The Physiology of the Senses
Lecture 3: Visual Perception of Objects

www.tutis.ca/Senses/

Contents

Revised 19/09/2012

Objectives

1) Select the key difference in the way the retina is mapped in the visual areas V1, V2 and V3.
2) Specify the mechanism that allows the features encoded by primary visual cortex to be grouped into objects?
3) State the 2 key areas within the ventral stream involved in the coding of objects.

Where does visual information go to after V1 (primary visual cortex)?

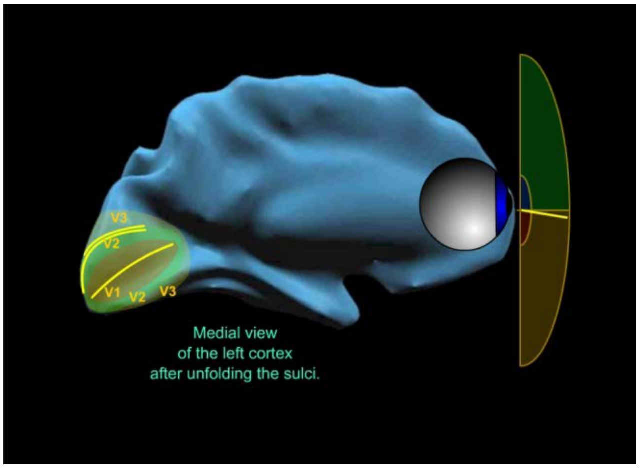
From V1 information is sent to higher order visual areas, first V2 and then to V3.

In each area the retina is re-represented.

Note that the  is seen in the right lower visual field and is represented in the left upper visual cortex. It is represented 3 times, once in each area V1, V2, and V3.

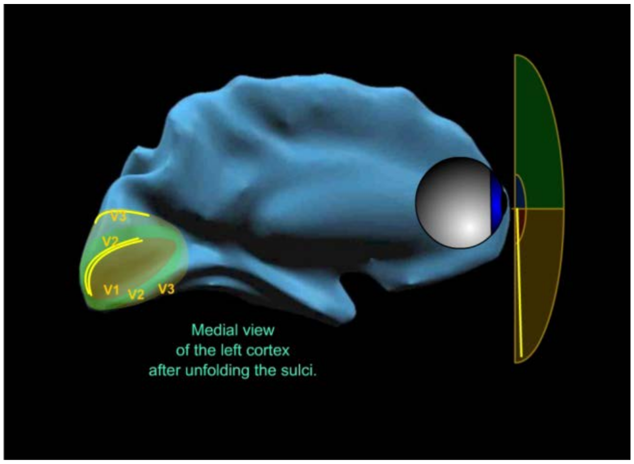
Notice the odd way in which three lines are represented in V1, V2, and V3.

A line just below horizontal is mapped first just above the calcarine sulcus in V1 and again along both sides of the V2 and V3 border.



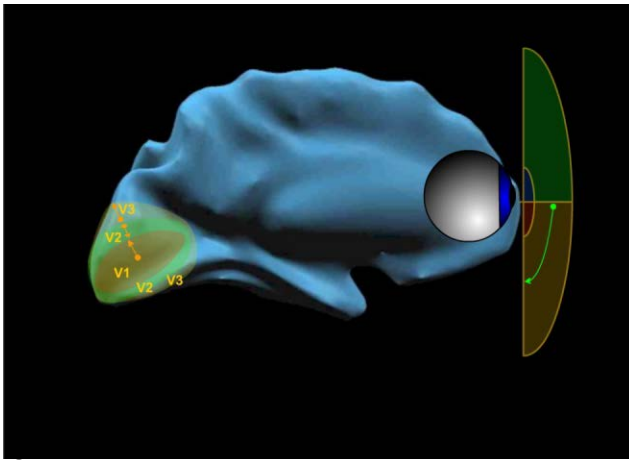
Also, a near vertical line is represented at the V1/V2 border and again on the far side of the V3 border.

These landmarks are used by scientists to map the edges of V1, V2, and V3.



This occurs because the representation of the arrow is mirrored at the V1/V2 border and again at the V2/V3 border. The same mirroring occurs for everything mapped in V1.

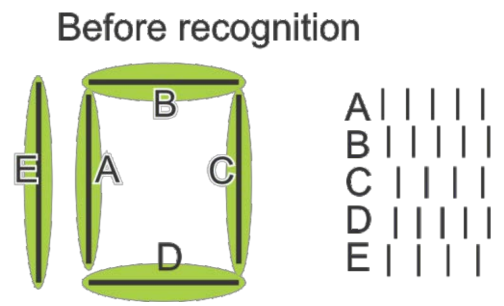
Note that the arrow heads are mapped near each other as are their tails. As we see again "like" cells like to be near each other. This reduces the length of the axons and of the wiring in the brain.



Revised 19/09/2012

Higher order visual areas begin assembling simple features into objects.

For example, it is here that the visual system starts to assemble lines and edges into objects. Here 4 lines form a box.

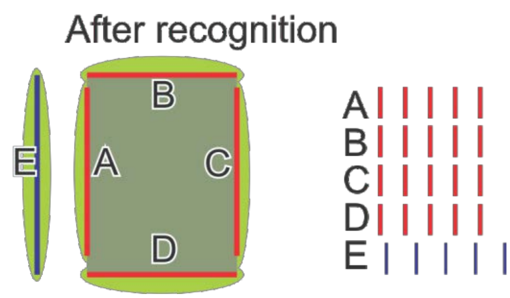


When one sees a box, a number of orientation specific simple cells in V1 are activated.

How might these cells signal the fact that they are all activated by the same box?

One theory is as follows:

1) Suppose each line activates one of the five V1 cells shown here. Before perceiving that 4 of the 5 lines belong to a box, all the cells are activated but asynchronously (i.e. they fire at different times).



2) After binding, 4 of the cells that are grouped with the box and begin to fire synchronously (i.e. they fire at the same time).

3) Thus V1 first "see" the elementary features of objects while higher areas, such as V2 and V3, begin grouping the features that belong to the same object.

4) This grouping is fed back to V1 producing synchronous activity.

The "Binding Problem"

In general, binding involves grouping features into objects.

Sometimes deciding which features should be grouped is obvious, as it is in the case of the lines that make up a square.

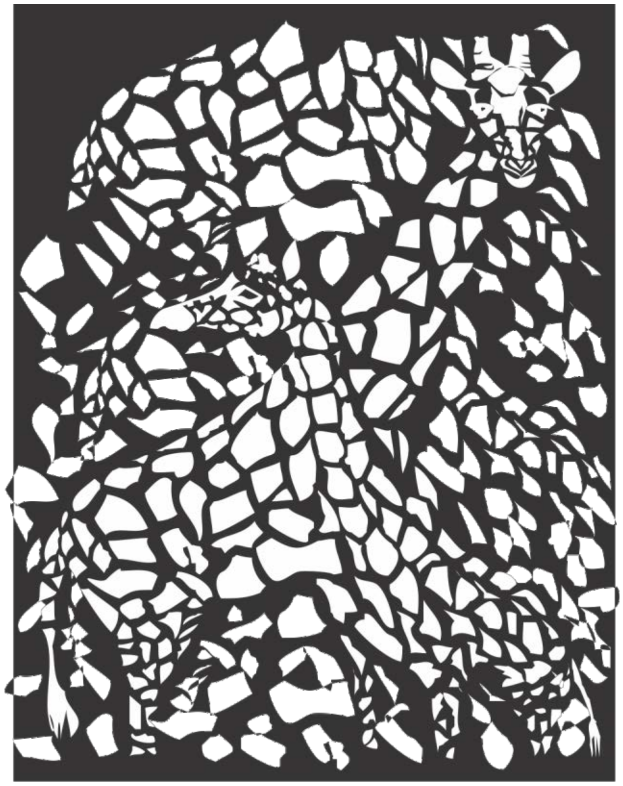
Sometimes it is more ambiguous, such as with the two giraffes shown here.

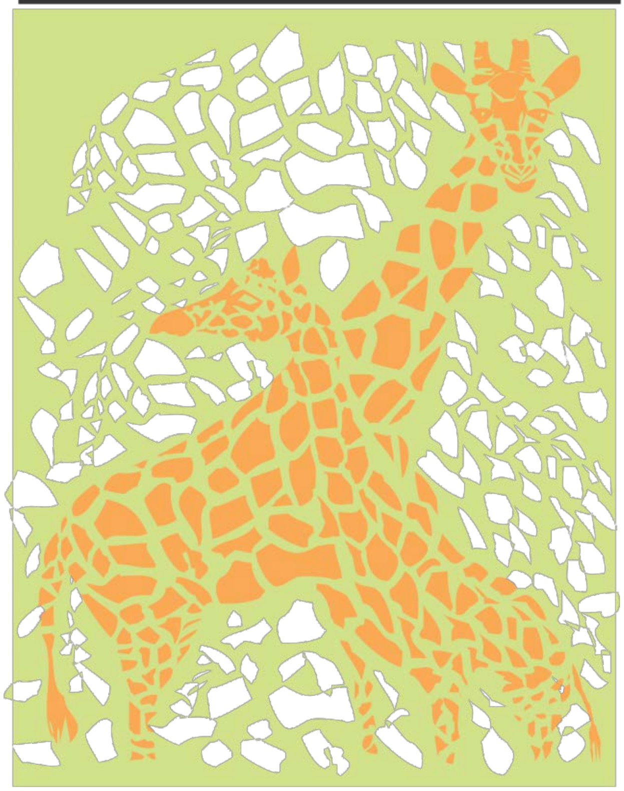
The brain uses common color, motion, or form (e.g. the nearness of lines or shapes) to group features that are common to an object. Note how much easier it is to see the giraffes when you add differences in shading or color.

Past experience is another important factor in binding. We see two giraffes because we remember what giraffes look like. Also once you make out where the giraffes are the in color, it is much easier to find them again later in black and white.

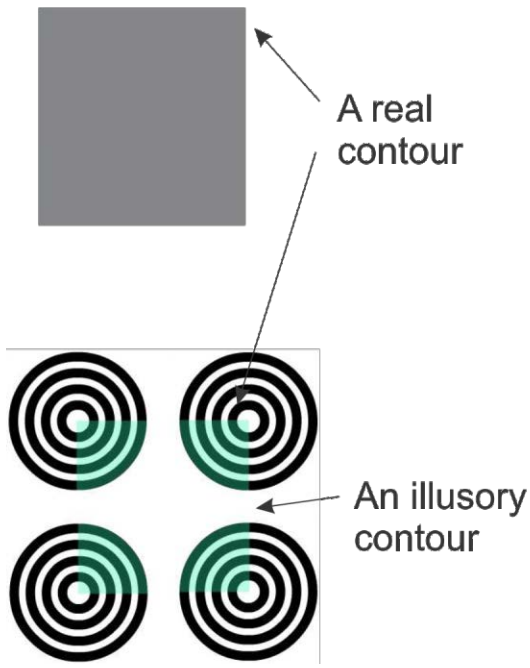
Which features are bound together is indicated by synchronous activity in cells that encode these features.

To allow for this synchronous activity to develop, the columns of cells in V1 (and other regions) have extensive reciprocal interconnections.





Illusory Contours



Note that one can see a square on the right even when there is no real square; an illusory contour. The visual system fills in a line from the corners.

Cells in V2, and some in V1, are activated by both a real contour and an illusory (or subjective) contour.

One has the illusion of a square formed by four lines in the figure when in fact there are no lines. Higher areas try to assemble objects by filling in the gaps between lines, as in this square.

You may also see a faint green color fill the entire square.

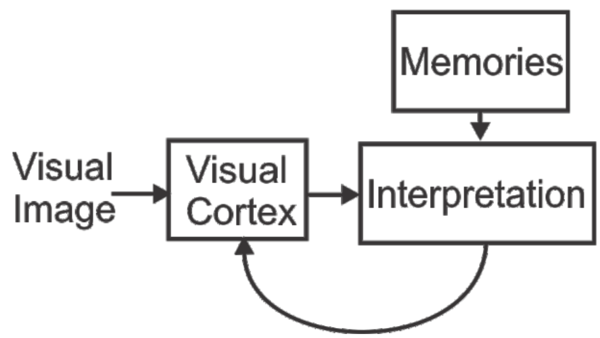
The color in the center of the square is defined from the corners. It is filled in from the edges.

Recall that retinal ganglion cells extract the edges.

As a result, color information from the center is lost.

This is reversed by this "filling in" process.

What we perceive depends on our interpretation of what we see.



Interpretation based on our memories modifies what we see. For example, if we expect to see the letter m in “exanple” we may not notice that is has been misspelled.

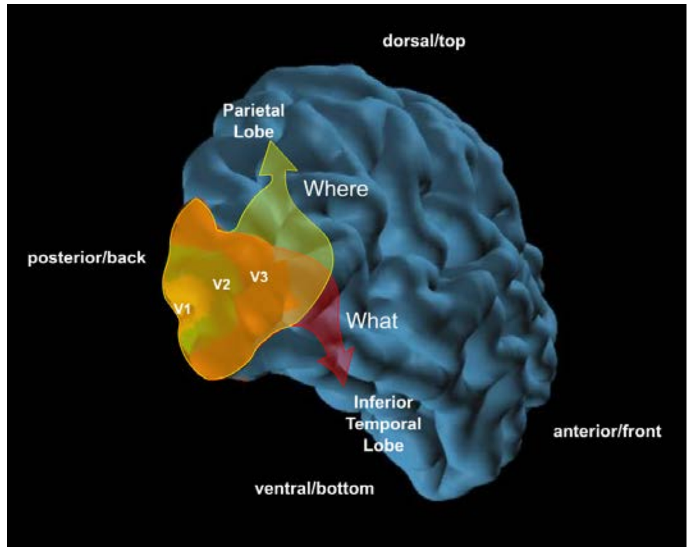
Another example

Acocdrnig to an elgnsih unviesitry sutdy the oredr of letetrs in a wrod dosen’t mttaer, the olny thnig thta’s iopmrantt is that the frsit and lsat ltteer of eevry word is in the crcreot ptoision. The rset can be jmbueld and one is stlil able to raed the txet wiohtut dclftfuiiy.

Visual Areas beyond V3.

From V3, information diverges to over 3 dozen higher order visual areas. Each processes some special aspect of visual information. These visual areas are like a multi-screen cinema. The main difference is that each screen is showing a different attribute of the same movie; some just the motion, others the colors, etc.

Here two processes separate into 1) the perception of edges and colors as objects and 2) the coding of their spatial attributes; for example, their location, orientation, and motion. Information flows along two main streams.

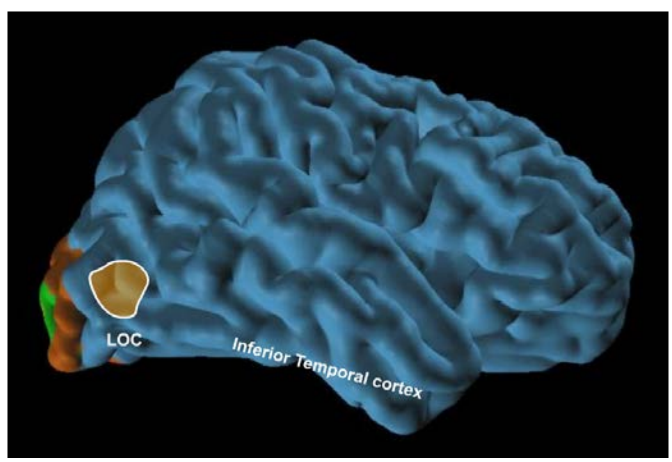


1) The dorsal stream (top surface), along the intra parietal sulcus, is concerned with selecting actions to particular spatial locations. For this reason this stream is called the "**where**" stream. We will discuss this stream in more detail in session 5.

2) The ventral stream (bottom surface), projecting to the inferior part of the temporal lobe, is concerned with the perception and recognition of objects, e.g. faces. This is called the "**what**" stream. We will concentrate on this stream in the remainder of this session.

Object perception begins in V1 which extracts simple features that are common to all images, e.g., lines. It ends in the inferior temporal cortex (IT) where cells respond to a particular combination of complex features, for example those that define a particular face.

In V2 & V3, the upper and lower visual quadrants are separated by V1. In V1 lines of the same orientation activate pinwheels of the same orientation. In V2 and V3, features that share common cues such as the same orientation are bound together. The Lateral Occipital Complex (LOC) combines object parts in the upper and lower visual quadrants but not, as yet, those in the left and right. Finally in IT regions such as FFA, the left and right sides are brought together and the object is recognised.



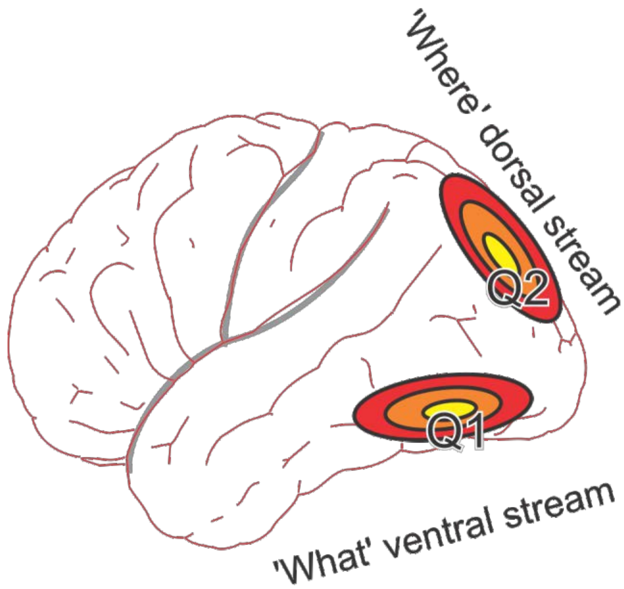
In area LOC, elements of objects are extracted from the background. LOC codes that something is an object part while areas of IT code a particular object (e.g. a rhinoceros). Lesions of LOC result in visual agnosia, the inability to perceive all objects through vision. Lesions in small areas of IT can result in visual agnosia of a particular class of objects (e.g. rhinoceros-agnosia).

Evidence for dorsal and ventral pathways.

In a functional imaging experiment, Ungerleider & Haxby asked subjects one of two questions.

Q1) Is this face the same face as that shown previously? This produced activity in early visual areas and a greater activity in the ‘what’ stream than in the "where" stream.

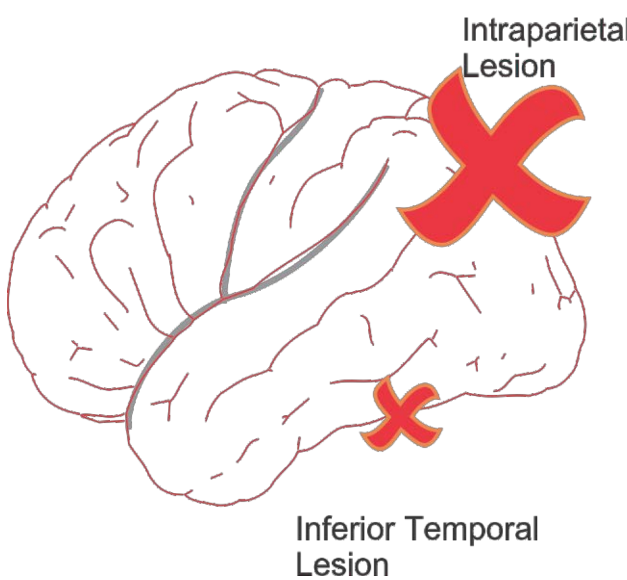
Q2) Is this face in the same location as that shown previously? This now produced a greater activity in the ‘where’ stream. The stimuli were identical! Remarkably, simply changing the task shifted which areas were most active.



If two streams exist, then one should find patients with selective loss of one or the other. This is indeed the case.

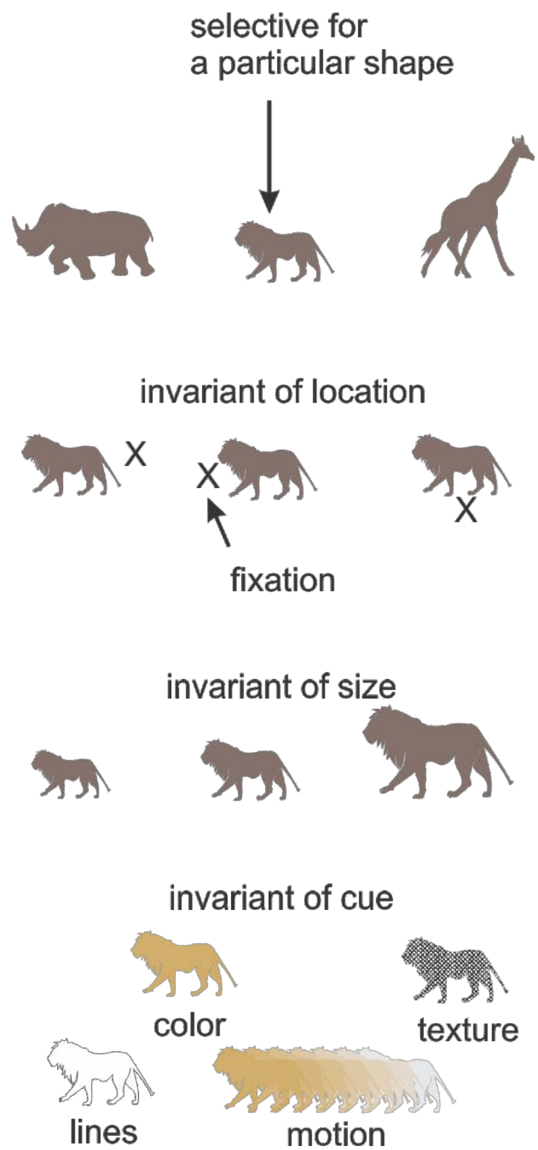
Patients with lesions of the intraparietal sulcus have difficulty in pointing or grasping accurately.

Small lesions in the inferior temporal cortex produce prosopagnosia, a specific loss of face recognition (a particular type of form agnosia)



What is the function of the inferior temporal cortex?

Recent studies of Nancy Kanwisher and others suggest that faces are represented by a small region located in the inferior temporal cortex called the Fusiform Face Area (FFA). Here all neurons respond preferentially to faces and a particular face is stored by a small number of highly selective neurons (sparse clustered population). A similar visual representation may hold for all objects.



Lesions of the FFA leads to prosopagnosia. Patients with prosopagnosia cannot recognize friends (or even themselves in a mirror) from visual clues but can recognize them through other modalities such as their voice or gait. Visual acuity and the recognition of colors and movement are not impaired. Patients can recognize features such as eye brows, lips, etc.

In the inferior temporal cortex;

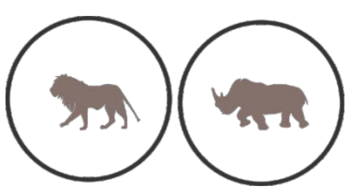
1) Cells respond selectively to a particular class of object e.g. faces, body parts, animals, etc. Cells in some regions of IT respond more to the shape of hands than to faces or animals. In other regions they respond more to animals than body parts or faces.

Within each region, some cells are tuned to particular instances of object, e.g. a particular animal.

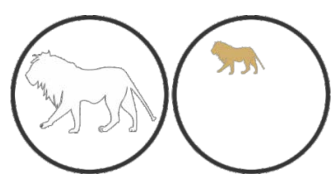
2) The response of cells is the same independent of

i) the location of the object's image on the retina because these cells have large bilateral receptive fields, ii) the size of the image, and iii) the cue that defines the objects shape (e.g. lines, color, texture, motion).

Computational processes involved in object recognition are remarkably fast. You can recognize a face in less than 100ms.



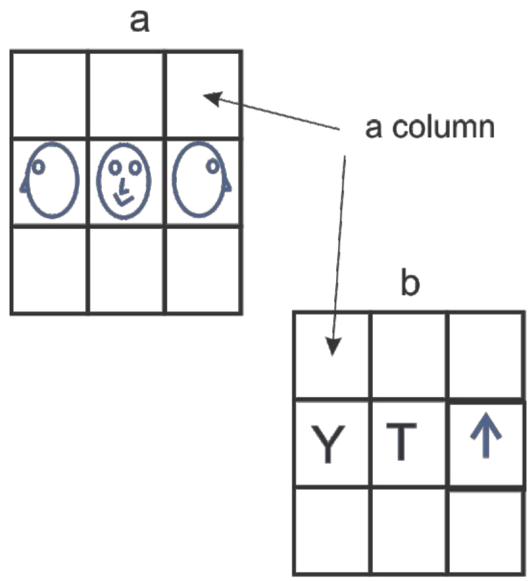
Summary: Some images look somewhat similar but represent different things. These fire many of the same cells in V1 but different cells in IT.



Other images look very different but are the same thing. These fire very different cells in V1 but the same cells in IT.

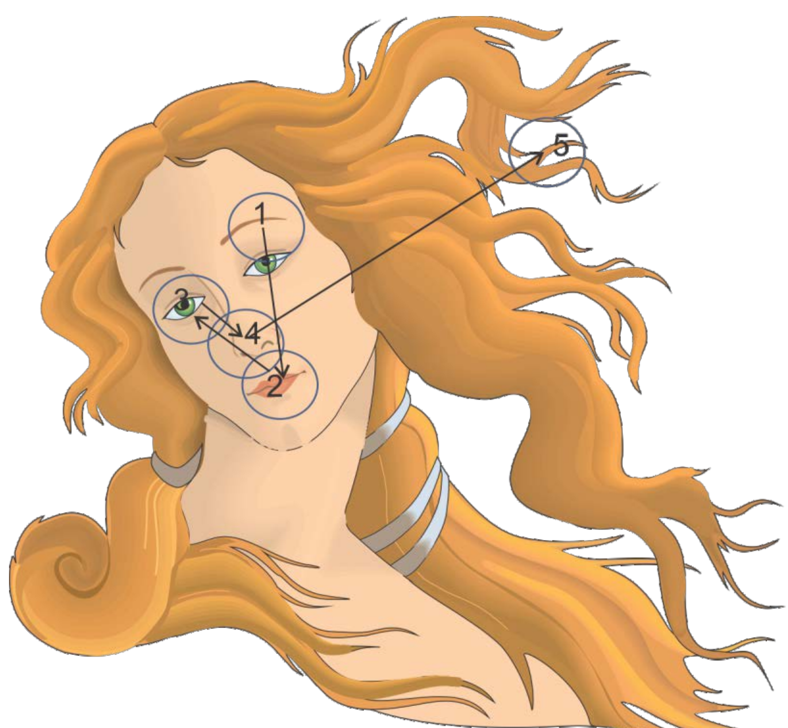
IT has a columnar organization

Cells within a column are activated by the same object. Neighbouring columns respond best to images of the same object from different viewpoints or objects of similar shapes, as in a and b.



When examining an object like a face, the eye scans it

This is because you see clearly only with the central 2 degrees of the retina, the fovea. To inspect the features of a face, you scan it with saccades. Saccades point the fovea to each important feature.



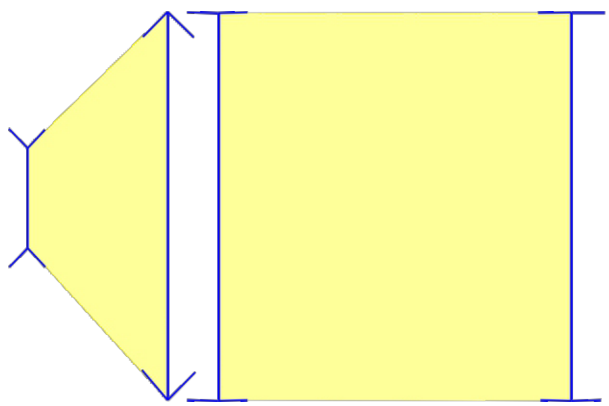
An Interesting Property of the What Stream

If you are like most, the two yellow lines will seem about the same length. But if you check with a ruler you will see that they are not.

Why is that?

This is because the “What” stream perceives objects independent of our view point. An object, like a window, is seen as a rectangle in spite of the distortions created by the viewed perspective. Because the yellow lines are captured by this perspective of the scene, their true length is distorted.

This is why drawing what we see takes so much practice. We have to learn to draw what is viewed, not what we perceive.



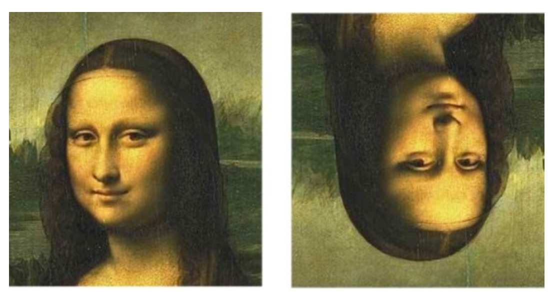
Notice anything odd?

If you turn the page so that the face is in its normal upright position, the problem becomes clear.

The features, the eyes and mouth, are not normal.

It difficult to tell this when the face is upside down. Why?

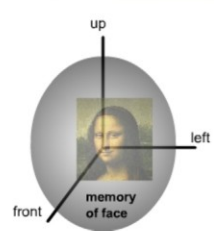
Objects, like faces, are stored in their usual orientation: an object centered representation.



When we see features like the eyes, we re-map what our eyes see into this representation.

For reasons that are not fully understood, this re-mapping fails when these features are seen in unusual orientations.

Perhaps it is simply a lack of practice.



Summary

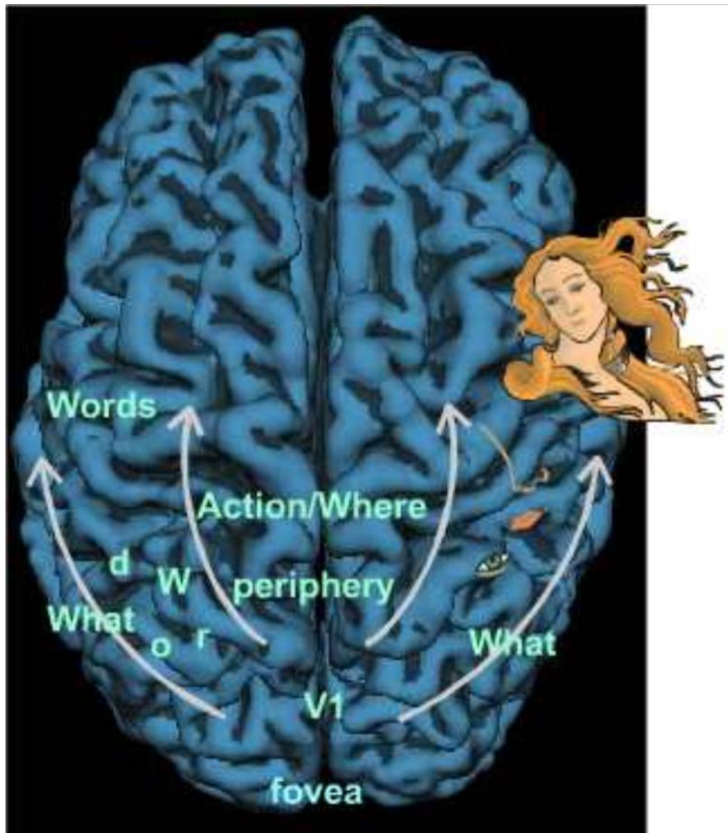
Visual information from V1 divides along two streams:

1) a dorsal "Action" or "Where" stream which is concerned with the spatial relationships between objects for the largely unconscious guidance movements.

2) a ventral "What" stream which is concerned with conscious object recognition and perception.

The "What" stream gets its input primarily from the small ganglion cells in the fovea.

The Action/Where stream gets much of its input from large ganglion cells in the peripheral retina.



In a later session we will see that the two sides are not equal.

The left side usually specializes in language, i.e., the recognition of words and sentences

The right side usually specializes in objects with spatially organized features, e.g., faces.

Perhaps this explains why it is sometimes difficult to associate a name with a face.

See problems and answers posted on

http://http://www.tutis.ca/Senses/L3VisualObjects/L3VisualObjectsProblem.swf