Object Perception

- Perceiving an object involves many cognitive processes, including recognition (memory), attention, learning, expertise.
- The first step is feature extraction, the second is feature grouping which leads to basic perceptual experience. This is followed by recognition (matching with memory).
- Object perception will be examined from multiple perspectives.

The Challenge

1. When you view an object from different angles and distances, the image on the retina changes, but you perceive the same object.
2. Different “objects” can be members of the same category. Imagine different types of chairs. Do they have basic visual elements in common?
3. Part of the view of an object can be occluded (hidden) yet recognition still works.

All of this means that the image on the retina is *ambiguous* yet you (as perceiver) generally perceive the world accurately.

Perspective 1 - Perceptual Organization

Perceptual organization refers to the mental and physiological steps that group parts of the visual world together to form objects.

In the visual system, there are separate pathways that work on different aspects of vision. Within each pathway, different units respond to different properties. Work on perceptual organization seeks to uncover the rules that govern how basic elements are combined to form objects and groups.

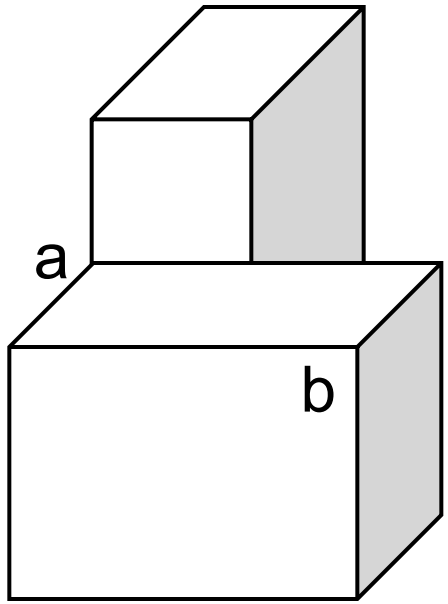
Perceptual Organization - Demo

What do you see?

Photograph by R. C. James

Demo - 2

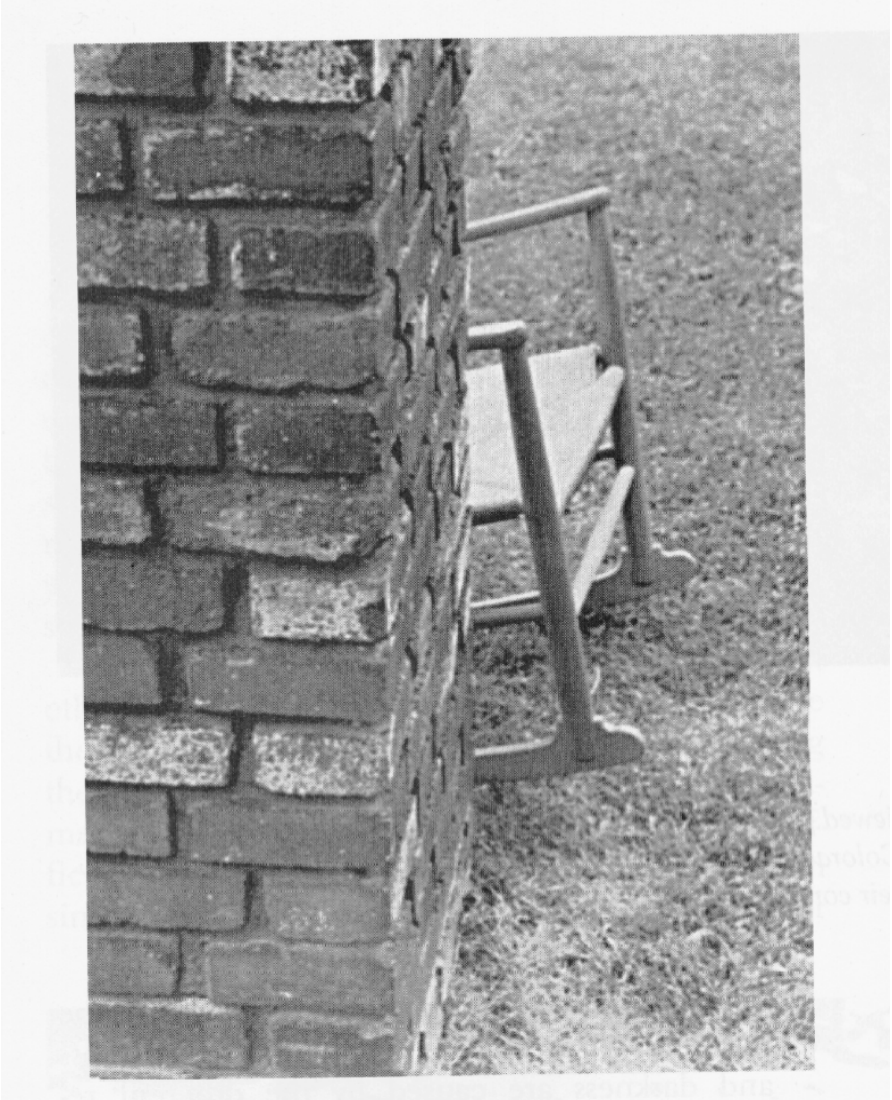
The intersection of three edges at point “a” is the same type of intersection as the three edges at point “b”. For point “a”, this is where two different objects meet. For point “b”, these three edges are all part of the same object. How does the visual system sort this out?



Demo - 3

The object in the photograph is partly hidden. What is its shape? What is the object?

How do humans perceive shape and recognize objects in this type of situation?



Gestalt Psychology

In the early 20th century, European psychologists such as Wertheimer proposed that perception of an object was based on a set of laws of perceptual organization. Their basic idea was that perception was a result of the mind (brain) imposing organization upon the pattern of stimulation that was received by the senses.

These laws became known as the Gestalt Laws of Perceptual Organization.

Apparent Motion

(1)

Flash line on left

(2)

50 msec of darkness

(3)

Flash line on right

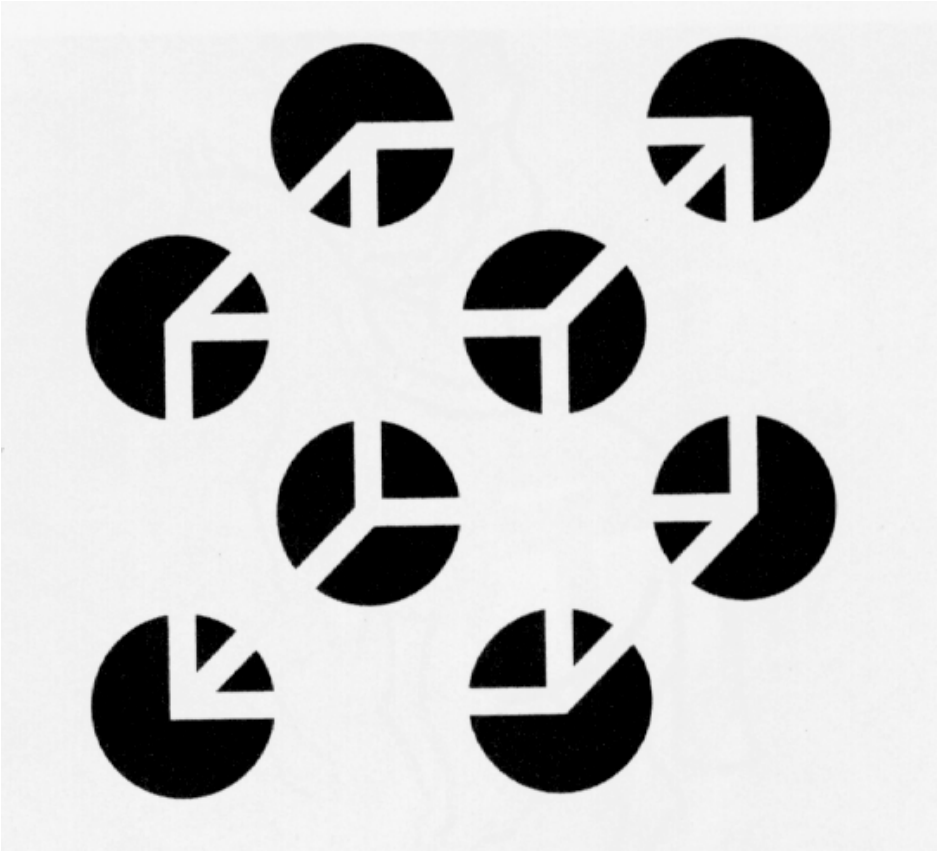
(4)

Subject reports movement, left to right

Illusory Contours

If you see a cube, does it appear that the edges continue across the white areas between the black circles?

These are illusory contours.



Gestalt Basics

Both the apparent motion and illusory contours are illusions. In both cases, we "see" things that are not present. If perception were simply the result of summed sensations, this could not happen.

These illusions reveal the operation of the laws of perceptual organization. The laws describe how the mind imposes structure on the output of the senses. That is, perception is more than the sum of its parts.

The Gestalt Laws

1. Pragnanz - also known as good figure or simplicity. Organize to make the simplest resulting object(s).

2. Similarity - Elements with similar properties are grouped together.

3. Good Continuation - Elements that form a straight or smooth curve are grouped together.

4. Proximity - also known as nearness. Elements close to one another are grouped together.

5. Common Region – elements within a single region (defined by lightness, color, contour) are grouped.

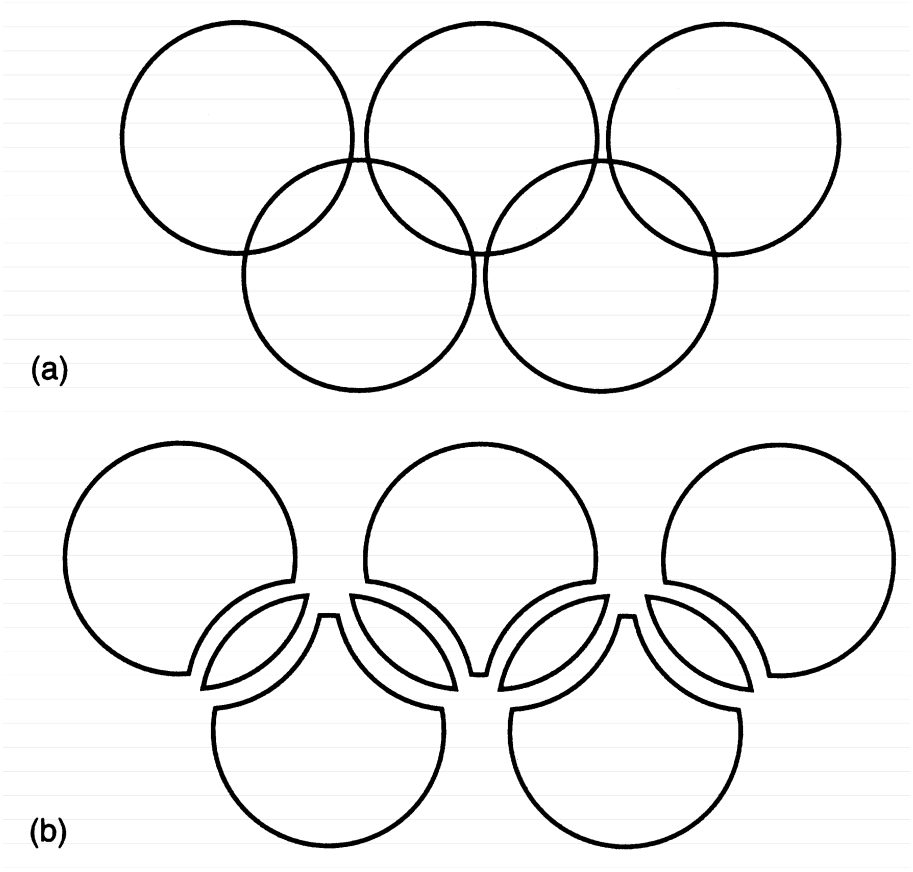
The Gestalt Laws

6. Common Fate - also known as common motion. Items moving together (speed and direction) are grouped together.
7. Familiarity – also called meaningfulness. Once a scene is organized a particular way, when the scene is repeated, the organization will also be repeated. If a scene is inspected for a particular “object” and it is found, on repetition, it is easier to see the object.

Pragnanz (simplicity)

The figure at the top is perceived as five rings rather than as the nine objects at the bottom.

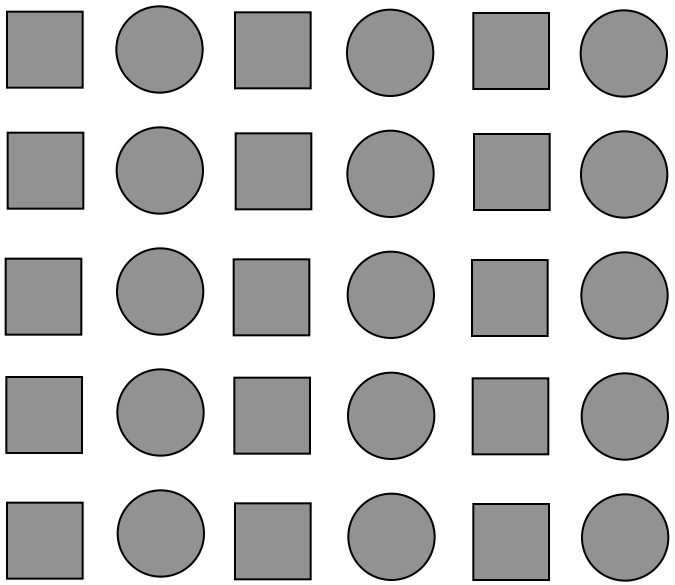
The five rings is "simpler" - it is fewer objects and they don't require accidental alignment.



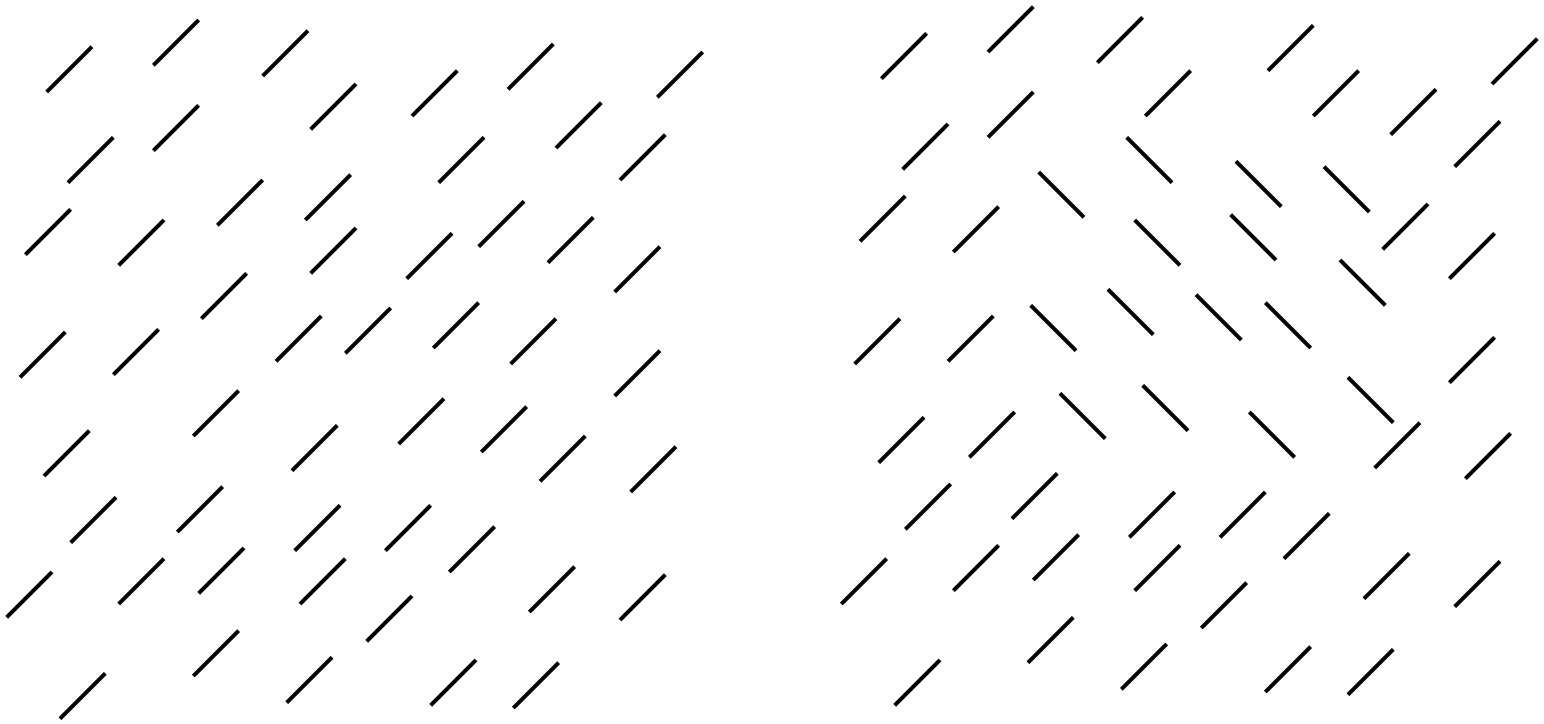
Similarity

Similarity refers to elements with common basic attributes such as hue, lightness, orientation, width (spatial frequency), and retinal disparity (stereopsis).

Do you see the grid at right as rows or columns?



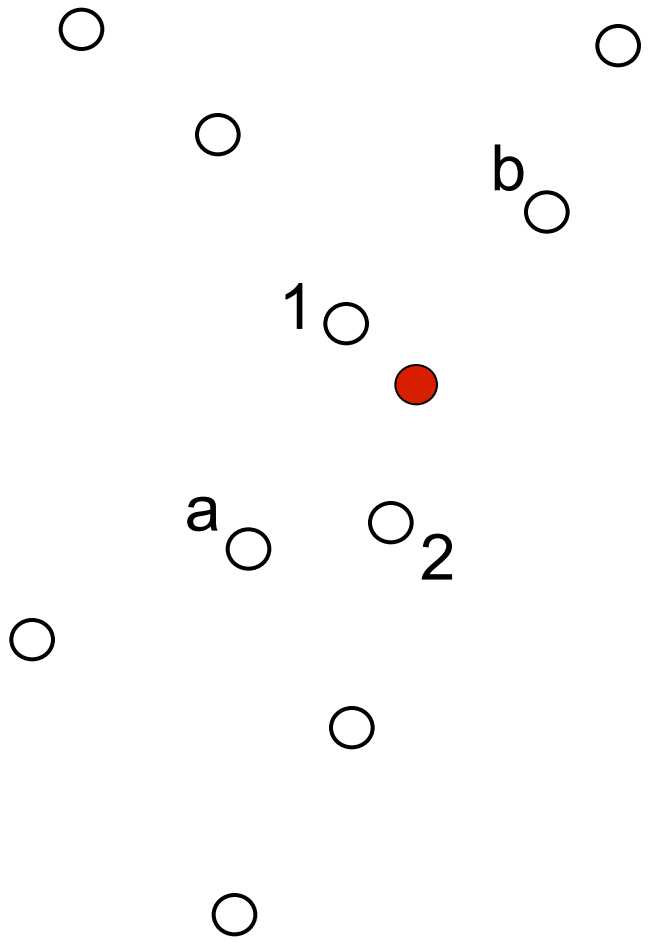
Pop-out Boundaries



Good Continuation

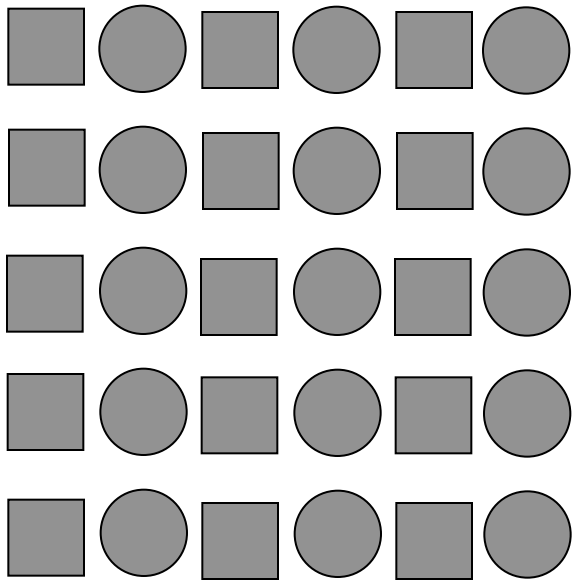
Does the red (gray) dot belong to the same group as dots a and b or 1 and 2?

The red dot is closer to 1 and 2, yet many observers say that it belongs with a and b. This is good continuation.



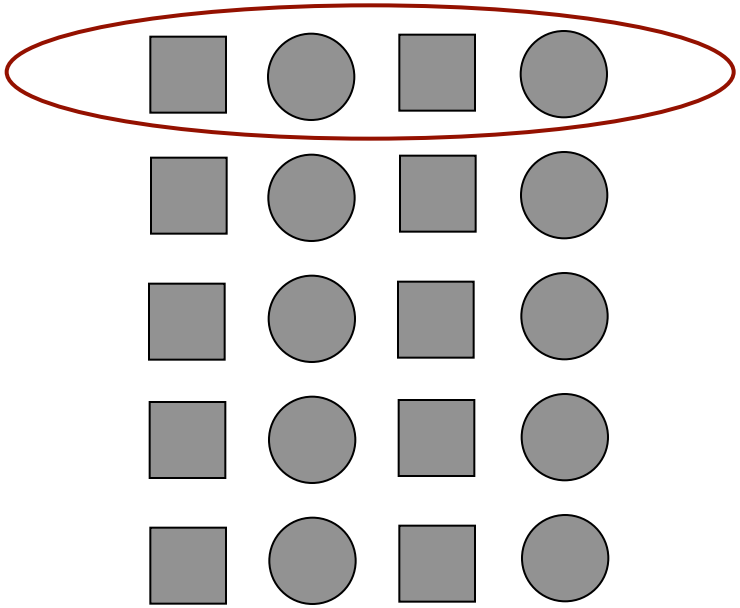
Proximity (nearness)

In this display, do you see rows or columns. Here, there is a tendency to see rows. However, if grouping were by similarity, it would be columns. Proximity, in this case, dictates organization as rows.



Common Region

In this display, the items within the red ellipse tend to group. This happens even though columns would be preferred by similarity.



Common Fate (common motion)

In common fate, elements that move together are seen as part of the same object. The five dots below on the left are seen as a single group because they move in the same direction as the same time. The five dots on the right are seen as two separate groups. The first, third and fifth are one group and the second and fourth form the second group based on common motion. Here, common motion overrides proximity.

Meaningfulness and familiarity

The Gestalt Psychologists also recognized a role for learning in perceptual organization. This has been called meaningfulness or familiarity. All other things being equal, a set of elements will be organized in the same way that it was previously organized or that a similar, recent set was organized.

For example, once you have seen the high contrast photograph by R. C. James as a Dalmatian, it is virtually impossible to see it as anything else.

Figure and Ground

Another important principle is that of Figure-Ground segregation. This refers to the process of separating the elements that form an object from the background.

1. The figure is more "thinglike".

2. The figure is in front of the ground.

3. The ground is seen as uniform and continuing behind the figure.

4. The contour separating figure from ground belongs to the figure.

Figure - Ground 2

This image can be seen as a white vase on a blue background or two blue faces against a white background. However, it can not be seen as both at once.

Figure - Ground 3

What governs when something is seen as Figure versus Ground?

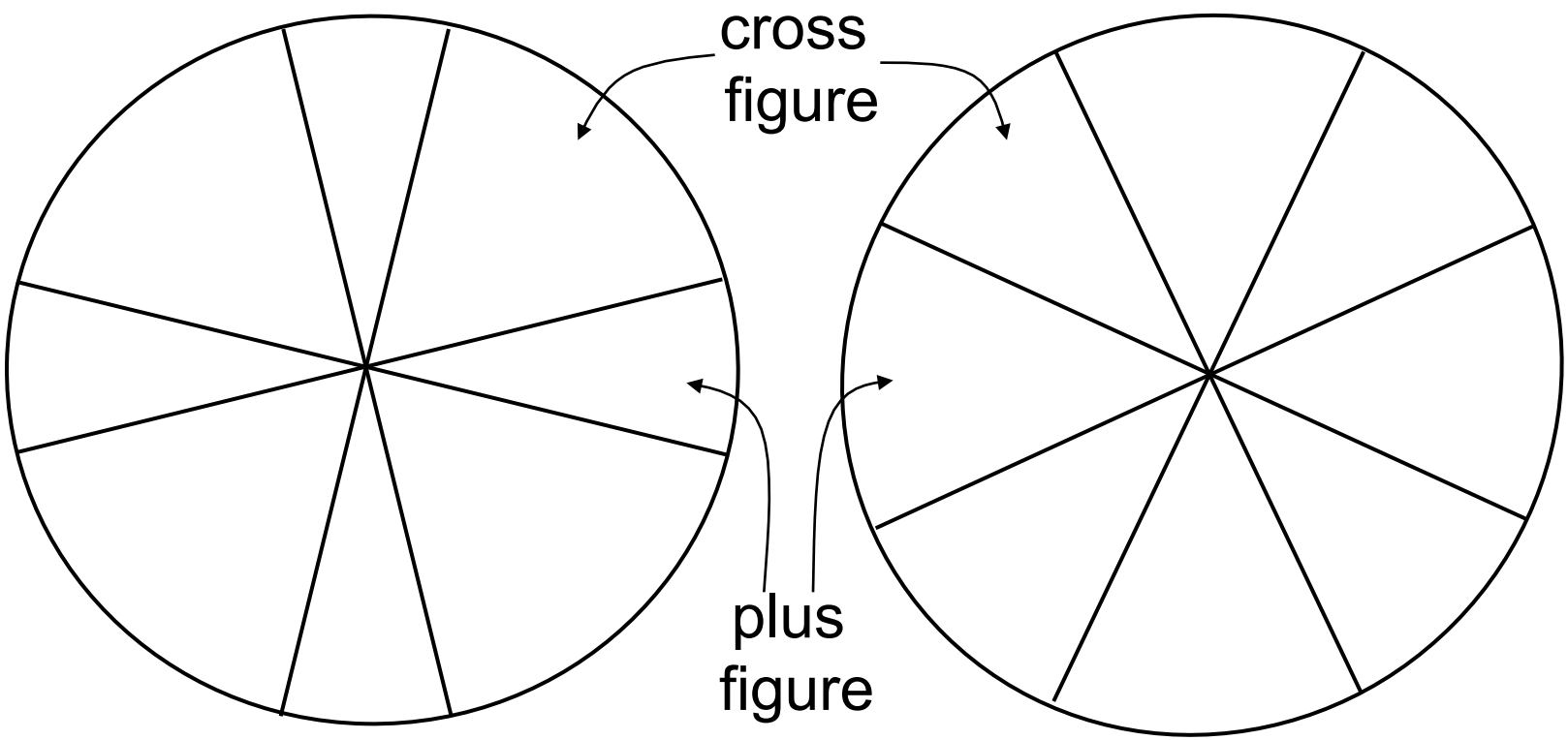
1. The smaller area tends to be seen as Figure.

2. Symmetrical areas tend to be seen as Figure.

3. Vertical and horizontal areas are more likely to be seen as Figure than other orientations.

4. Meaningful objects are more likely to be seen as Figure.

Figure - Ground Area



The smaller area plus is more likely to be seen as Figure (the plus).

The smaller area cross competes with the vertical-horizontal plus.

A critique of the Gestalt approach

1. Are the Gestalt Laws anything more than a statement of the obvious?

Yes and No. They are a description of the regularities in the world around us. As such, it is important to codify the rules by which perception operates.

In addition, they do lead to some testable predictions about situations where illusions occur and perception does not mimic the real world.

Gestalt Critique - 2

2. Some of the laws are vague and uninterpretable.

Some of the terms in the original laws were very vague. Terms like simplicity and similarity are very difficult to define. Lacking a definition, the laws are little more than an “after-the-fact” description.

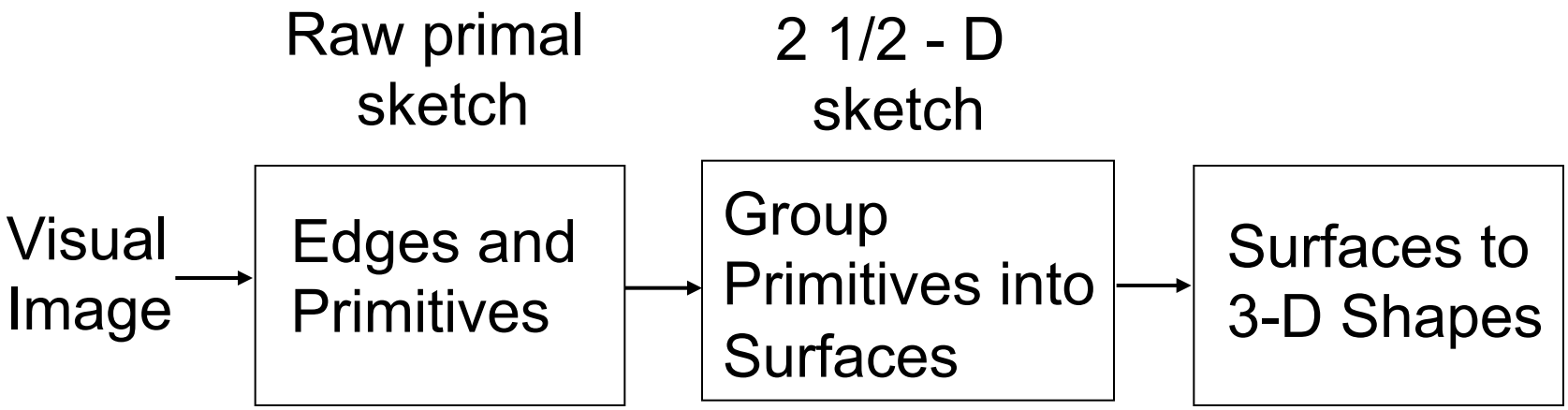
Further research over the past 30 years has provided precise, quantified definitions of some of these terms. For similarity, we specified lightness, color, orientation, width and retinal disparity: the qualities extracted by simple cells in the primary visual cortex.

Gestalt Critique - 3

3. The laws offer only after-the-fact explanation.

Until all of the laws are precisely specified, this is true. Only when all are quantitatively specified can we figure out how they combine to determine perception. Until then, the laws are "work in progress" that help us in understanding perception, but are incomplete.

Computational Approach (from Marr)



Computational Approach - 2

The computational approach, as represented in the work of Marr, emphasizes that there are "ecological constraints" - basic properties of the environment. The visual system uses these to convert a representation based on primitives into a viewer-centered representation of surfaces and depth.

These constraints, which include common motion, common orientation, proximity, and similarity are very much like the Gestalt laws of grouping.

This 2 1/2 - D representation is then mapped to an object-centered 3-D representation. This is the point at which object recognition occurs.

Object Recognition

The very interesting question remains, however, as to how an observer manages to recognize an object in spite of viewing it from different angles.

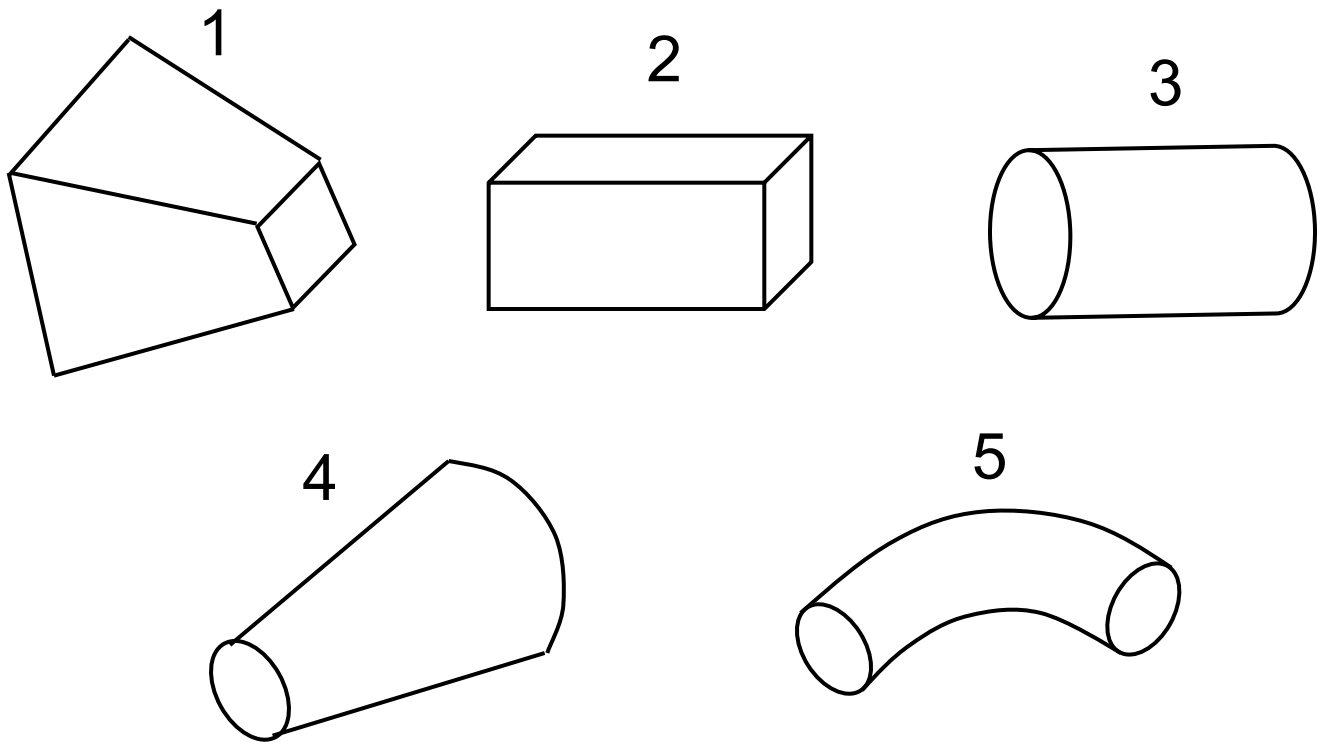
One approach is to construct an object-centered, 3-D representation using volumetric primitives. This is what the last step in Marr's approach seems to require. Biederman has proposed such a model of object recognition - Recognition by Components.

RBC

Objects are represented a a group of connected volumetric primitives (GEONS). Each object is composed of a set of GEONS in a specific arrangement.

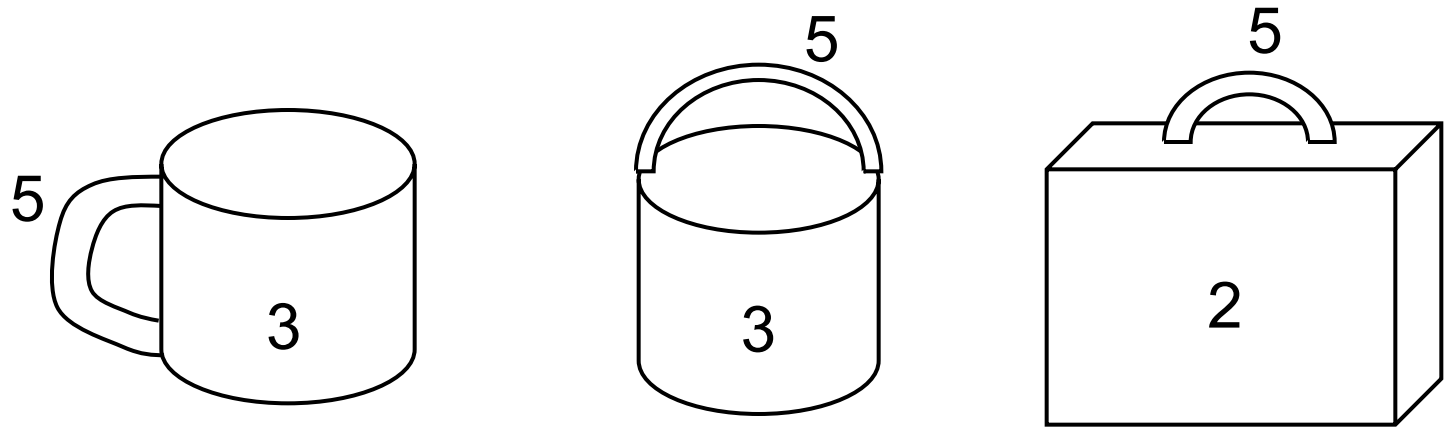
The GEONS represent simple 3-D shapes based on Non-Accidental Properties of edges. For example, a curved edge in the visual image corresponds to a curved edge on the object. Except when viewed edge-on, a curved edge in the real world always projects a curved edge on the retina.

Object Recognition - 2



These are GEONS - basic volumes (3-D shapes) from which objects are built (the components).

Object Recognition - 3



Each different object is composed of a set of GEONS that are arranged together in a particular way. The arrangement is *view - invariant*.

Object Recognition - 4

Some basic data:

1. If the edges of GEONS are obscured, object recognition is still relatively fast and accurate.

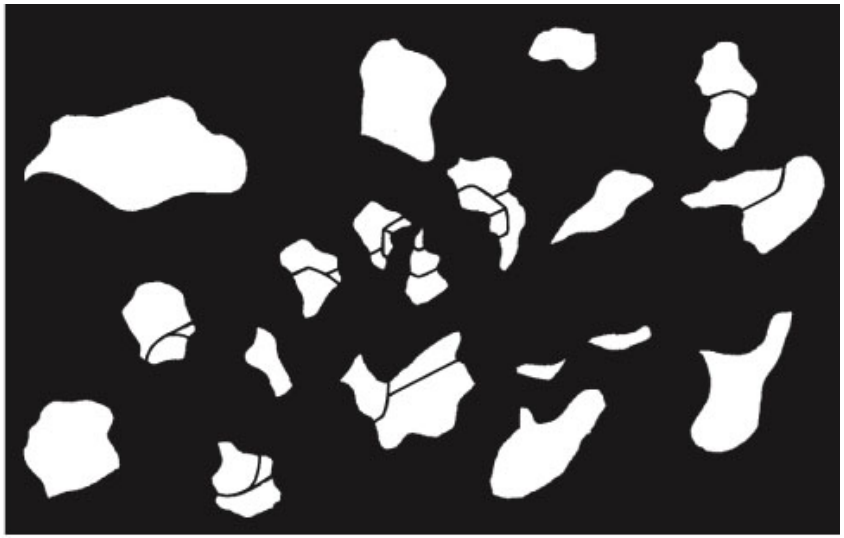
2. If the junctures where GEONS join are obscured, object recognition is very slow and difficult.

3. Object recognition is often faster for a basic line drawing than for a color picture for many objects.

4. Object recognition is difficult if the angle of view obscures the GEONS or the arrangement of the GEONS. See Figure 5.32 in text.

The object below is the same in the left and right drawings but the object on the left has the junctures between Geons obscured and is harder to identify. The same total amount of contour (edges) is present in both displays.

(a)



(b)

More examples of easy and hard to recognize images.

Knowledge and Object Recognition

Thus far, all of the process that are part of object perception have been "bottom-up". That is, they are based on the information in the visual world as represented on the retina. Put another way, we have discussed perception as a "data-driven" process.

Perception can also be influenced by knowledge. For example, perception could be influenced by what is likely to appear in a visual scene. This is "top-down" processing because the influence is from knowledge that the person already has. This is also referred to as "conceptually-driven" processing.

Scene Perception

In the real world, objects are typically seen in the context of other objects – a scene. How is perception of a scene similar to or different from that of objects?

Some basics:

1. It takes about 250 ms to identify a scene that you are looking for. About 500 ms to identify details (based on flashing pictures briefly).

2. Scenes have global image properties (regularities in the world).

Image Properties

1. Natural scenes tend to have undulating contours. Man-made scenes (a city street scene) has straight line edges.

2. Natural and man-made scenes have more horizontals and verticals than oblique edges.

3. Some scenes have characteristic colors (blue sky).

4. Open scenes typically have a visible horizon line.

Image Properties (cont.)

5. Shape from shading. Most of the time, light comes from above. If this is the case, then the presence of shadows (shading) can inform us about the nature of a surface (see examples in 5.37 and 5.38 in text).

6. Familiarity (meaningfulness).

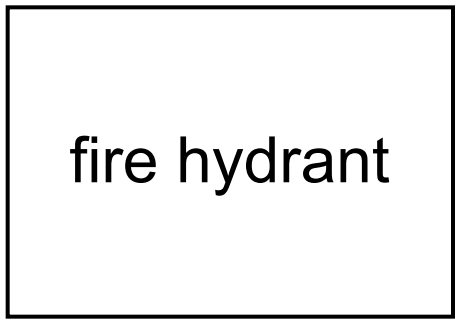
Influence of scene familiarity (meaningfulness)

If an object were to occur in an unusual environment, or an unusual location in an environment, would it be easier or harder to identify?

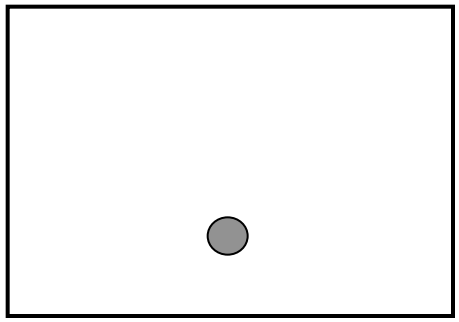
If we were to preserve the visual information right around an object, but scramble the rest of the visual world, would the object be easier or harder to identify?

Experiment

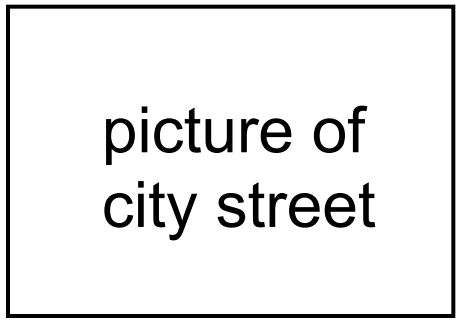
Biederman did a series of studies looking at the effect of context (the visual scene) on the speed and accuracy with which people recognize objects. The observer's task is to indicate (yes or no) whether the object named (slide 1) occurs at the location (slide 2) in the visual scene (slide 3) and respond as fast as possible.



slide 1

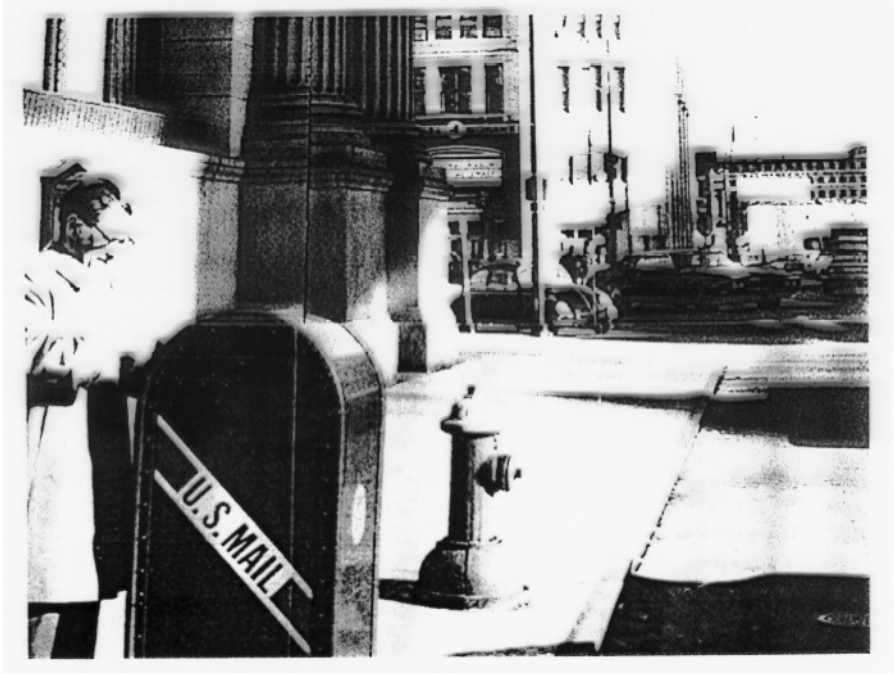


slide 2

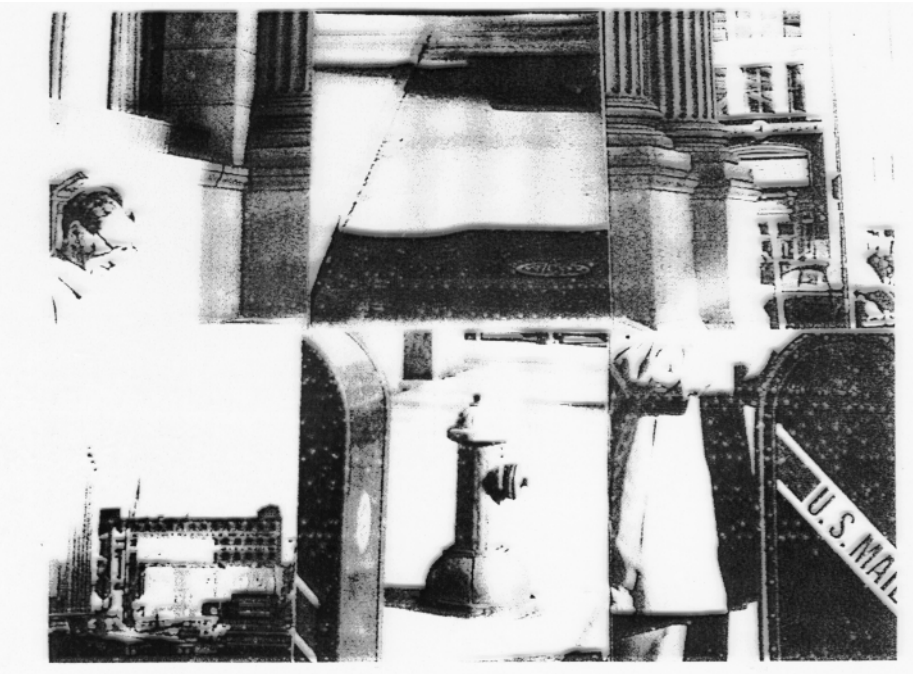


slide 3

Experiment Scenes (from Biederman, 1973)



Normal scene - observers are fast and accurate to identify fire hydrant.



Scramble scene - observers are slower and less accurate to identify fire hydrant.

Results

In the scrambled street scene, the observer knows were to look for the object and the image right around the object (the local visual image) has not been changed.

However, observers are much slower (and make more errors) identifying the objects in the scrambled scenes.

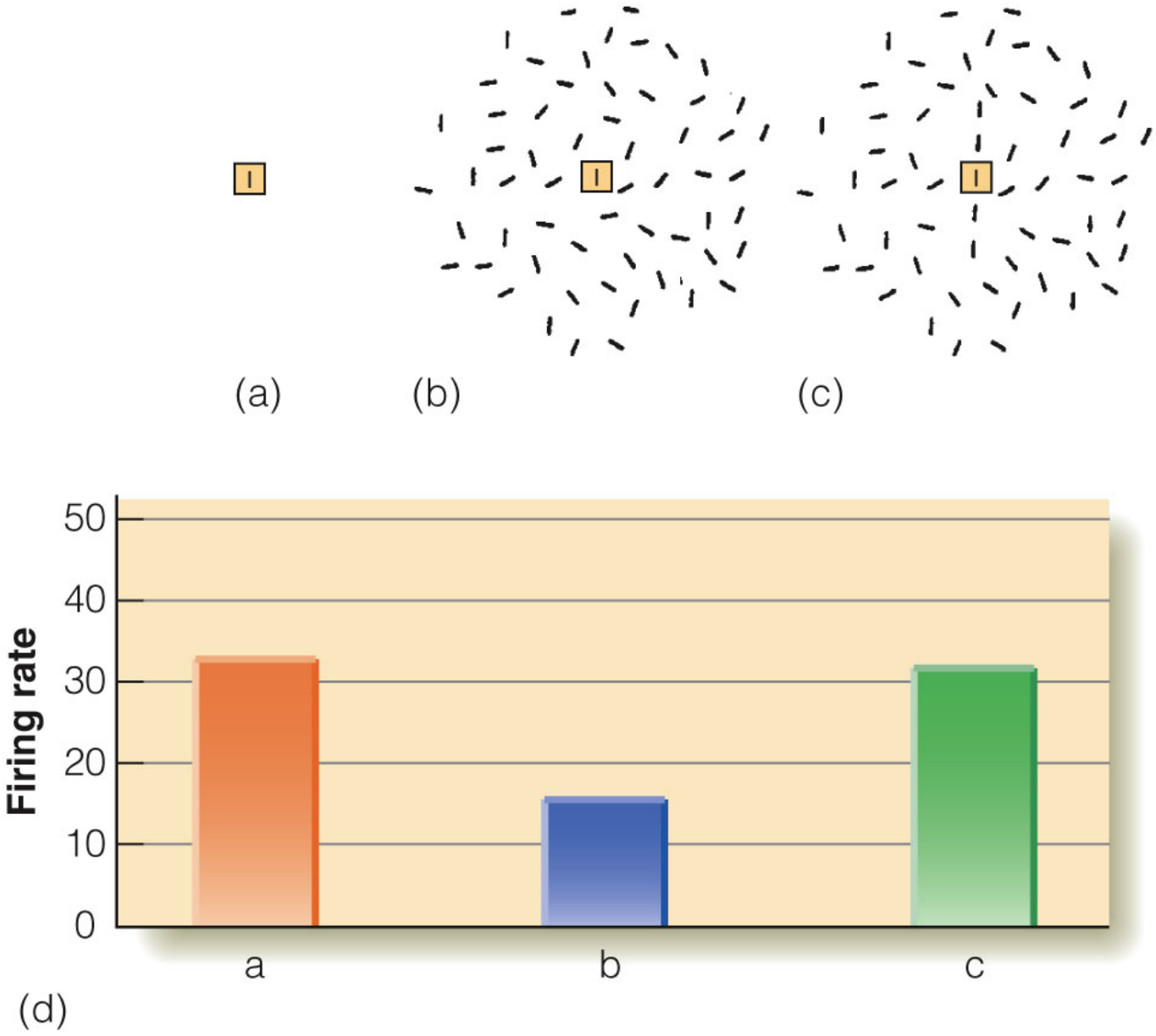
Using line drawings, Biederman also found that putting an object in an unusual position in a scene or in a scene that it did not belong in made observers slow and error prone in identifying the object. (See also Palmer, 5.40 in text).

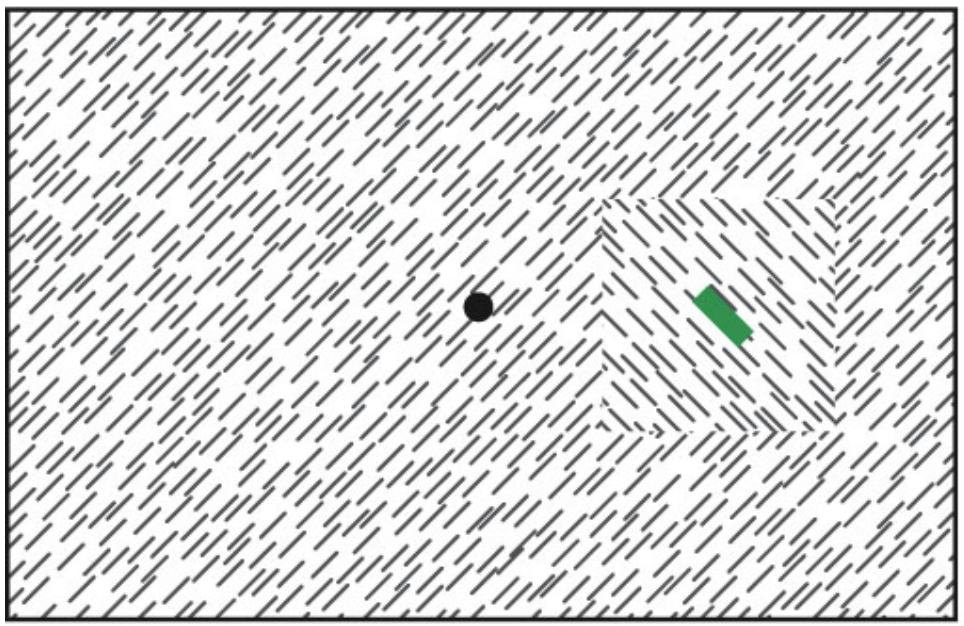
Neural Coding and Object/Scene Perception

1. Individual cells, starting in V1, exhibit response properties that are like the Gestalt laws.

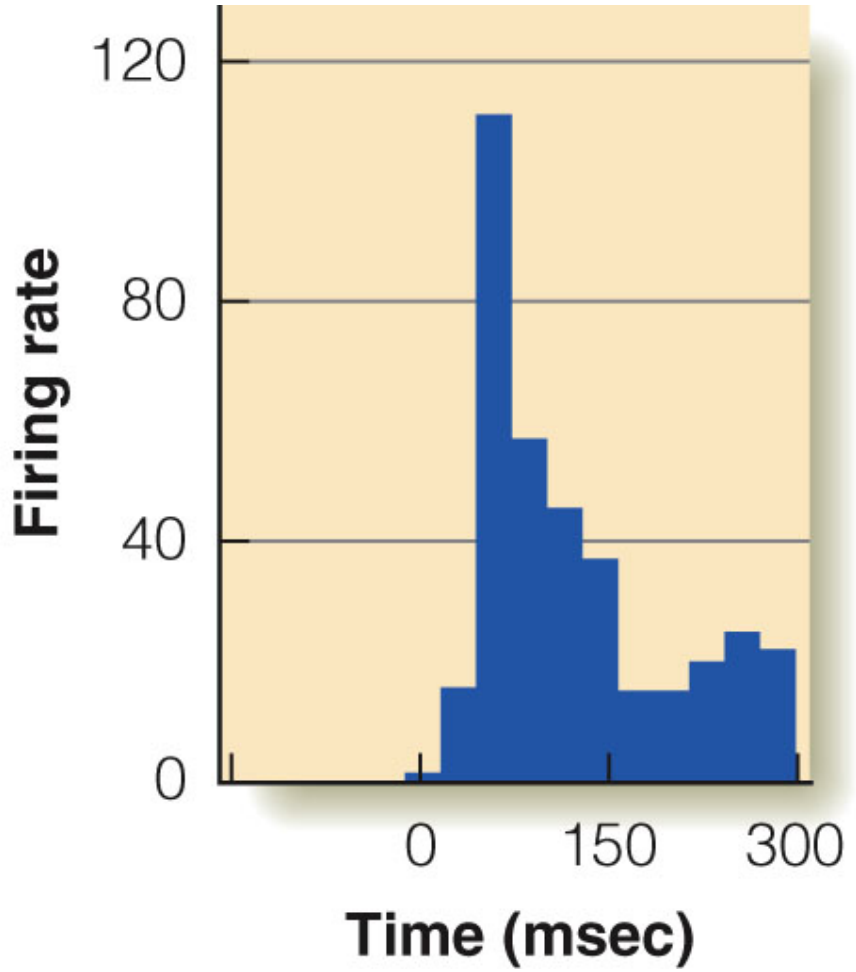
2. Some cells respond only when the “feature” in their receptive field is a part of a Figure (not ground).

3. In humans, the fusiform face area seems to code facial recognition.


(a)
(b)
(c)
Firing rate
50
40
30
20
10
0
a
b
c
(d)



(a)



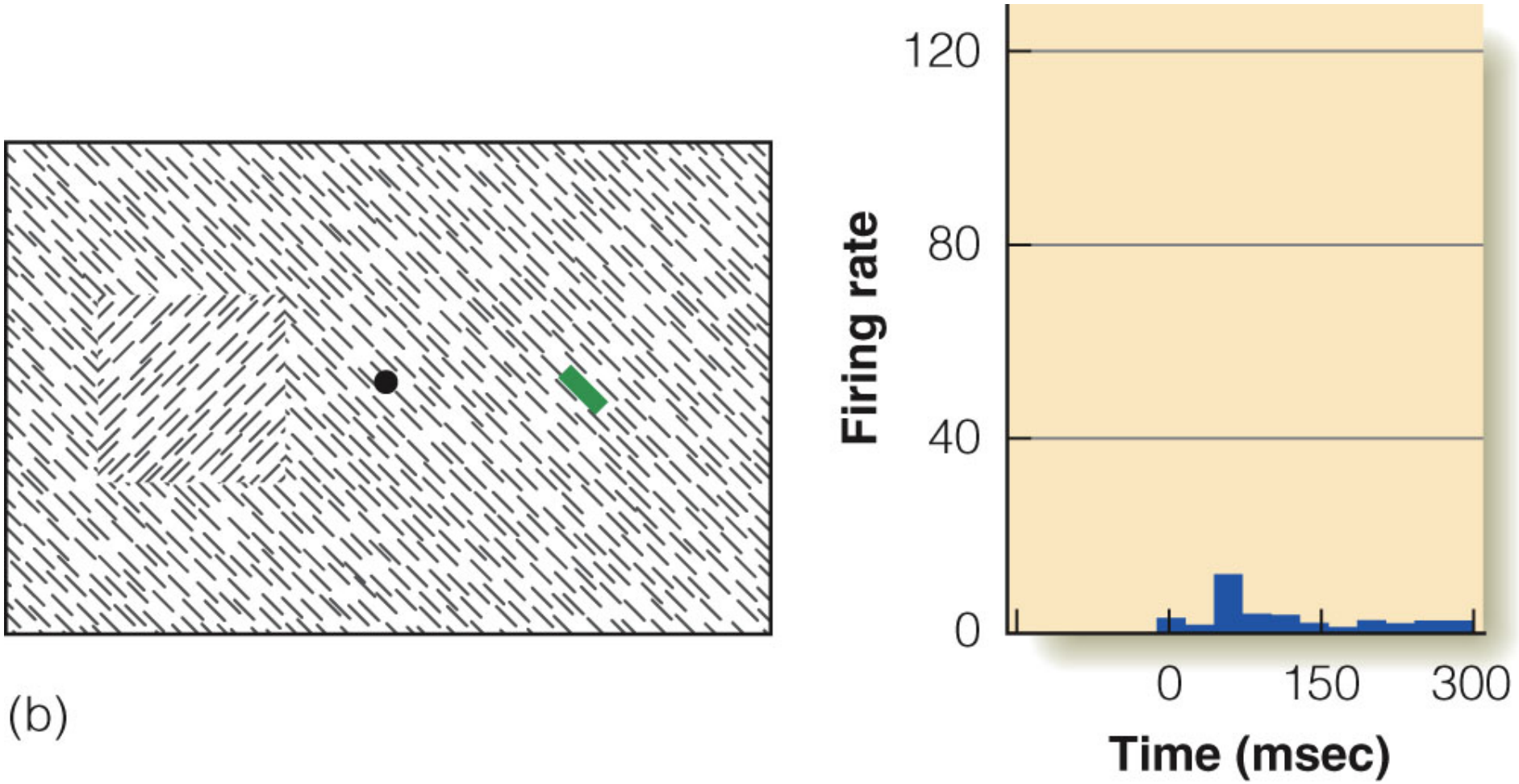


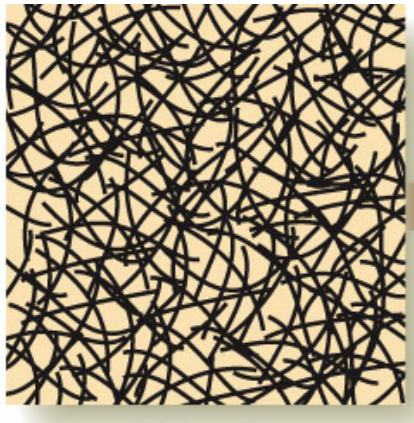
Fig. 5-44b, p. 121

Facial Recognition and fMRI

Grill-Spector asked participants to respond to rapidly presented pictures while fMRI imaged their brain activity. On each trial, a picture of Harrison Ford, a different face or a random texture were presented. The pictures were presented for 50 ms and followed by a pattern mask.

Viewers responded “Harrison Ford”, “Another face”, or “Nothing”.

Stimulus

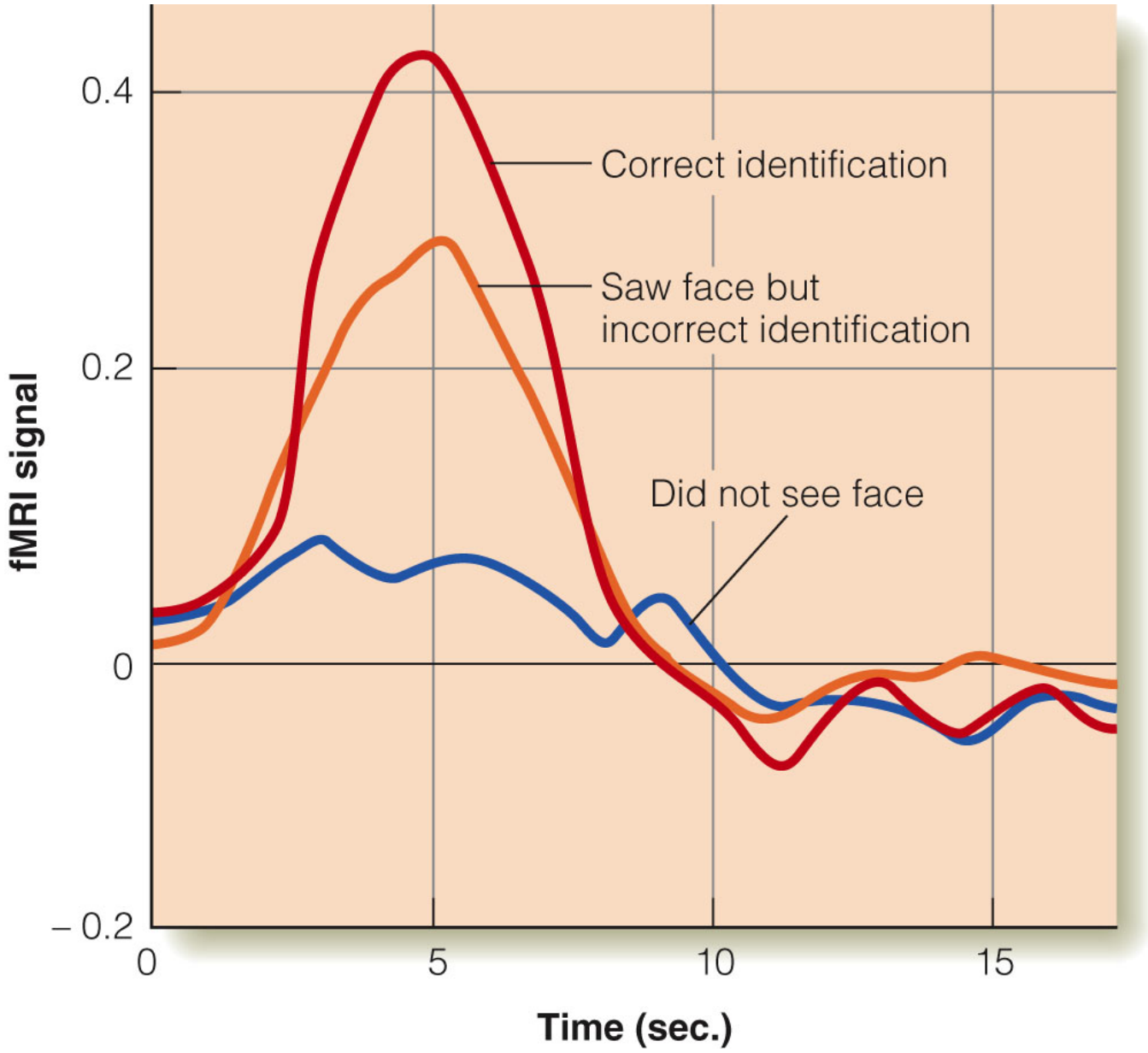


Mask

Observer's response

See either
(a) Harrison Ford
(b) Another person's face
(c) A random texture

Indicate either
(a) "Harrison Ford"
(b) "Another object"
(c) "Nothing"


0.4
0.2
0
– 0.2
fMRI signal
0
5
10
15
Time (sec.)
Correct identification
Saw face but
incorrect identification
Did not see face