: border ownership

Rubin faces-vase stimulus

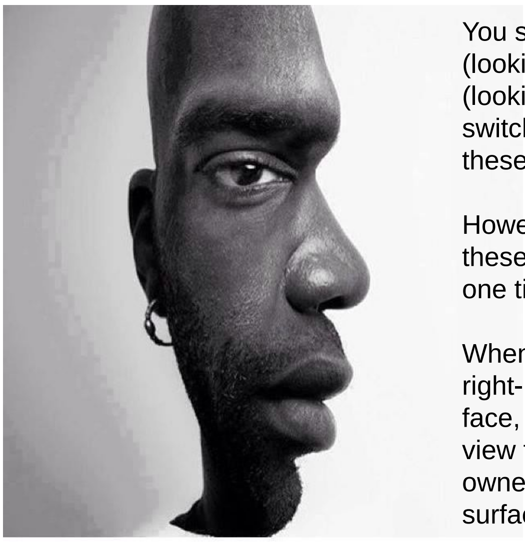
There are no strong Gestalt cues as to how to segment this image, and both sides are supported by top-down knowledge.

You thus see either two faces in profile OR a vase, and perception switches spontaneously between these different interpretations.

However, you see only one of these two interpretations at any one time.

When you see a face, the common border is owned by the face, and hence the vase becomes borderless and shapeless.

Object Segmentation: border ownership



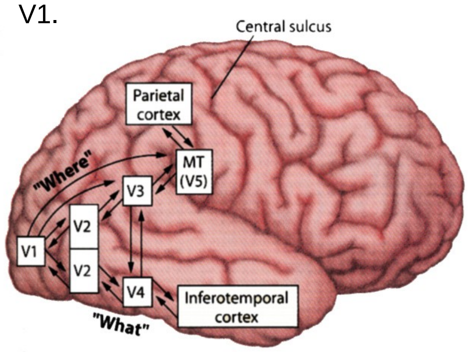
You see the face as a profile (looking right) OR a frontal view (looking at you), and you can switch your perception between these different interpretations.

However, you see only one of these two interpretations at any one time.

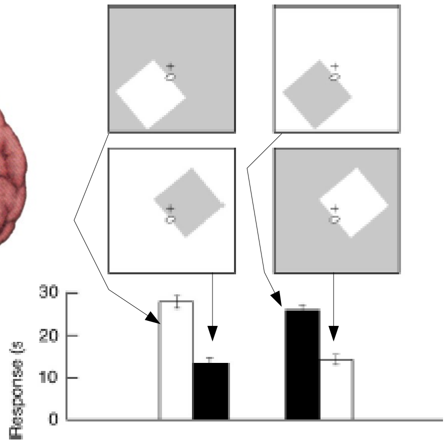
When you see the profile view the right-hand border is owned by the face, when you see the frontal view the border seems to be owned by a white occluding surface.

V2 Border-ownership cells

V2 is the cortical area which comes next in the hierarchy to V1.



V2 contains cells that encode border-ownership.



e.g. This cell is selective for a particular edge orientation, but only responds strongly when the perceived figure is to the lower left of its RF.

V2 Border-ownership cells

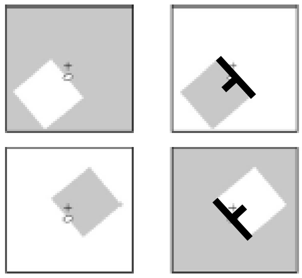
One mechanism by which V2 cells could compute border-ownership, is via lateral connections within V2.

Imagine that at each location there are multiple V2 neurons selective to different orientations.

At each location and orientation there are a pair of neurons that prefer the foreground object to be on opposite sides of the border.

These neurons compete with each other.

Denote orientation preference as a line and the side of border ownership using a shorter perpendicular line.

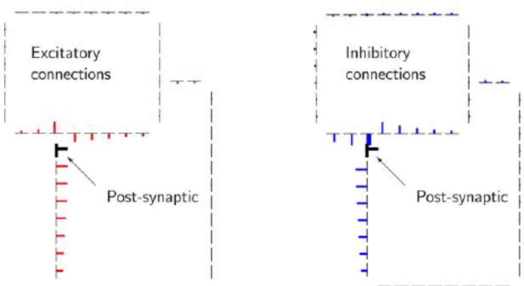


V2 Border-ownership: model

One mechanism by which V2 cells could compute border-ownership, is via lateral connections within V2.

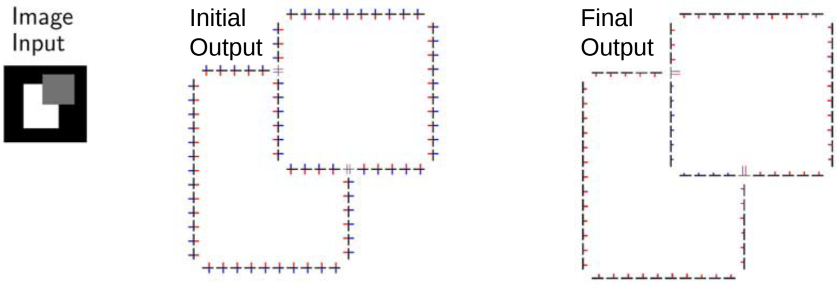
Excitatory connections link neurons encoding segments consistent with a probable object.

Inhibitory connections link neurons encoding segments inconsistent with a probable object



V2 Border-ownership: model

This pattern of excitation and inhibition results in the model settling into a state in which border-ownership (and hence the grouping of edge segments to form object boundaries) is consistent with human perception.



Summary

Mid-Level Vision = grouping image elements together

Which elements are grouped together is influenced by:

- prior experience and object knowledge (top-down influences), and
- Gestalt Laws applied to the image properties (bottom-up influences)

Interpreted as:

- Increasing Simplicity: the grouping generated simplifies the image
- Increasing Likelihood: the grouping generated is the most likely

Lateral connections in V1 and V2:

- enhance neural responses that are probable/consistent
- implement some Gestalt principles and produce border-ownership.