Object vision (Chapter 4)

Lecture 8

Jonathan Pillow
Sensation & Perception
(PSY 345 / NEU 325)
Princeton University, Spring 2019

What do you see?

What do you see?

What do you see?

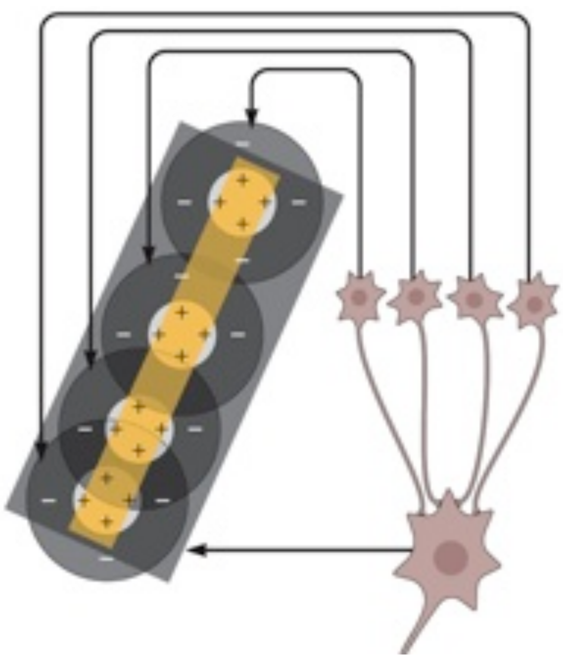
How did you recognize that all 3 images were of houses?

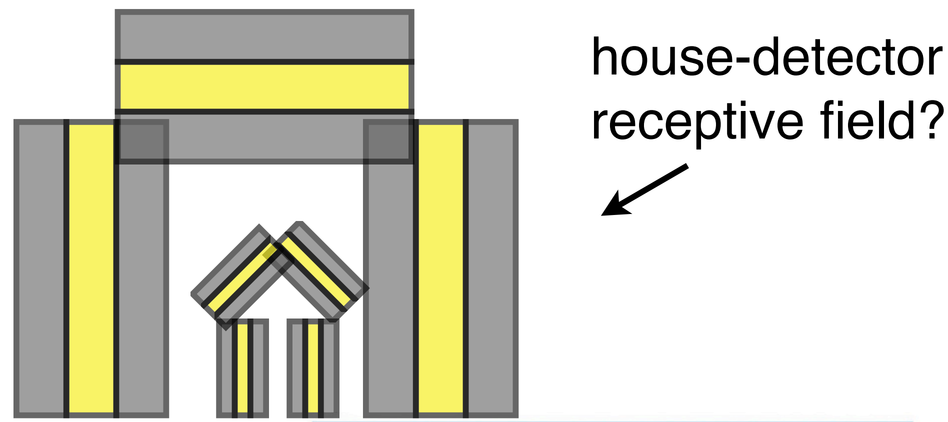
How did you know that the 1st and 3rd images showed the same house?

This is the problem of *object recognition*, which is solved in visual areas beyond V1.

Unfortunately, we still have no idea how to solve this problem.

Not easy to see how to make Receptive Fields for houses the way we combined LGN receptive fields to make V1 receptive fields!



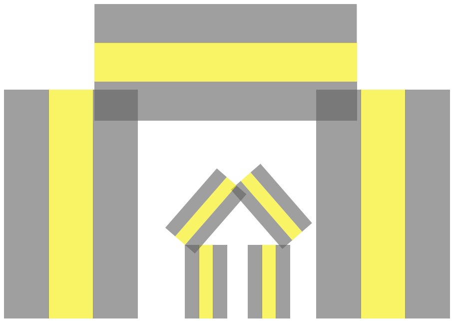


Viewpoint Dependence

View-dependent model - a model that will only recognize particular views of an object

- **template-based model**

e.g.



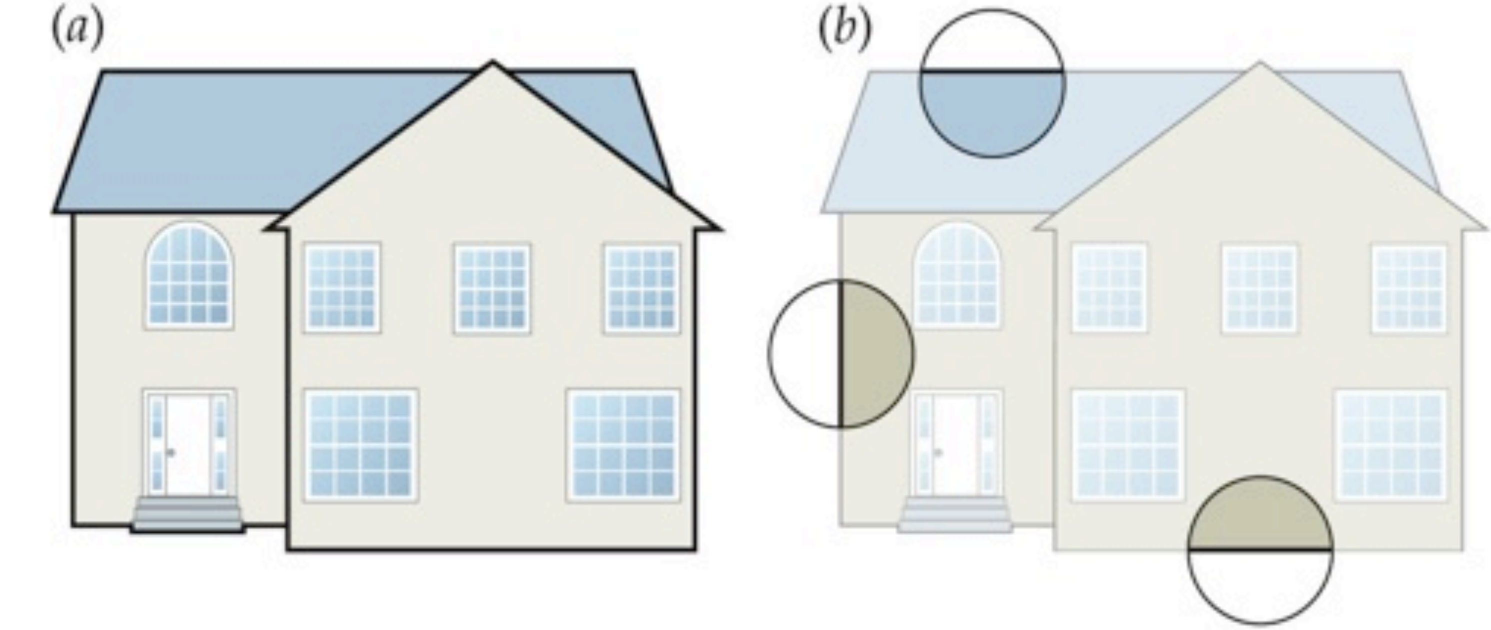
"house" template

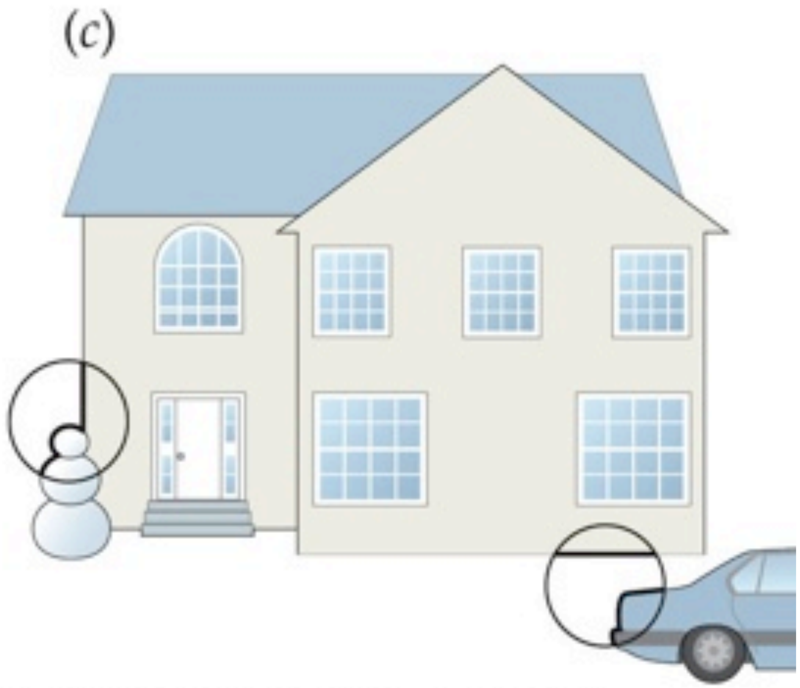
Problem: need a neuron (or "template") for every possible view of the object
- quickly run out of neurons!

Middle vision:

– *after* basic features have been extracted and *before* object recognition and scene understanding

- Involves perception of edges and surfaces
- Determines which regions of an image should be grouped together into objects



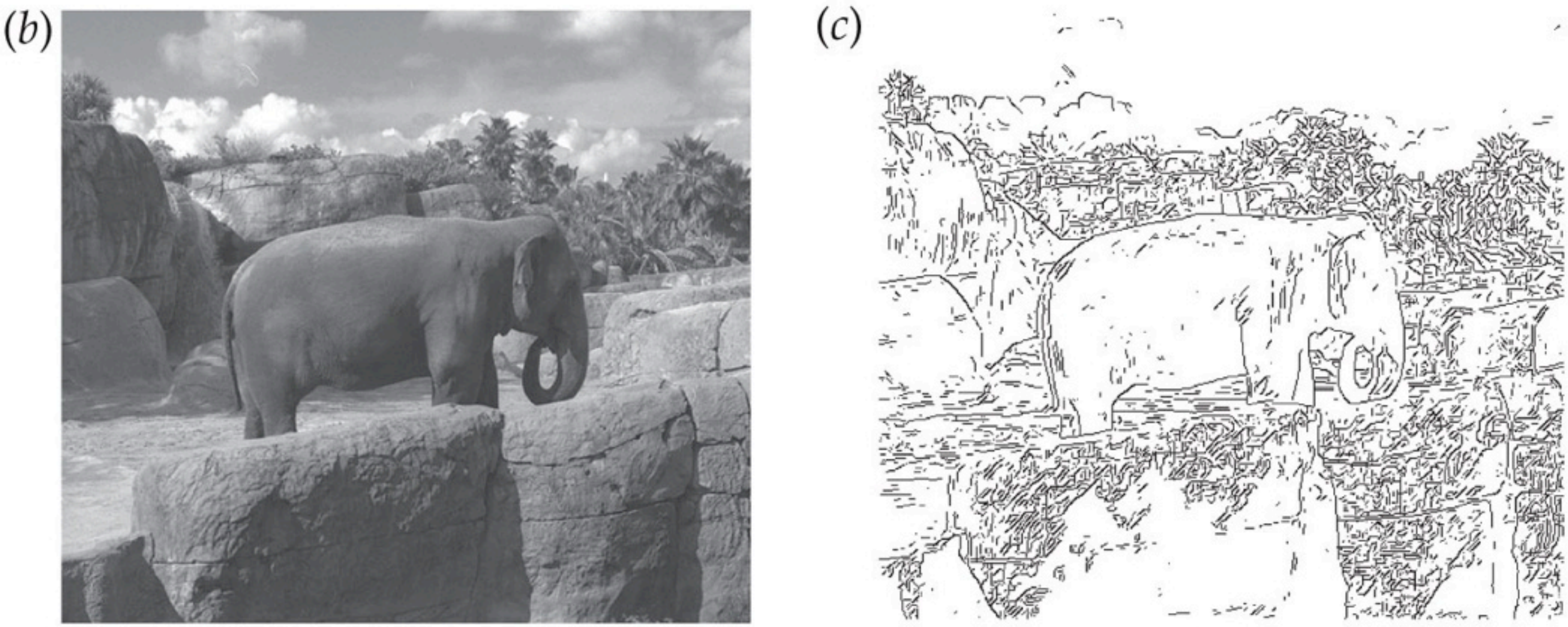


Finding edges

- How do you find the edges of objects?
- Cells in primary visual cortex have small receptive fields
- How do you know which edges go together and which ones don't?

Computer-based edge detectors are not as good as humans

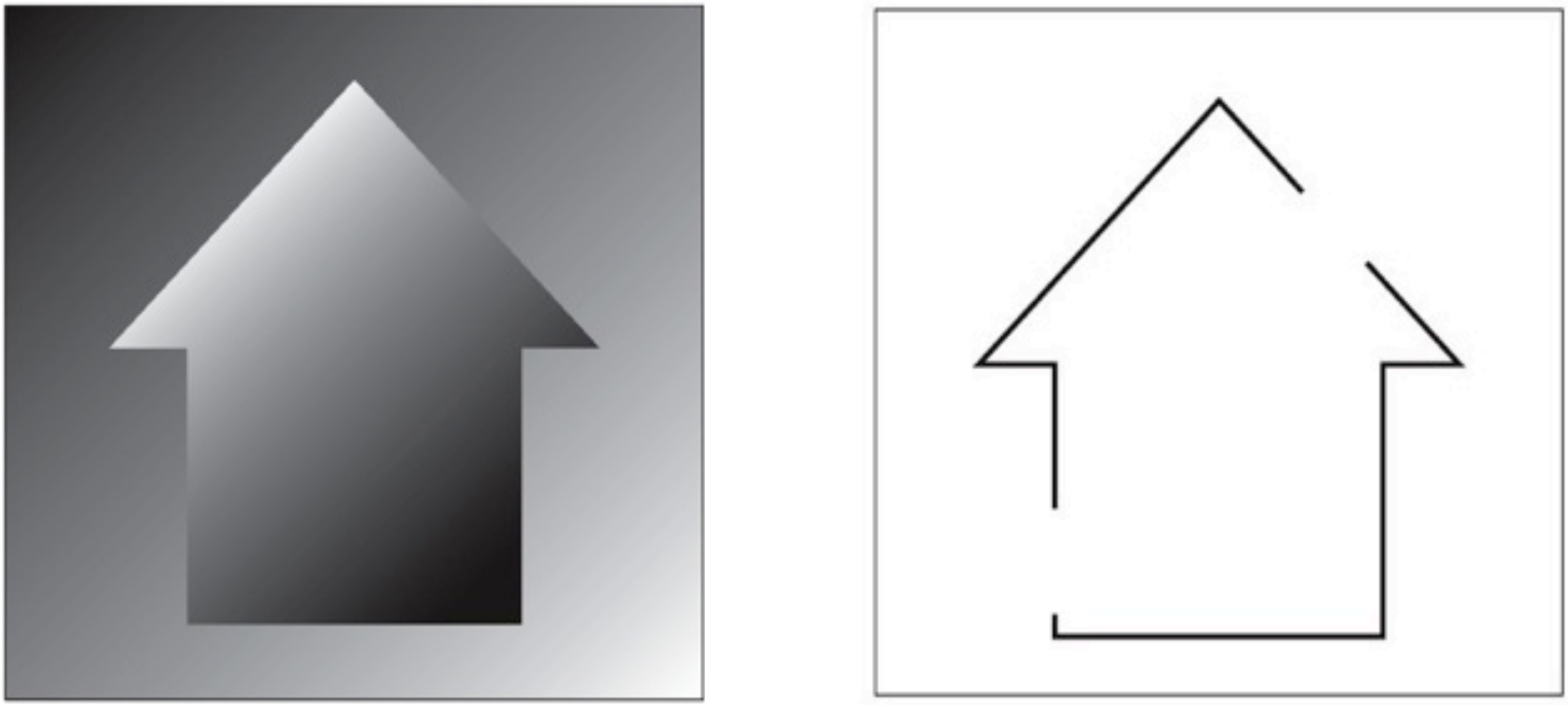
- Sometimes computers find too many edges



- "Edge detection" is another failed theory (along with Fourier analysis!) of what V1 does.

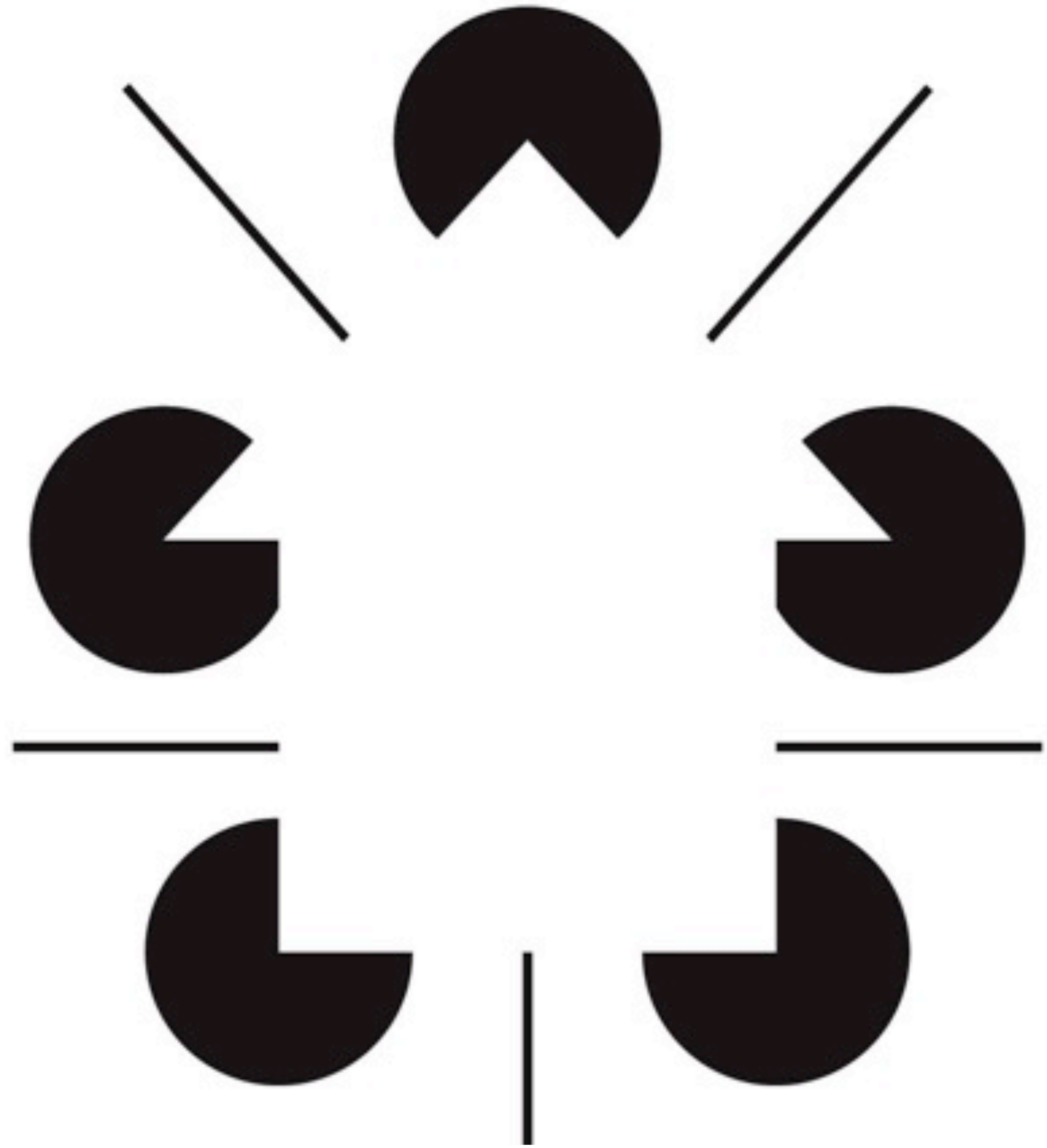
Computer-based edge detectors are not as good as humans

- Sometimes computers find too few edges



"Kanizsa Figure"

illusory contour: a contour that is perceived even though no luminance edge is present



- **Gestalt**: In German, “form” or “whole”
- **Gestalt psychology**: “The whole is greater than the sum of its parts.”
- Opposed to other schools of thought (e.g., structuralism) that emphasize the basic elements of perception

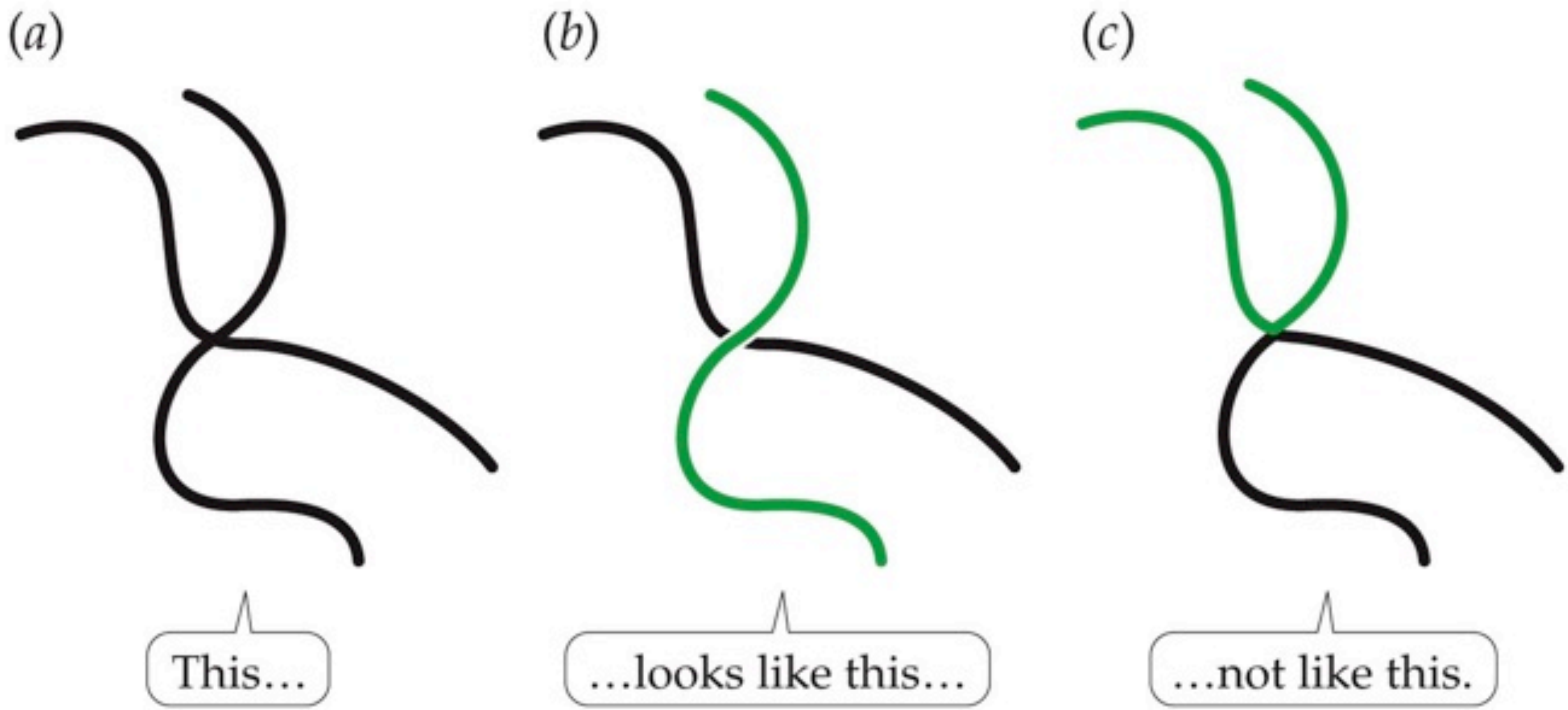
structuralists:

- perception is built up from “atoms” of sensation (color, orientation)
- challenged by cases where perception seems to go beyond the information available (eg, illusory contours)

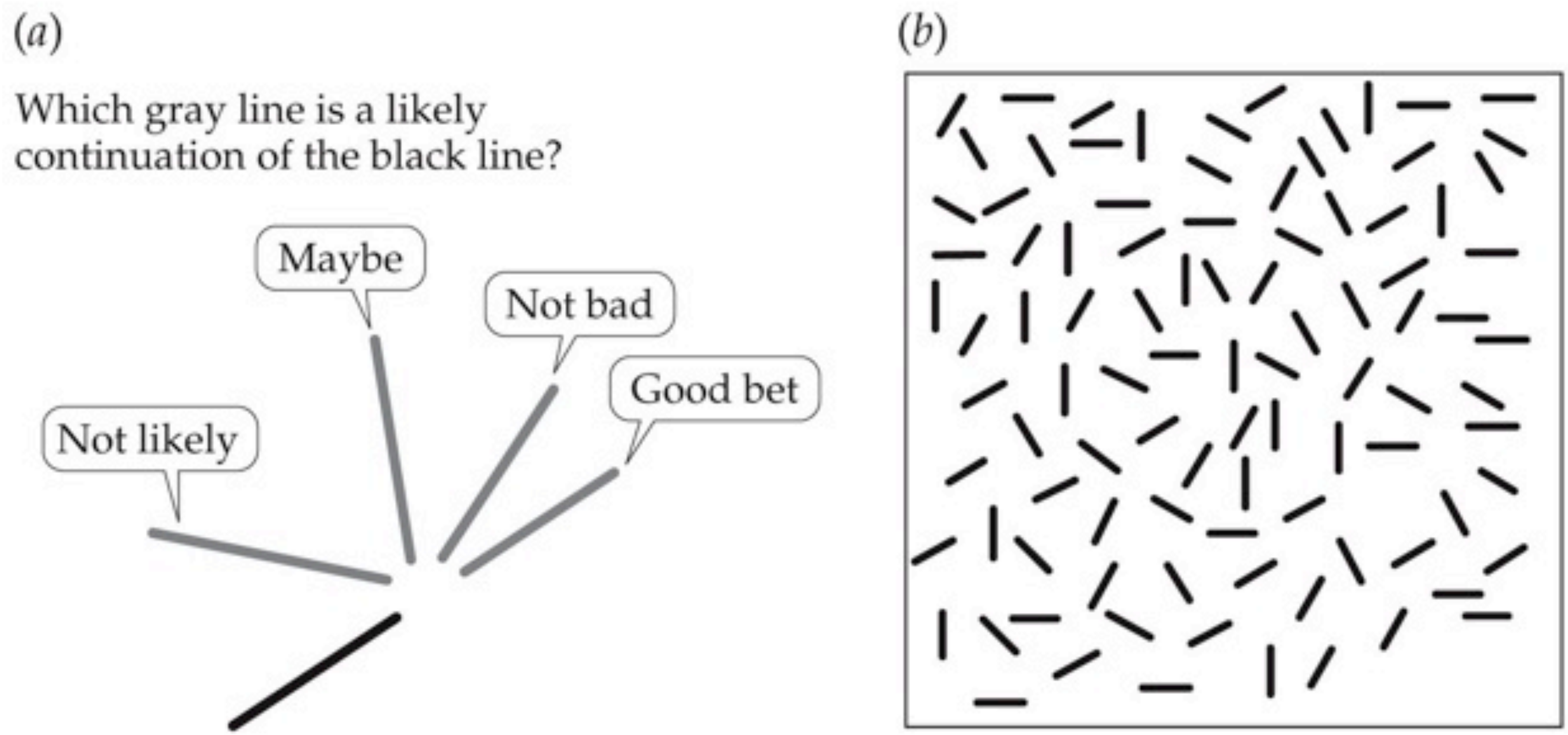
Gestalt grouping rules:

a set of rules that describe when elements in an image will appear to group together

Good continuation: A Gestalt grouping rule stating that two elements will tend to group together if they lie on the same contour

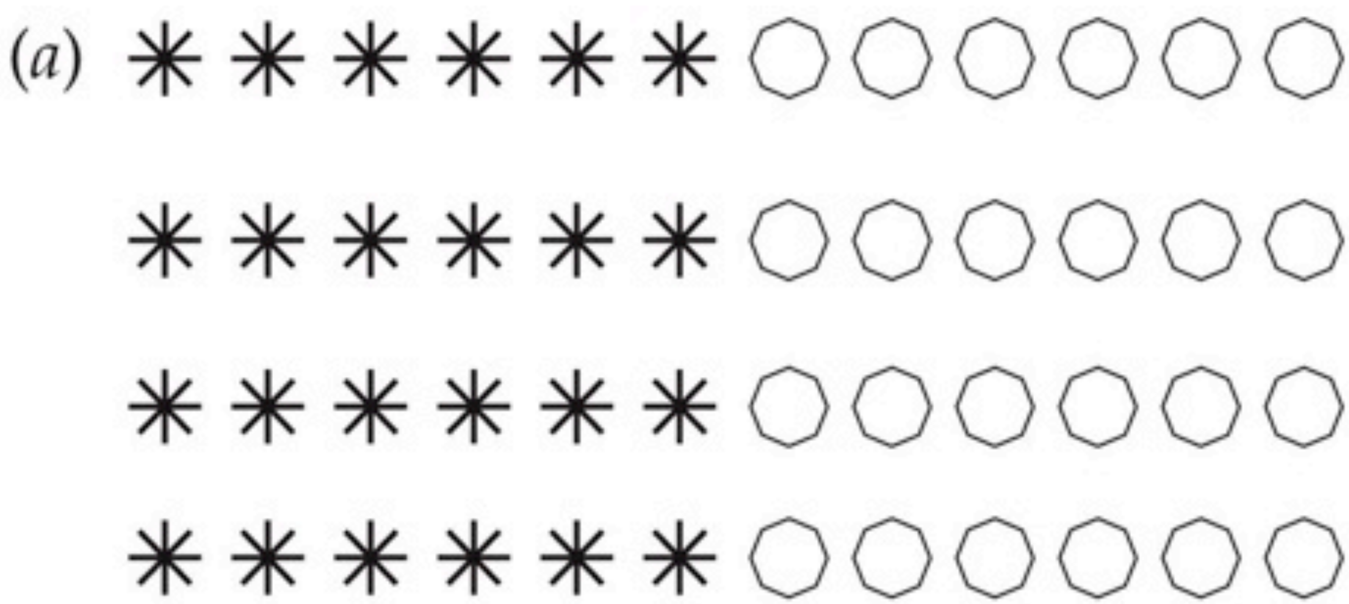


Good continuation: A Gestalt grouping rule stating that two elements will tend to group together if they lie on the same contour



Gestalt grouping principles:

- **Similarity**
- **Proximity**



Gestalt Principles

Dynamic grouping principles

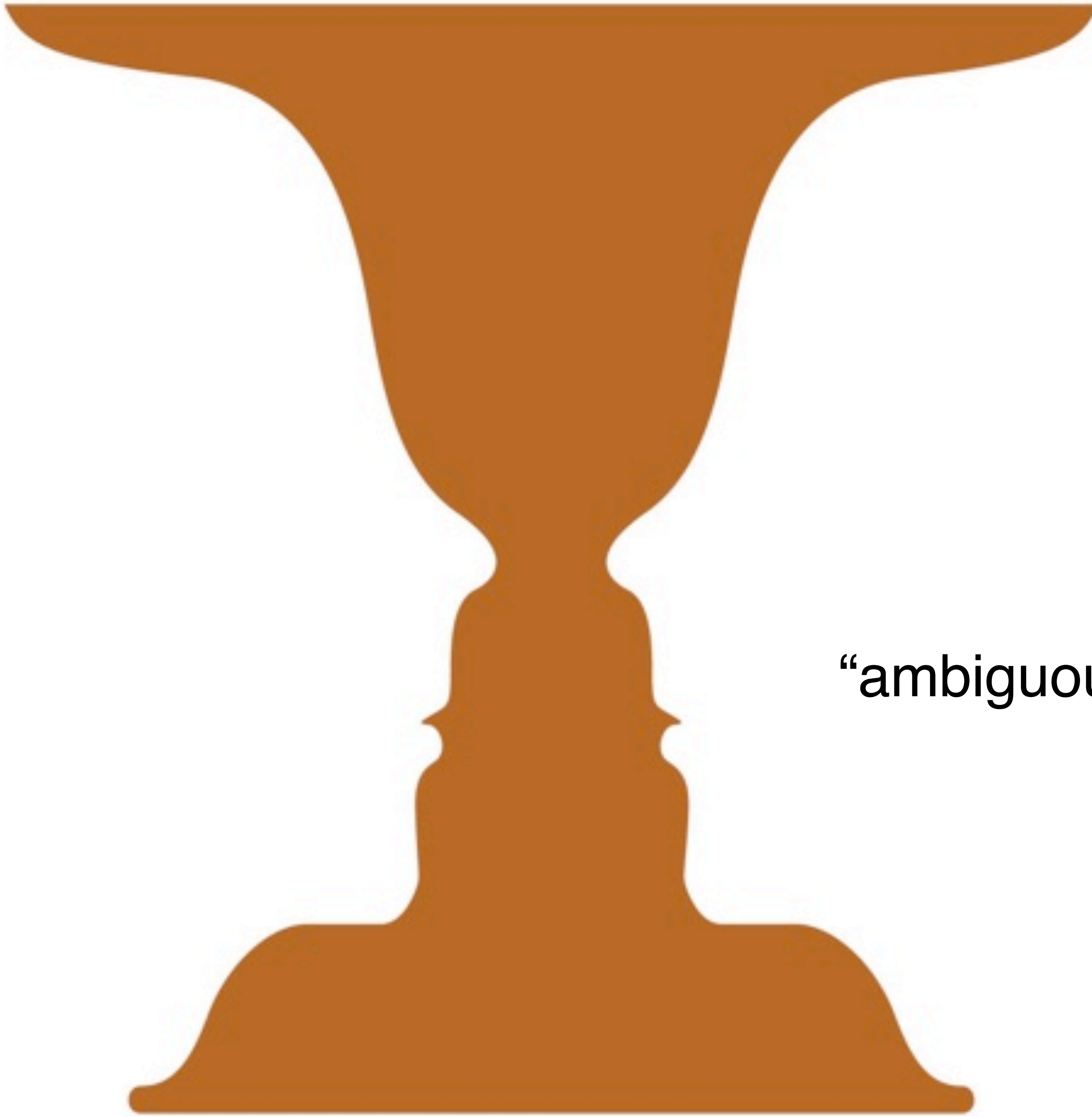
- **Common fate**: Elements that move in the same direction tend to group together

- **Synchrony**: Elements that change at the same time tend to group together

(See online demonstration: book website)

https://oup-arc.com/access/content/sensation-and-perception-5e-student-resources/sensation-and-perception-5e-activity-4-2?previousFilter=tag_chapter-04

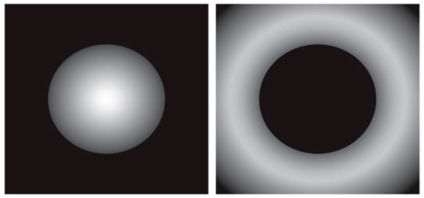
Figure/Ground Segregation: Face/Vase Illusion



“ambiguous figure”

Gestalt figure–ground assignment principles:

- **Surroundedness**: The surrounding region is likely to be ground
- **Size**: The smaller region is likely to be figure
- **Symmetry**: A symmetrical region tends to be seen as figure
- **Parallelism**: Regions with parallel contours tend to be seen as figure
- **Extremal edges:** If edges of an object are shaded such that they seem to recede in the distance, they tend to be seen as figure

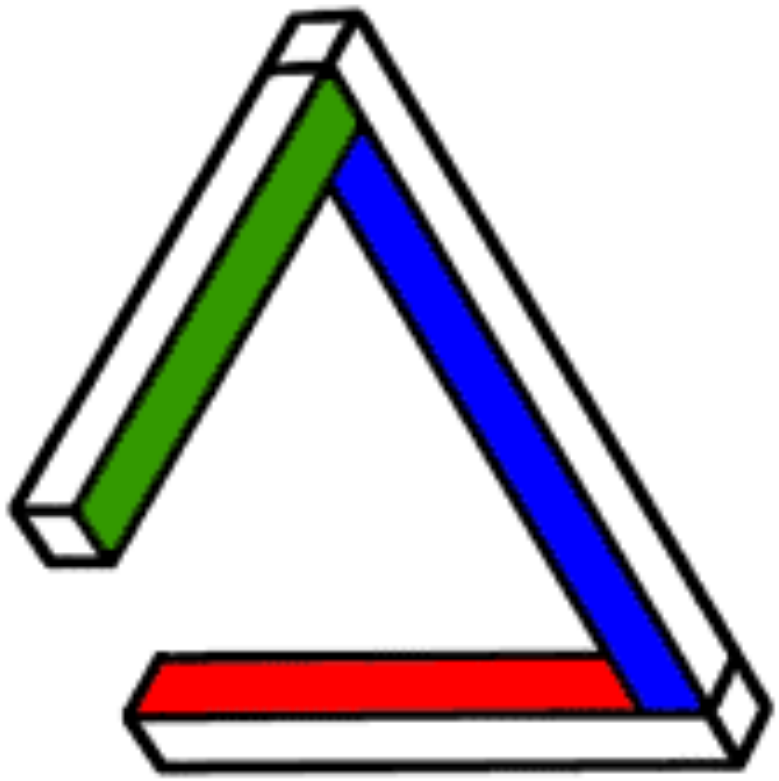


- **Accidental viewpoint:** produces a regularity in the visual image that is not present in the world
- **Visual system will *not* adopt interpretations that assume an accidental viewpoint!**

- **non-accidental viewpoint**

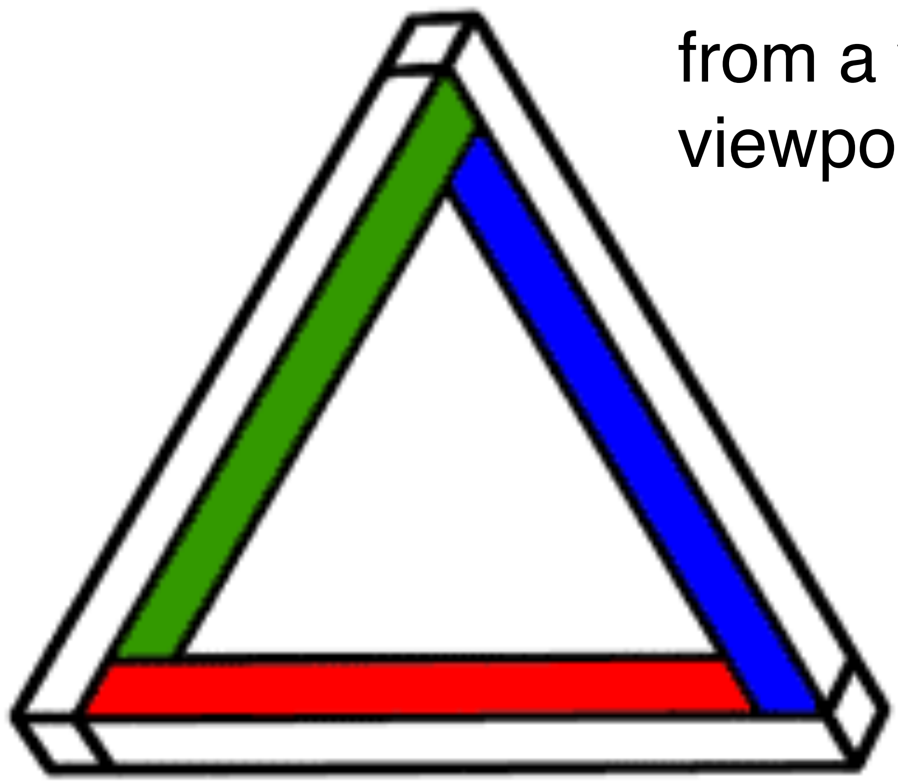
  "typical" viewpoint: interpretation won't change if you move the camera a little bit

- **Belivable 3-d figure:**



- **Unbelievable figure**

You could build a 3D object that would lead to this 2D image, but would need to take the picture from a very specific viewpoint



Impossible triangle
(Perth, Australia)

Impossible triangle
(Perth, Australia)

Accidental Viewpoints in street art


früh
Gmb
als da


YOUNGEAGERMINDS.COM


WORLD
YOUNGEAGERMINDS.COM

...one more argument against Naive Realism:

West Vancouver

Speed Bumps of the Future: Children

Speed Bumps of the Future: Children

"the girl's elongated form appears to rise from the ground as cars approach, reaching 3D realism at around 100 feet, and then returning to 2D distortion once cars pass that ideal viewing distance. Its designers created the image to give drivers who travel at the street's recommended 18 miles per hour (30 km per hour) enough time to stop before hitting Pavement Patty–acknowledging the spectacle before they continue to safely roll over her."

- Joseph Calamia (Discover magazine blog)

"It's a static image. If a driver can't respond to this appropriately, that person shouldn't be driving…."

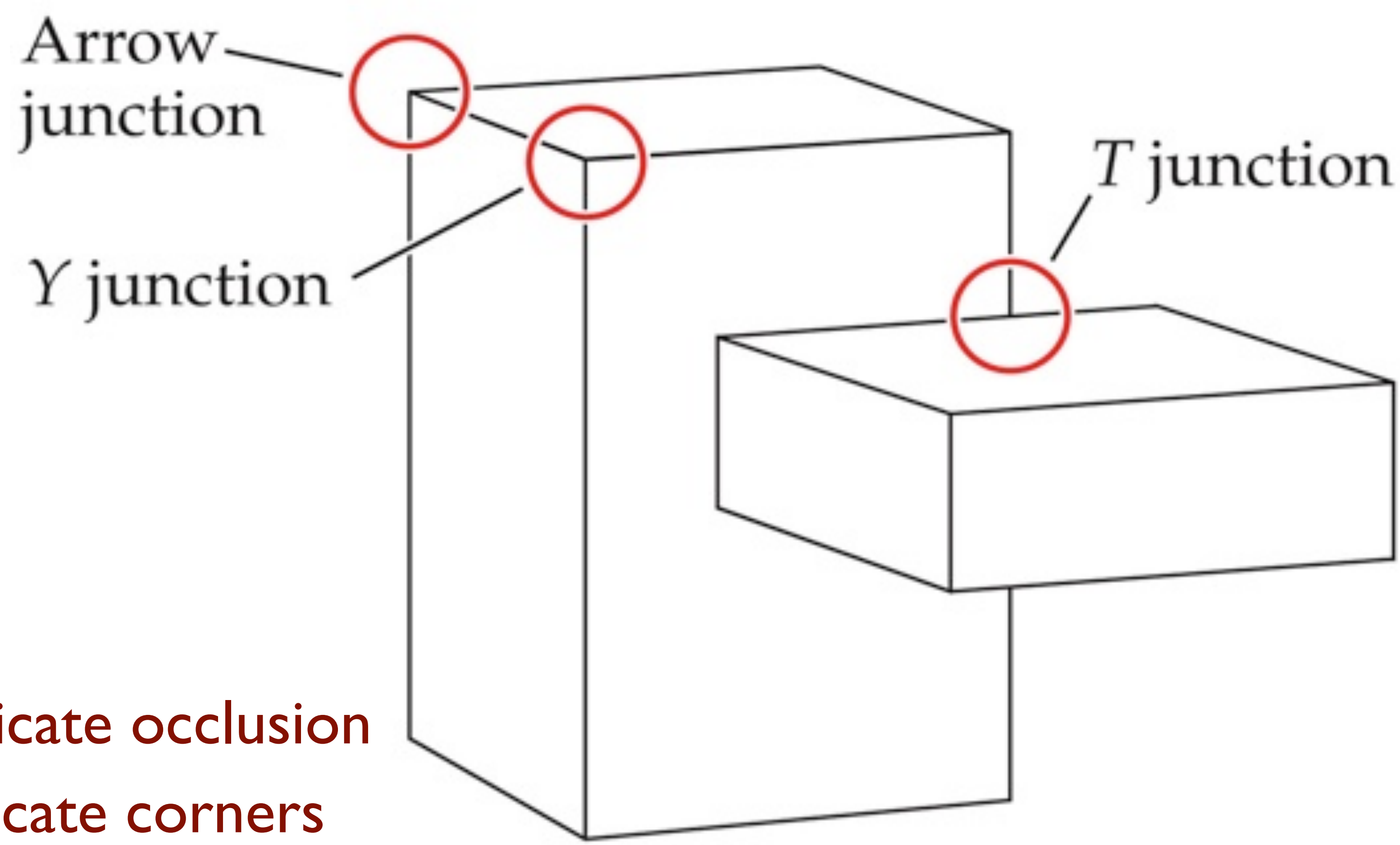
- David Duane, BCAA Traffic Safety Foundation

http://tinyurl.com/358r46p

Benard Pras - French sculptor

http://www.facteurcheval.com/creations/artistes/bernard-pras.html

Nonaccidental feature: features that do not depend on the exact (or accidental) viewing position of the observer



T junctions: indicate occlusion

Y junctions: indicate corners facing the observer

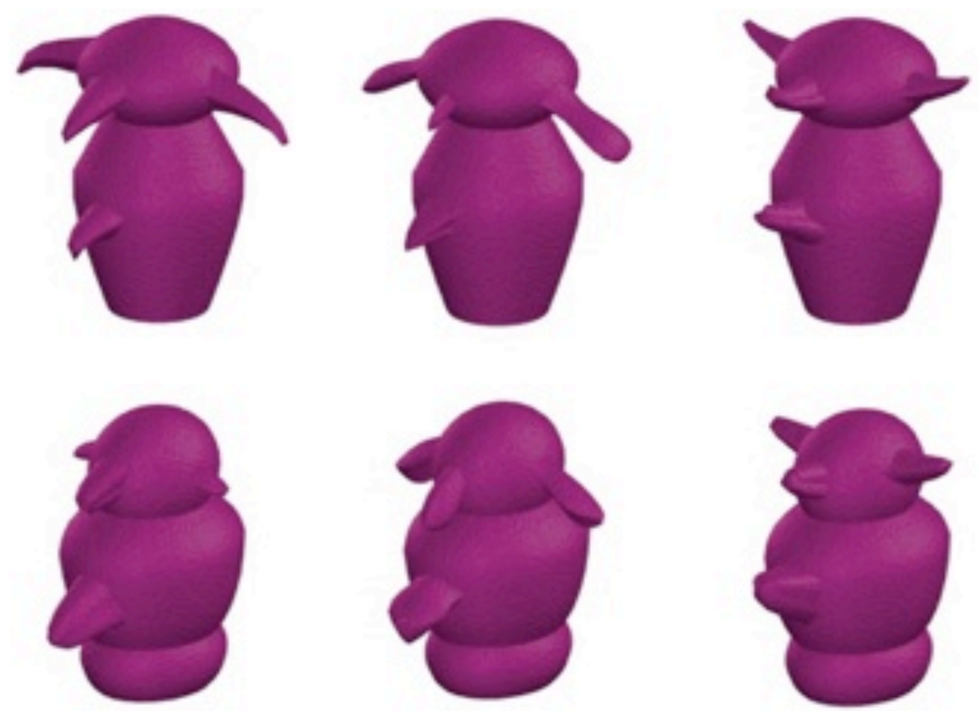
Arrow Junctions: corners facing away from observer

- these feature are still present if object is shifted, scaled or rotated by a small amount

Problems with view-invariant theories:

Object recognition = not completely viewpoint-invariant!

“greebles” (1998)

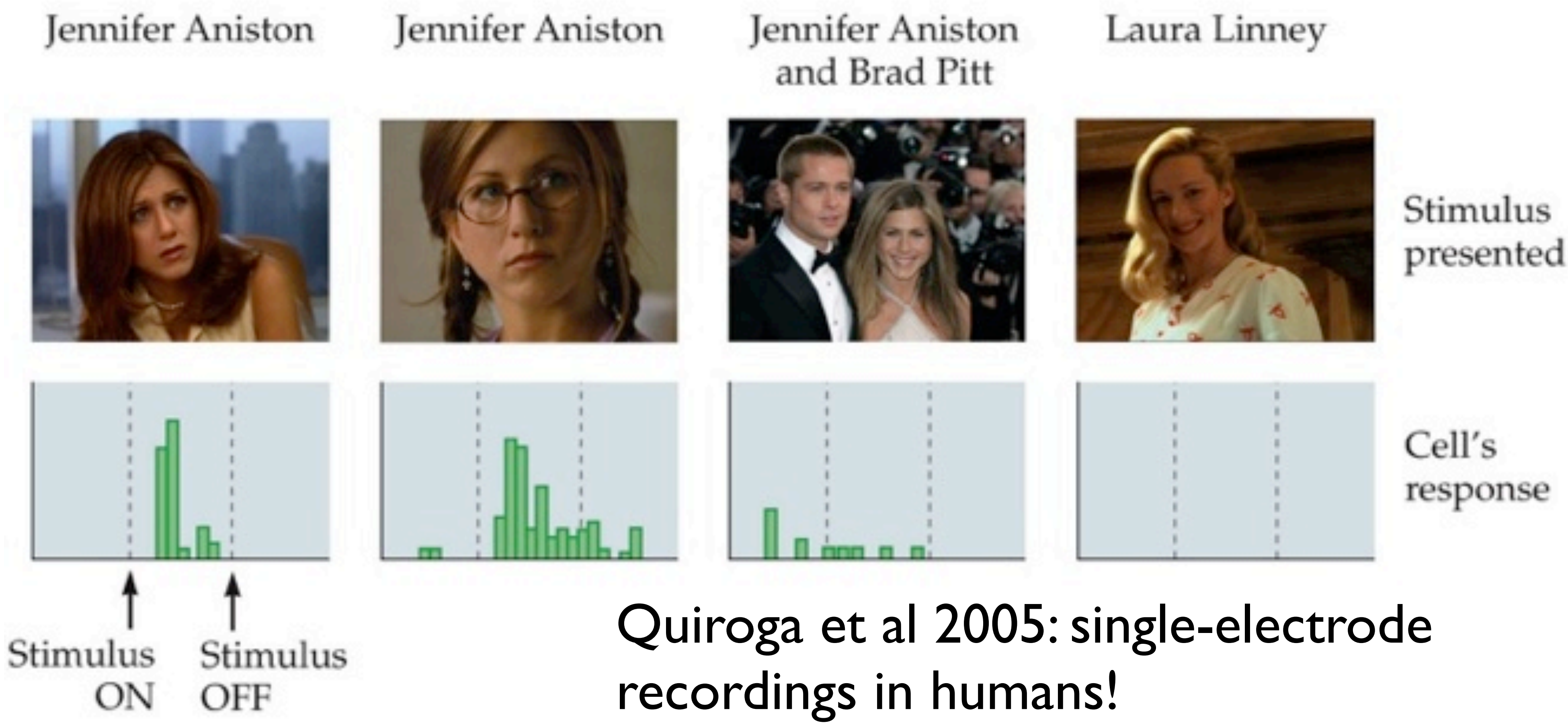


Viewpoint affects object recognition

- The farther an object is rotated away from a learned view, the longer it takes to recognize

Viewpoint-invariance in the nervous system

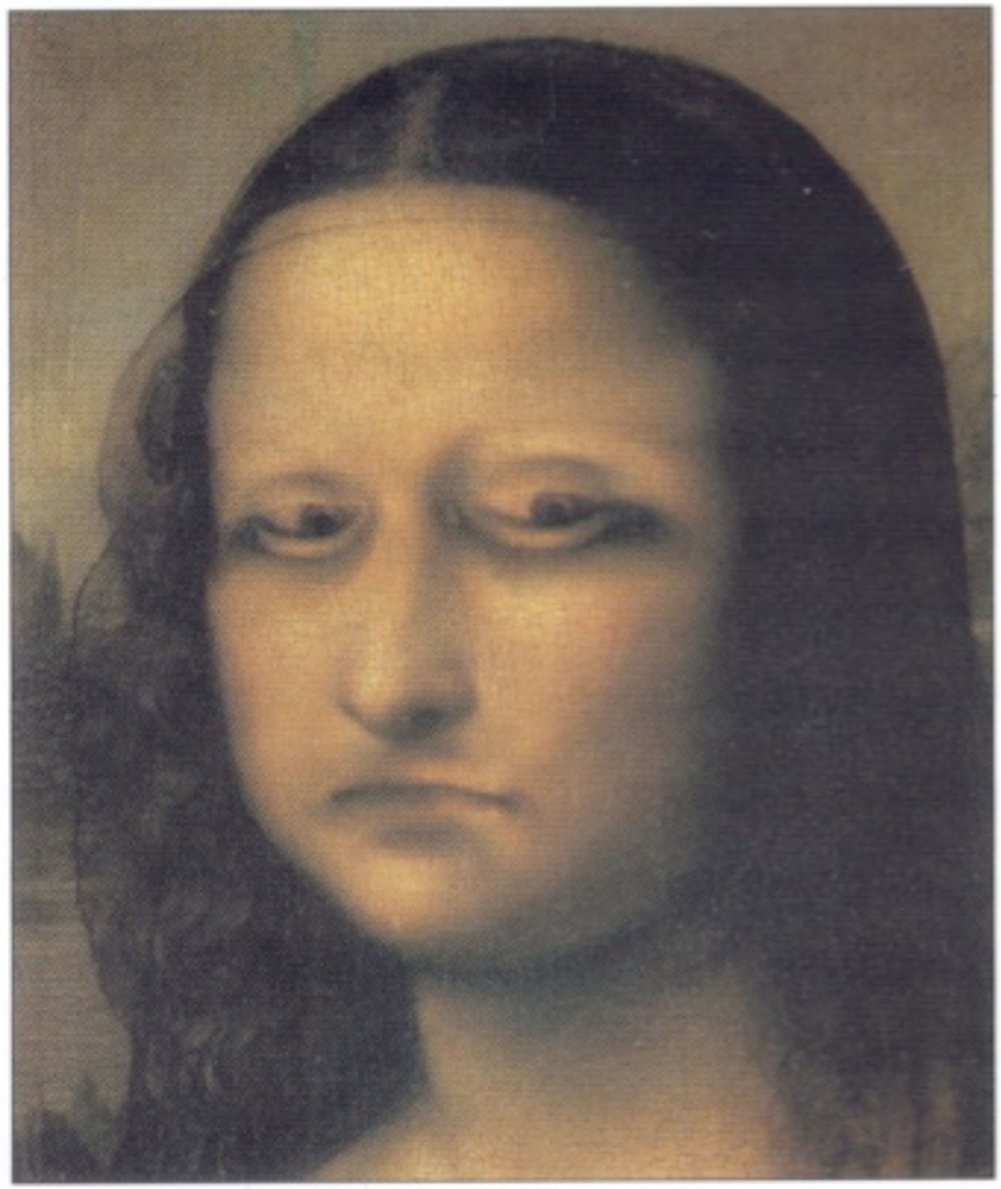
Inferotemporal (IT) cortex
- high selectivity to people / things, independent of viewpoint
- e.g., "Jennifer Anniston neuron"



Quiroga et al 2005: single-electrode recordings in humans!

Face Recognition

Face Recognition: not entirely viewpoint-invariant!



Conclusion:

- object recognition is *somewhat* but not entirely viewpoint invariant
- observers *do* seem to store certain preferred views of objects.

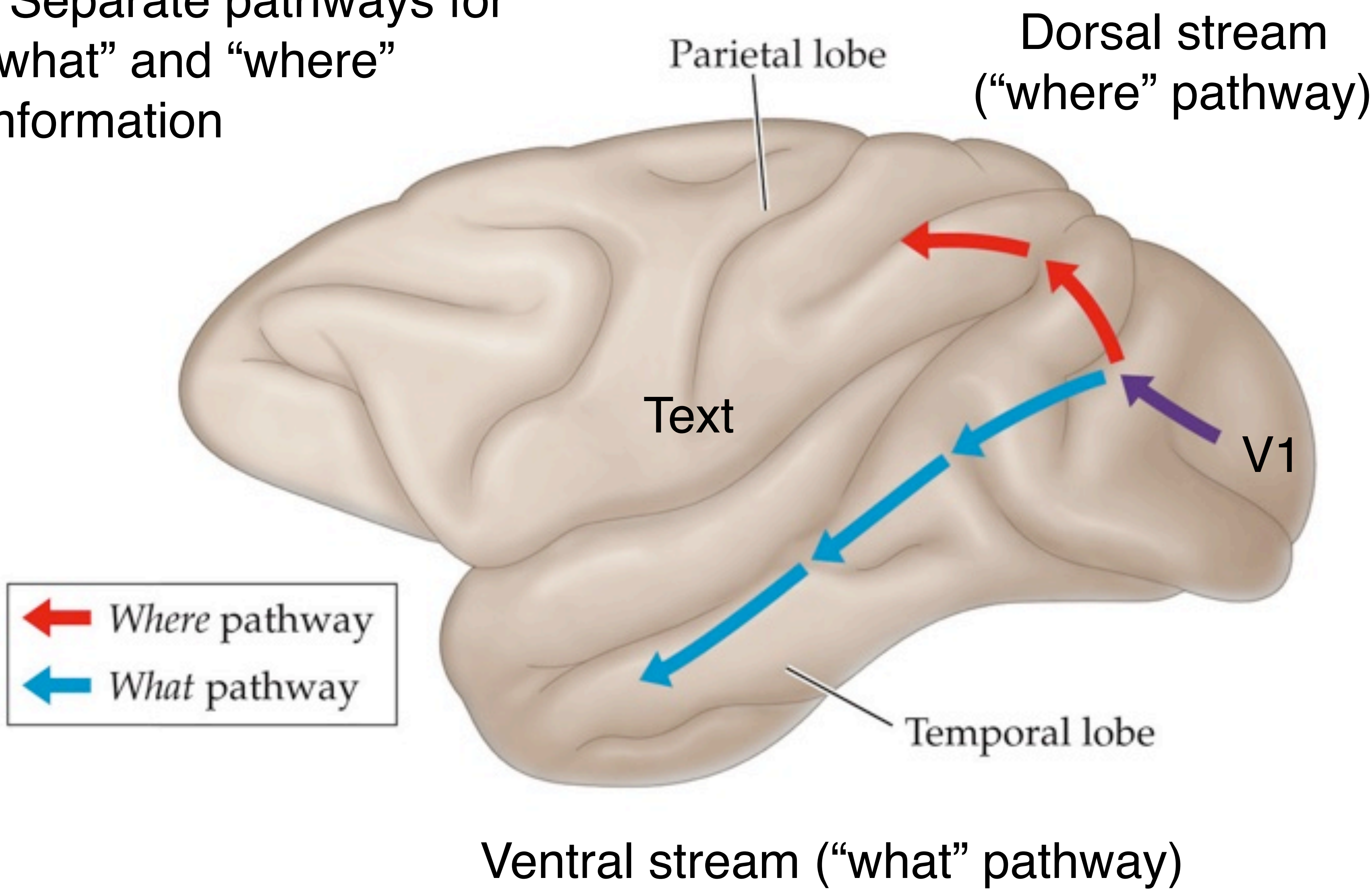
Makes sense from an evolutionary standpoint:

We generate representations that are as invariant as we need them to be for practical applications

Two facts that constrain any models of object recognition in the visual system

1. Visual processing divided into two cortical streams:

- Separate pathways for “what” and “where” information



2. Object recognition is *fast.* (100-200 ms)
Suggests operation of a feed-forward process.

(5 frames /s)

Feed-forward process: computation carried out one neural step after another, without need for feedback from a later stage

(Still debated, but it's agreed there's not much time for feedback).

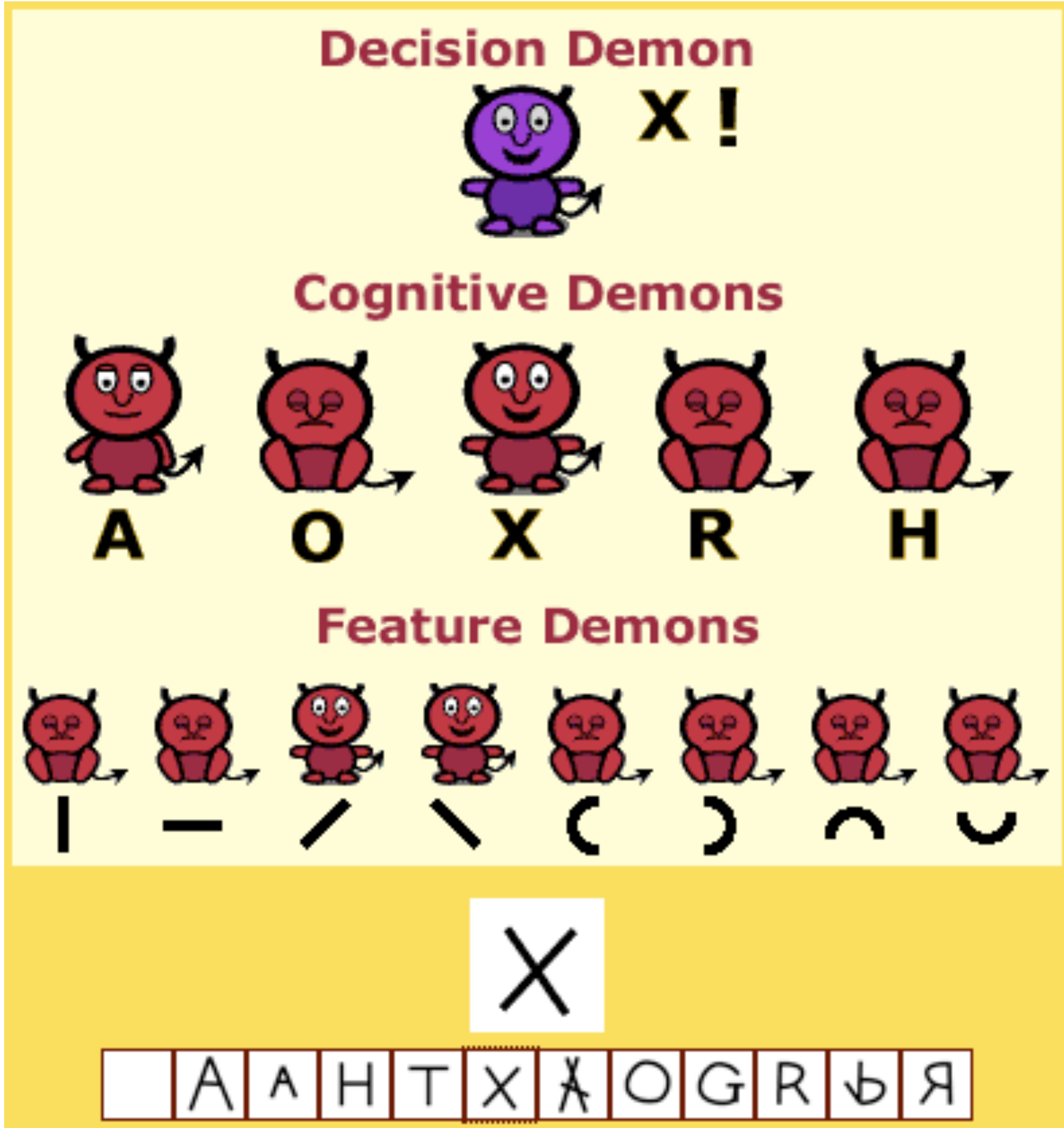
Models of Object Recognition

pandemonium model

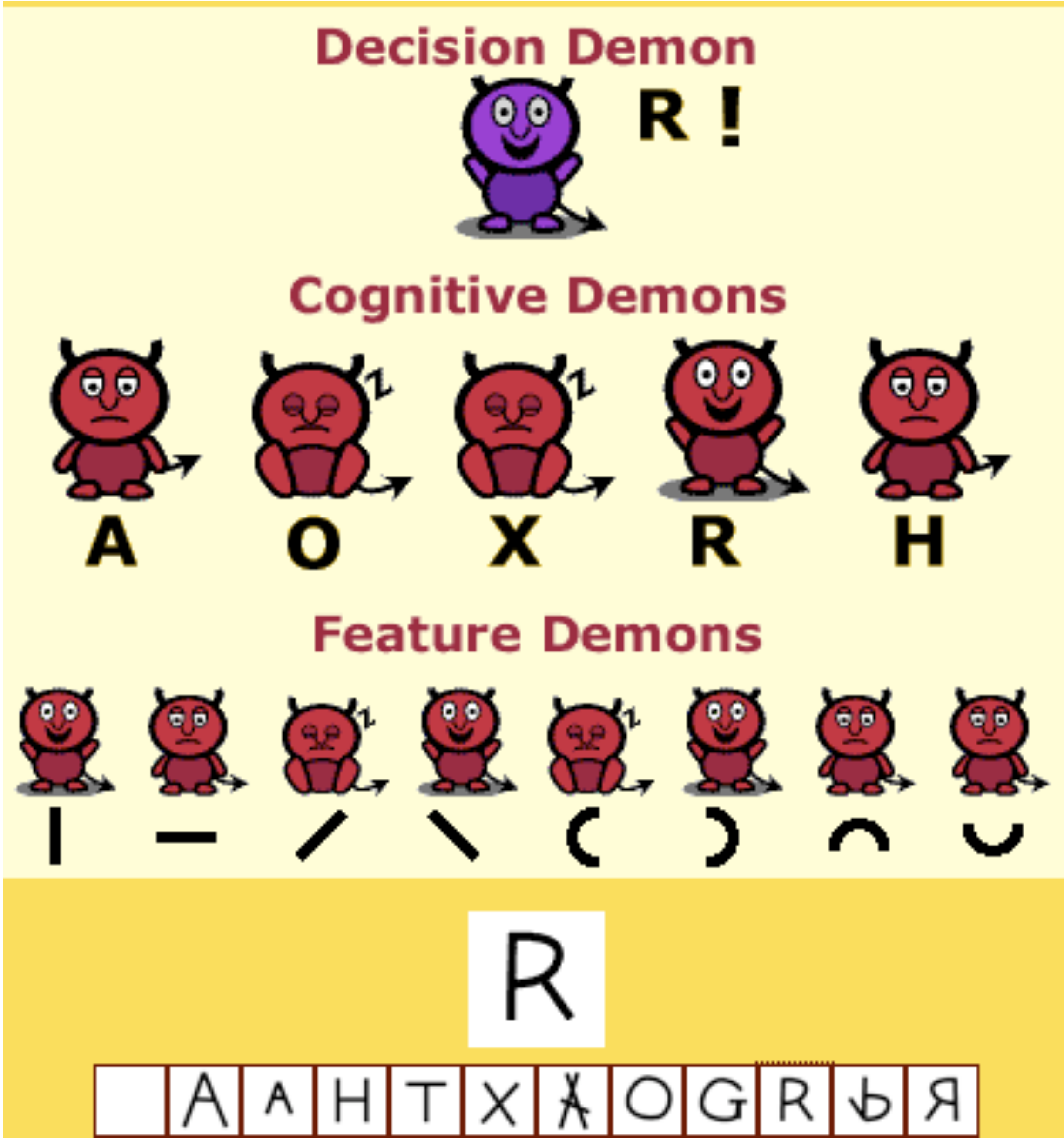
- Oliver Selfridge's (1959) simple model of letter recognition
- Perceptual committee made up of "demons"
    - Demons loosely represent neurons
    - Each level is a different brain area
- Pandemonium simulation:

https://oup-arc.com/access/content/sensation-and-perception-5e-student-resources/sensation-and-perception-5e-activity-4-5?previousFilter=tag_chapter-04

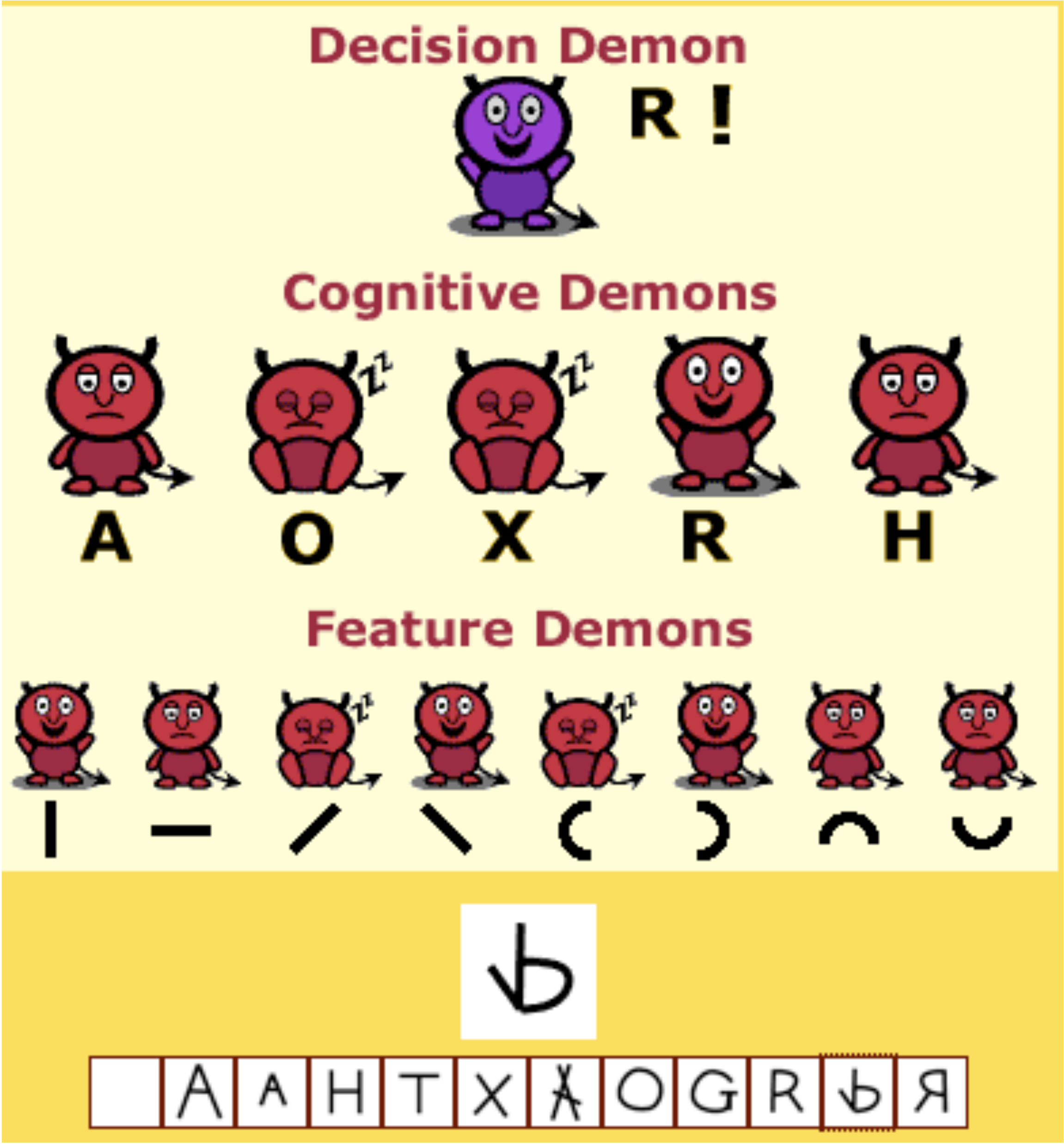
Models of Object Recognition



Models of Object Recognition



Models of Object Recognition



Models of Object Recognition

- Hierarchical “constructive” models of perception:
- Explicit description of how parts are combined to form representation of a whole

Metaphor: “committees” forming consensus from a group of specialized members

- perception results from the consensus that emerges

rapid progress in “deep learning” methods for object recognition & scene understanding

Researchers Announce Advance in Image-Recognition Software (NY Times, Nov 2015)

http://www.nytimes.com/2014/11/18/science/researchers-announce-breakthrough-in-content-recognition-software.html

Captioned by Human and by Google's Experimental Program

Human: "A group of men playing Frisbee in the park."
Computer model: "A group of young people playing a game of Frisbee."

Captioned by Human and by Google's Experimental Program

Human: "Three different types of pizza on top of a stove."
Computer: "A pizza sitting on top of a pan on top of a stove."

Captioned by Human and by Google's Experimental Program

Human: "Elephants of mixed ages standing in a muddy landscape."
Computer: "A herd of elephants walking across a dry grass field."

Captioned by Human and by Google's Experimental Program

Human: "A green monster kite soaring in a sunny sky."
Computer: "A man flying through the air while riding a snowboard."