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

www.tutis.ca/Senses/

Contents

Objectives

1) Select the key difference in the way the retina is mapped in each of 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.

What is after V1?

From V1 (primary visual cortex) 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 diamond is seen in the right lower visual field and is represented in the left upper V1. It is represented 3 times, once in each area V1, V2, and V3.

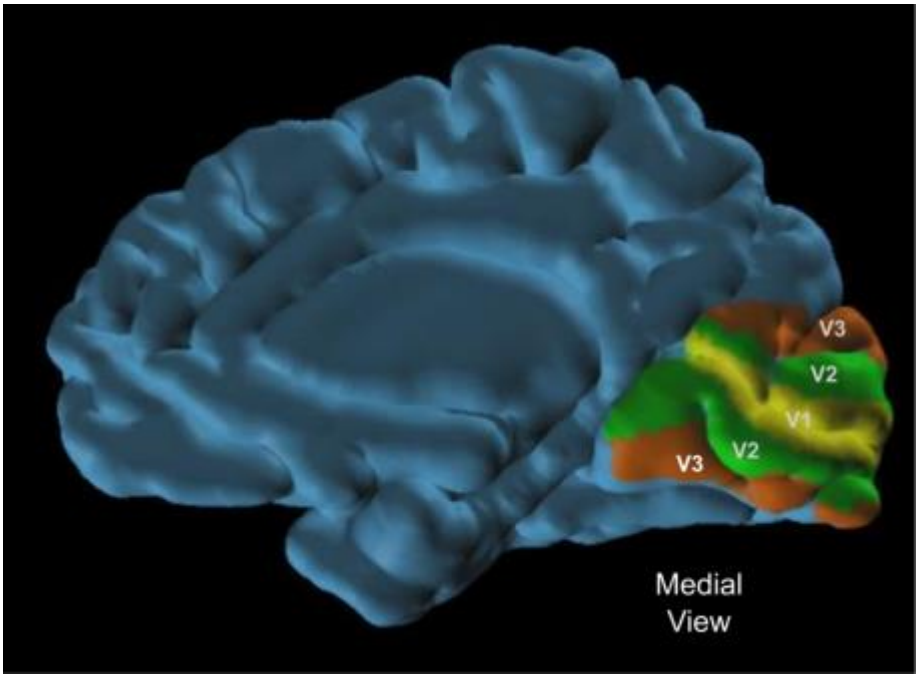


Figure 3.1 Higher Order Visual Areas V1, V2, and V3

The medial view of the right cortex with the calcarine sulcus (yellow) represented at the posterior end.

Notice the odd way the line in Figure 3.2A is represented in V1, V2, and V3. The line just below horizontal is mapped first just above the calcarine sulcus in V1 and again along both sides of the border where V2 and V3 meet. Also, a near vertical line (Figure 3.2B) is represented at the V1/V2 border and again on the far side of the V3 border. The borders formed by these locations of horizontal and vertical lines are used as landmarks to define the locations of V1, V2, and V3 in human subjects.

This odd pattern of lines occurs because the representation of the arrow (shown in the lower visual field in Figure 3.2 C) 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 arrowheads are mapped near each other (on the V1/V2 border), as are their tails (on the V2/V3 border).

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 (as in the length of the black arrows in Figure 3.3). It has been suggested that during the brain's development these axons act as springs pulling areas that are heavily connected together and thus forming the cortical folds between V1 and V2 as well as elsewhere in the cortex.

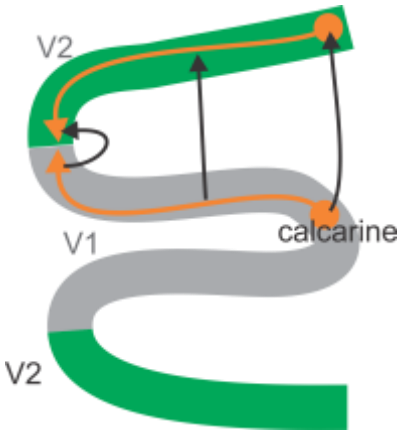


Figure 3. 3 A Possible Reason for Mirroring
The head of the arrow in V1 connects the head of the mirrored arrow in V2. The same holds for the tail and every point in between. The length of these axons is less than if the arrow was not mirrored.

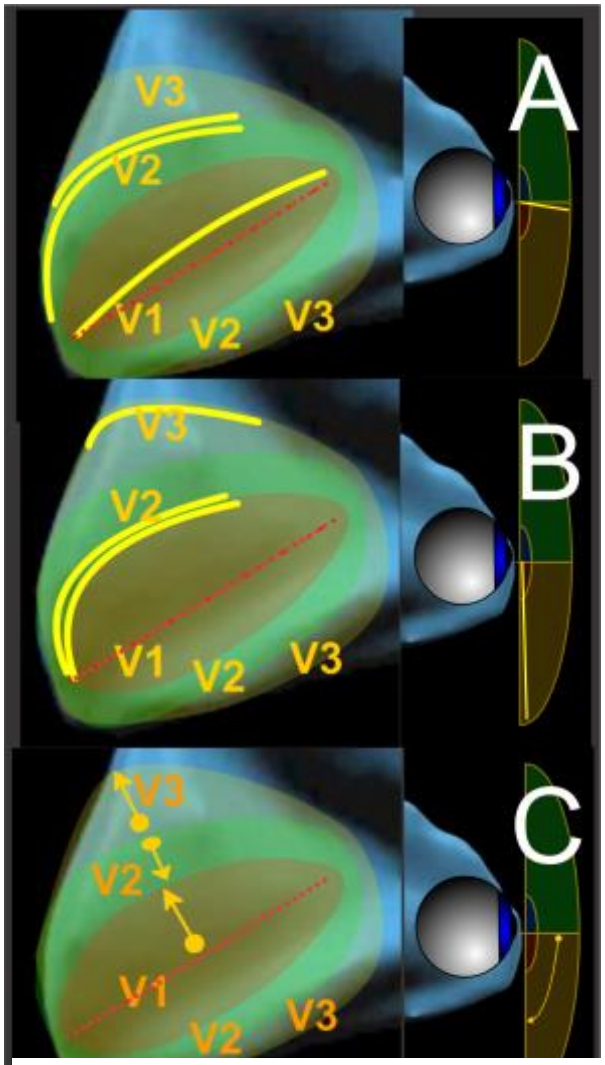


Figure 3. 2 The Projection of Lines in the Bottom Right Quadrant of the Visual Field to V1, V2, V3 in the Left Visual Cortex above the Calcarine Sulcus (red doted line)

A: The projection of a line on the left horizontal meridian.
B: The projection of a line on the lower vertical meridian.
C: The projection of an arrow whose tail starts on the right horizontal meridian and ends on the lower vertical meridian.

Assembling Simple Features into Objects

In V2 and V3 the visual system starts to assemble objects from the lines and edges extracted in V1. For example, let us look at how 4 lines of different orientations can be integrated into the contours of the 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 are grouped with the box and begin to fire synchronously (i.e. they fire at the same time).

3) Thus V1 cortex first "sees" 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 and therefore larger activity.

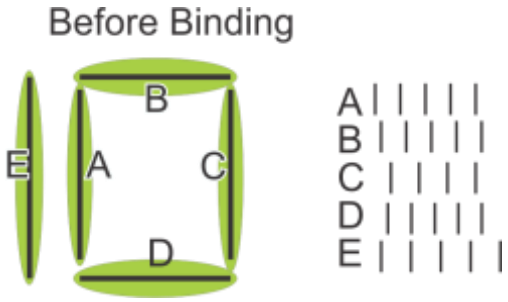


Figure 3. 4 **Before binding, the neurons representing the 5 lines A, B, C, D and E, fire at different times.**

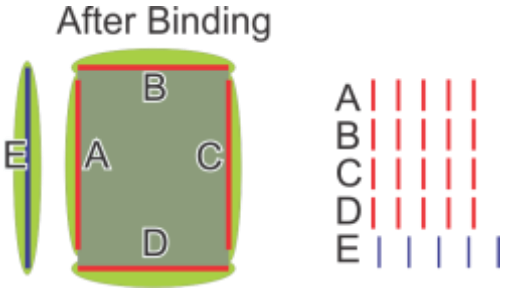


Figure 3. 5 **After binding the square shown on the left, the action potentials from neuron in the square's receptive field, shown on the right, occur at the same time.**

Synchronization Explained
Youtube Video

This video begins with 5 metronomes which are clicking asynchronously. The metronomes are placed on wheels. Now the motion of each causes a slight motion of the others in the opposite direction. Gradually these small motions of individual metronomes synchronizes the motion of all the metronomes.

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 in Figure 3.6.

The visual system 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 (Figure 3.7).

Figure 3. 6 Can you separate the white shapes belonging to the giraffes from those belonging to the background? Can you see two giraffes with their necks, bodies, and legs?

Past experience is another important for what features to bind. We see two giraffes because we remember what giraffes look like. Also once you make out where the giraffes are 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, V2, and V3 have extensive reciprocal interconnections.

Figure 3. 7 Color helps bind the shapes belonging to the giraffes and separating them from those of the background.

Illusory Contours

Note that one can see a blue square (Figure 3.8) even when there is no real square, just an illusory contour. The visual system fills in a line from the pie shaped 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 use this type of cell to bind lines separated by gaps and assemble these lines into objects such as this square.

You may also see a faint blue-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.

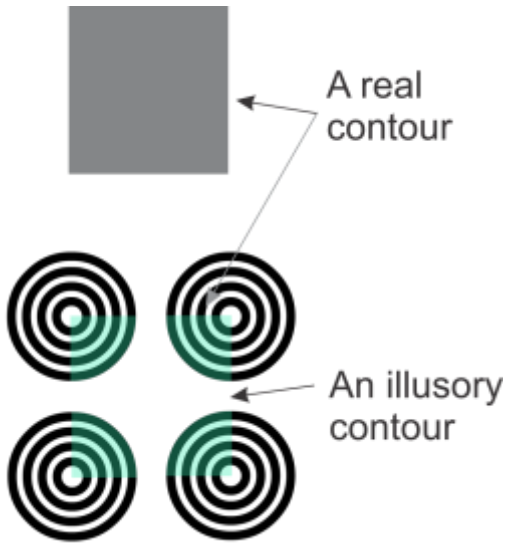


Figure 3. 8 Real contours are those depicting the grey square and blue corners of the circles. Illusory contours seem to extend from the blue corners and produce an illusion of a blue square.

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

Interpretation based on our memories modifies what we see (Figure 3.9). These top down influences (from higher to earlier areas) can be viewed as predictions. For example, if based on our memory of the word "example" we expect to see the letter m in "exanple" and we may not notice that is has been misspelled.

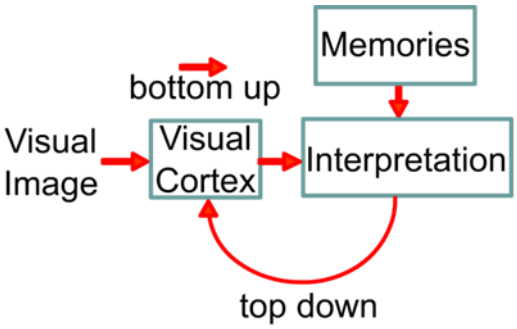
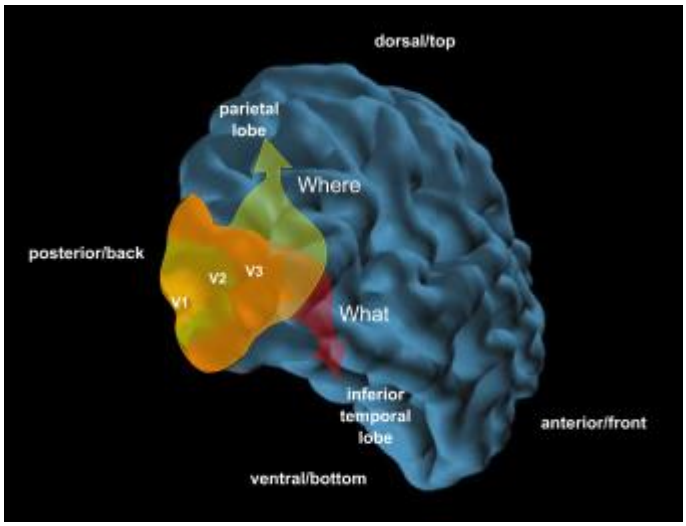


Figure 3. 9 Our visual interpretations are modified by memories which in turn produce top down tuning of the operation of neurons in visual cortex.

Another example of the predictive influence based on our memory of words is the following:

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 of these areas processes some special aspect of what we are seeing. These visual areas are like a multi-screen cinema. The main difference is that each of your brain's screens is showing a different attribute of the **same** movie: some just the motion, others the colors, etc.

Beyond V3 the two processes separate into 1) a stream that ends in the perception of edges and colors as objects, and 2) one that codes the spatial attributes of objects; for example, their location, orientation, and motion. 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),** the center for object perception, where cells respond to a particular combination of complex features, that define a particular object, for example a face.

In V2 and V3, the upper and lower visual quadrants are separated by V1. In V1 lines of the same orientation activate pinwheels of the same orientation. Thus an object centered in the visual field becomes divided into four parts in the cortex. In V2 and V3, features that share common cues, such as lines of similar orientations, are bound together. Next the **lateral occipital complex (LOC)** combines object parts seen in the contralateral visual field but not, as yet, those in the ipsilateral visual field. Finally, in regions such as the fusiform face area (FFA) located in IT, the left and right sides are brought together and the object is recognized.

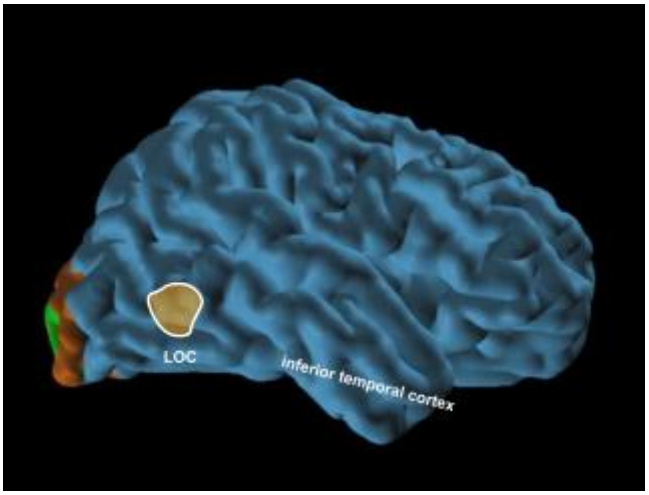


Figure 3. 11 The Lateral Occipital Complex (LOC) is seen in the back of the right cortex in front of V3.

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. A

bilateral lesion of LOC results in an inability to recognize any object including faces. Lesions in small areas of IT can result in visual agnosia of a particular class of objects (e.g. rhinoceros-agnosia).

Evidence for "What" and "Where" Pathways

In a functional imaging experiment, Leslie Ungerleider and James Haxby gave subjects two tasks.

Task 1) Press the button when the face you see now is the same face as that shown just previously. This produced activity in early visual areas and a greater activity in the 'what' stream than in the "where" stream.

Task 2) Press the button when the face you see now is in the same location as that shown just previously. This produced a greater activity in the 'where' stream. The stimuli and actions of the two tasks are identical! Remarkably, simply changing the task shifted which areas were most active.

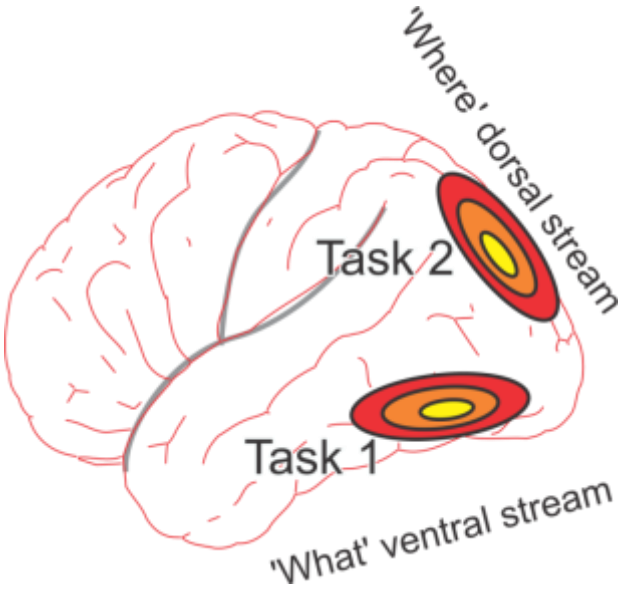


Figure 3. 12 Location of Activation to Task 1 and Task 2

If two streams exist, then one should be able to find patients with selective loss of abilities that are characteristic of each stream. This is indeed the case.

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

Small lesions in the IT produce a particular type of agnosia, for example prosopagnosia, the specific loss of face recognition.

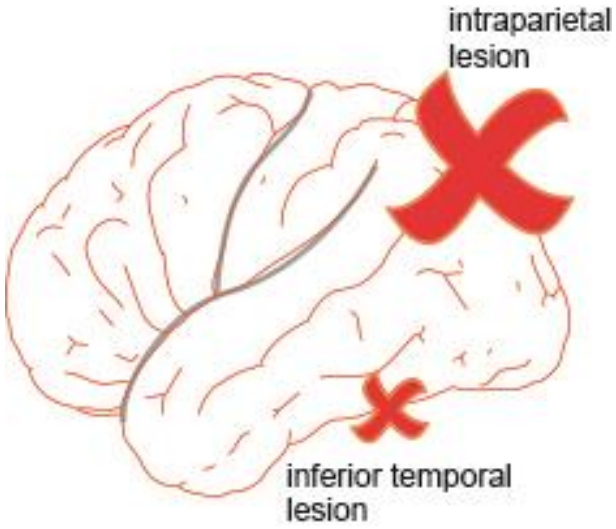


Figure 3. 13 Lesions of the "Where" and "What" Streams

The Inferior Temporal Cortex

The inferior temporal cortex (IT) stores the memories of a variety of objects e.g. animals. The studies of Nancy Kanwisher and others suggest that faces are represented by a small region located in IT called the fusiform face area (FFA). Here all neurons respond preferentially to faces and a particular face is stored by a cluster of a few highly selective neurons (sparse clustered population). A similar visual representation may hold for all objects.

Lesions of the FFA lead to prosopagnosia. Patients with prosopagnosia cannot recognize friends from visual clues, or even themselves in a mirror, 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 that a face is a face and features such as eye brows, lips, etc., but cannot recognize that a particular combination of features belongs to a particular person.

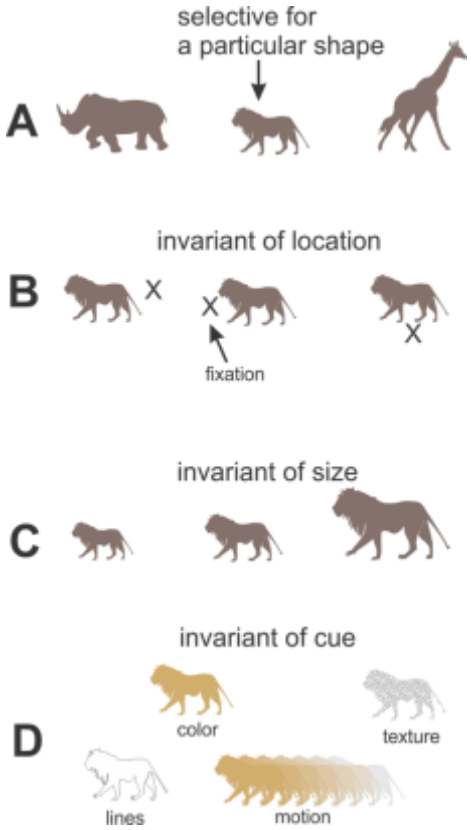


Figure 3. 14 Properties of Cells in the Inferior Temporal Cortex A: Cells in one area are selective for a particular shape. B: Their response is invariant of where the image appears on the retina. C: Or of the object size on the retina. D: Or of the cue used to display the object.

In areas of IT that include FFA:

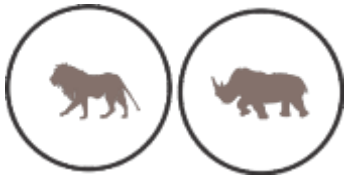
1) Cells respond selectively to a particular class of objects e.g. faces, body parts, animals, etc. Cells in some regions of IT respond more to the shape of hands than to faces or animals. In other regions they respond more to animals than body parts or faces. Within each region, some cells are tuned to particular instances of object, e.g. a particular animal.

2) Cells exhibit perceptual constancy. Their response 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 object's shape (e.g. lines, color, texture, motion).

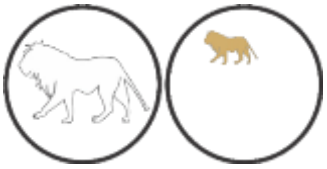
The area involved in the perception of a particular object is also involved in storing its associated visual memories.

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.



One of the most striking abilities of the ventral stream is that of identifying an object as rapidly as in 100ms. We can do this from a variety of viewpoints and images on our retinas (Figure 3. 15). "**Meaning**" is a word that is difficult to define, but it implies the ability to identify objects that are the same despite appearing from countless viewpoints.

Figure 3. 15 Photos from different viewpoints are all identified as President Obama (except perhaps those in the bottom row).

IT has a columnar organization.

Cells within a column in IT are activated by the same object. Neighbouring columns (shown from above as squares in Figure 3. 16) respond best to images of the same object from different viewpoints or objects of similar shapes, as in a and b.

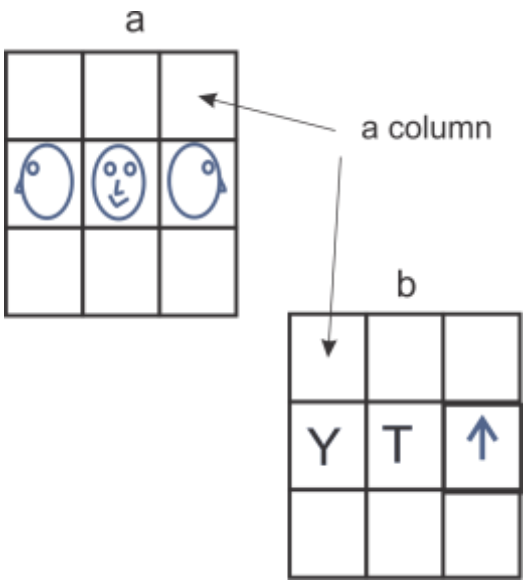


Figure 3. 16 Columns in the inferior temporal cortex. a: Three columns respond best to different views to a face. b: columns that respond to similar letter-like features.

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. Some yet unknown process reassembles these features into their correct positions when the face is recognized.

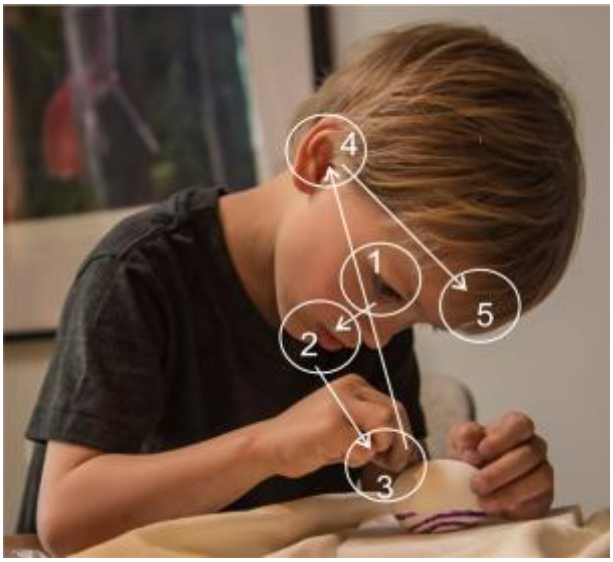


Figure 3. 17 Rapid eye movements called saccades reposition the fovea onto different features of the face.

An Interesting Property of the "What" Stream

Which of the two yellow lines in Figure 3.18 seem longer? If you check with a ruler you may be surprised that it is the one on the right.

Why is that?

Figure 3. 18 Our perception of perspective distorts our perception of the length of a line. The rightmost of the two yellow lines added to the window seems shorter. But actually it's the longer of the two.

This is because the "What" stream perceives objects independent of our viewpoint. An object, like a window, is seen as a rectangle in spite of its 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. But it is difficult to tell this when the face is upside down. Why?

Figure 3. 19 Mona Lisa seems normal But turn the picture upside down to see that she is not.

Objects, like faces, are stored in their usual orientation: an object centered representation (Figure 3.20C). 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 (Figure 3.20B) Perhaps it is simply a lack of practice.

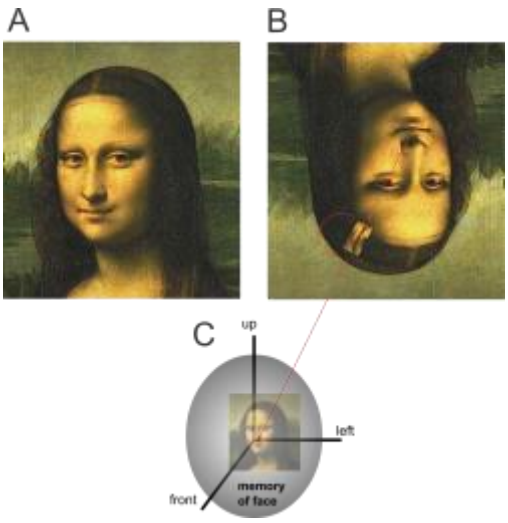


Figure 3. 20 Two Views of a Face (A or B) and its Stored Orientation C In order to compare the viewed orientation A or B to its stored orientation C, processes in the cortex must rotate the viewed features, such as the lips, from the viewed to the stored orientation.

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 and

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.

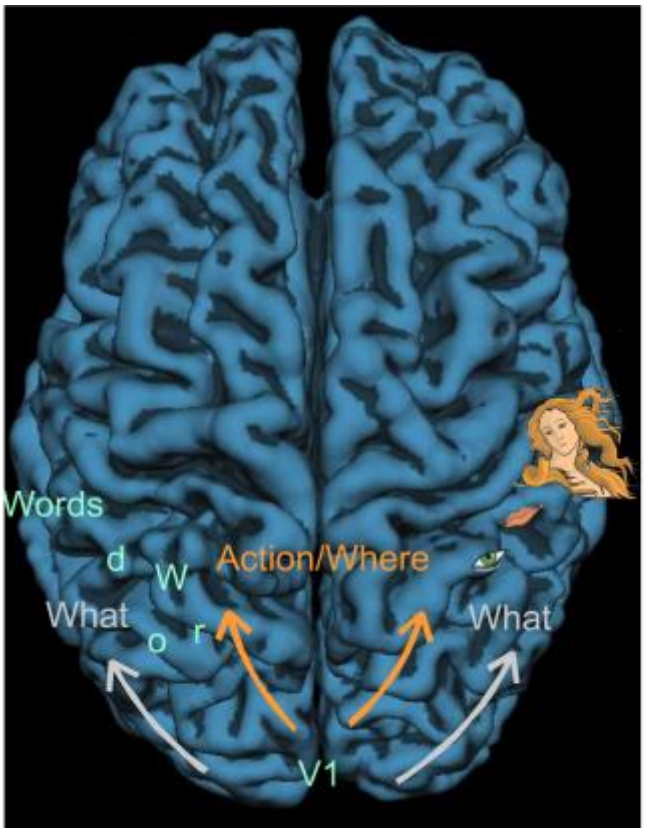


Figure 3. 21 The" where" stream projects to the parietal lobe and the "what" stream to the inferior temporal lobe. The "what" stream is asymmetric. The right side specializes in faces and the left in words.

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