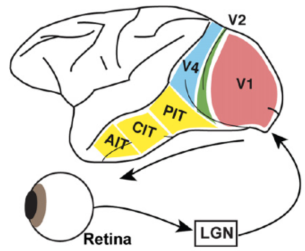
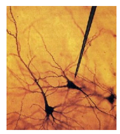
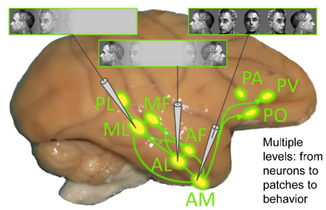
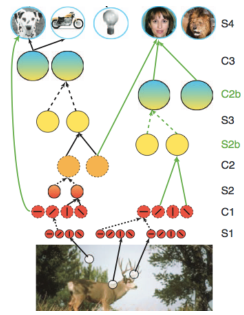
CS332 Visual Information Processing

Neural Processing in the Ventral Visual Pathway

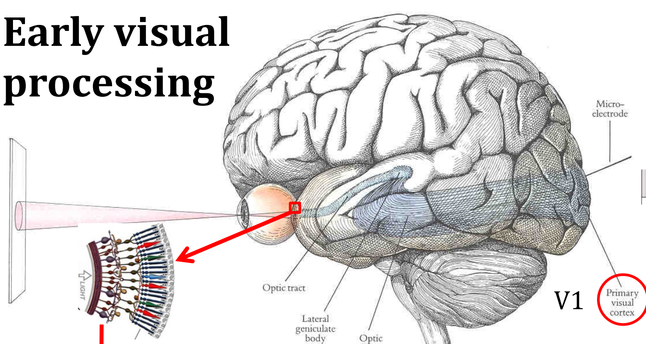


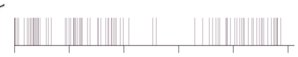




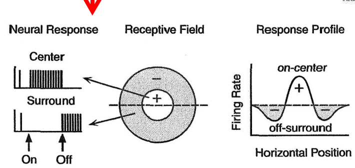


Early visual processing

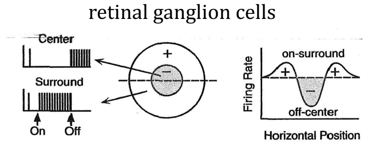


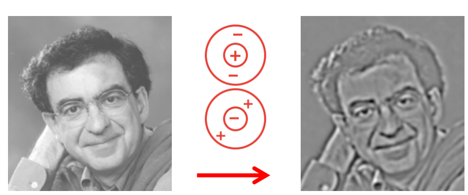


Hubel & Wiesel found 3 cell classes in V1:
- simple
- complex
- hypercomplex



retinal ganglion cells

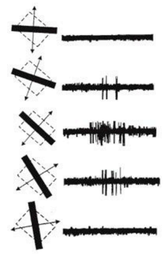




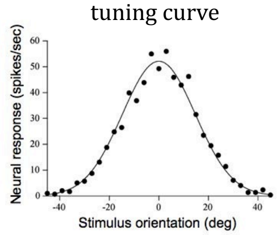
neural processing enhances intensity changes in the image projected onto the retina

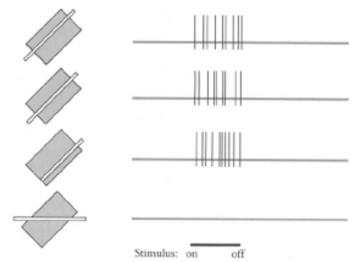
V1 simple & complex cells

simple cells respond best to edges or bars of a particular position, orientation, and sign of contrast

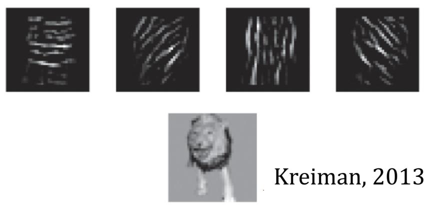


tuning curve

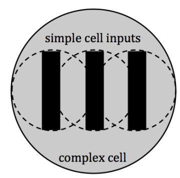




complex cells have larger receptive fields and are more tolerant to position

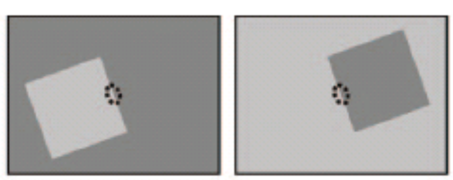


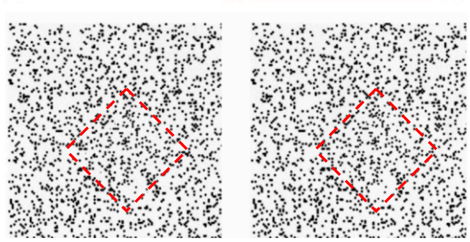
Kreiman, 2013



complex cell may "pool" inputs from many simple cells within receptive field

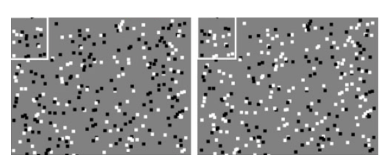
Selectivity for *stereo boundaries* in V2





Von der Heydt & colleagues:

Some V2 cells are selective for the orientation, contrast, and *side of border ownership* of an edge ... for edges defined by luminance *or stereo disparity*

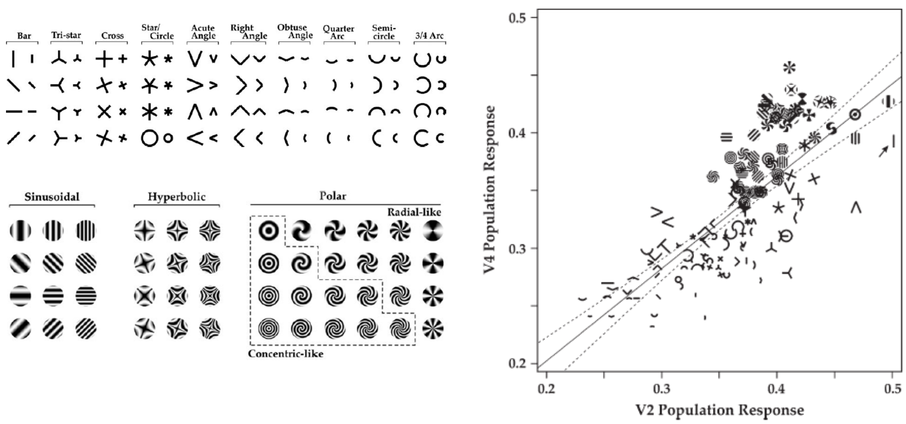


"anti-correlated" stereogram

Later, in area V4, neural responses to stereo disparity appear to correspond more closely to perceived depth

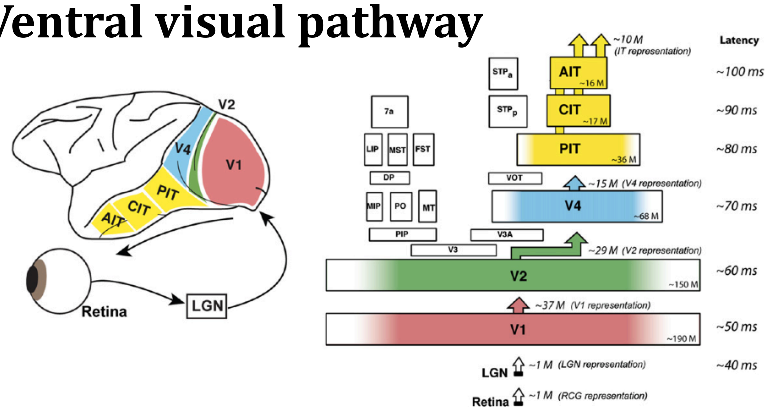
1-4

V2 and V4 responses to complex shapes



Hegde & Van Essen, 2007

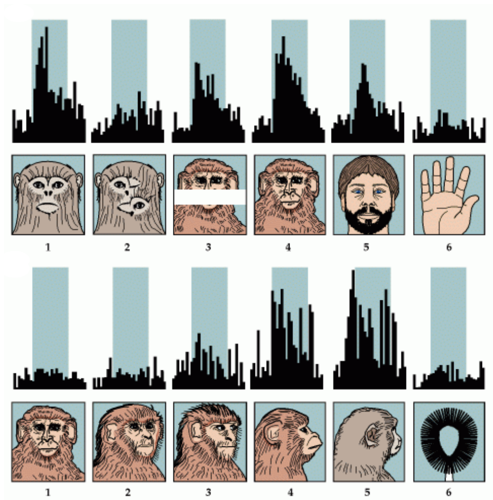
Ventral visual pathway



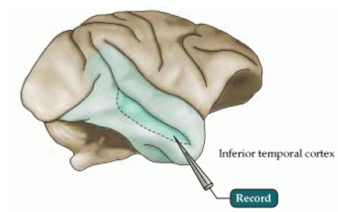
Progressing to higher areas along the ventral pathway:

- response latency increases
- receptive field size increases
- neurons become selective to more complex spatial patterns
- neural responses become more invariant to changes in position, scale, pose, etc.

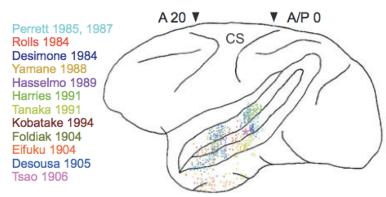
Face selective cells in IT cortex



Desimone et al., 1984

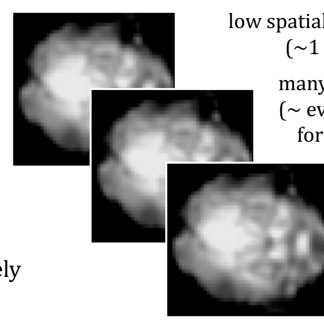


Locations of face selective cells in IT, from single cell recordings



Tsao & Livingstone, 2008

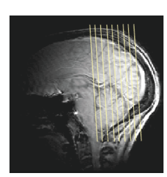
functional Magnetic Resonance Imaging (fMRI)

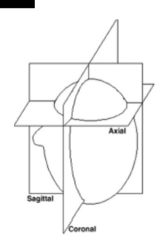


low spatial resolution
(~1 mm)

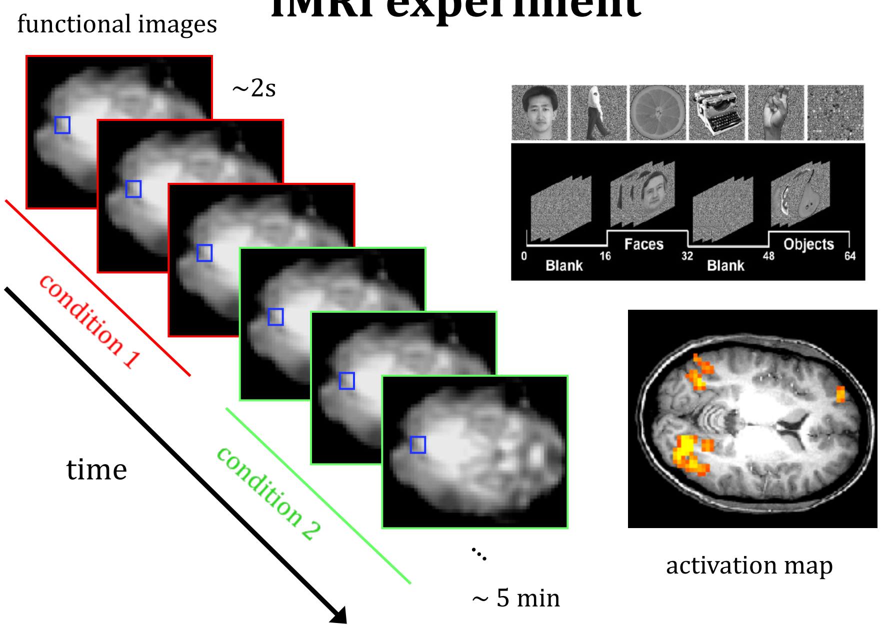
many images
(~ every 2 sec
for 5 mins)

- best spatial resolution available for measuring neural activity noninvasively in the whole human brain
- increased neural activity
  → increased local blood flow
    → change in oxygenation of hemoglobin
      → increase in MRI signal
- Blood Oxygenation Level Dependent (BOLD) signal is an indirect measure of neural activity
- raw data: ~30,000 3D "voxels"
  (each voxel: hundreds of thousands of neurons)





fMRI experiment



9

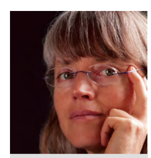
Fusiform Face Area (FFA) in human brain

The Journal of Neuroscience, June 1, 1997, 17(11):4302–4311

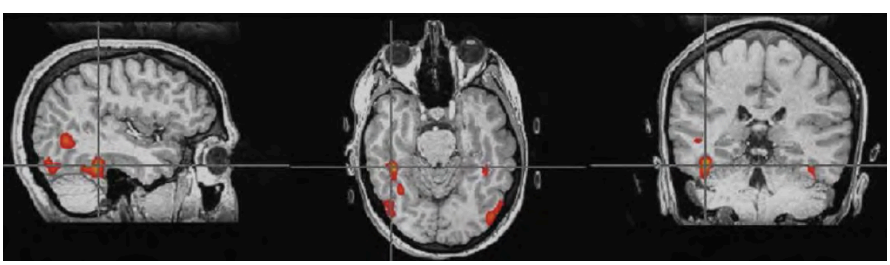
The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception

Nancy Kanwisher,[1,2] Josh McDermott,[1,2] and Marvin M. Chun[2,3]

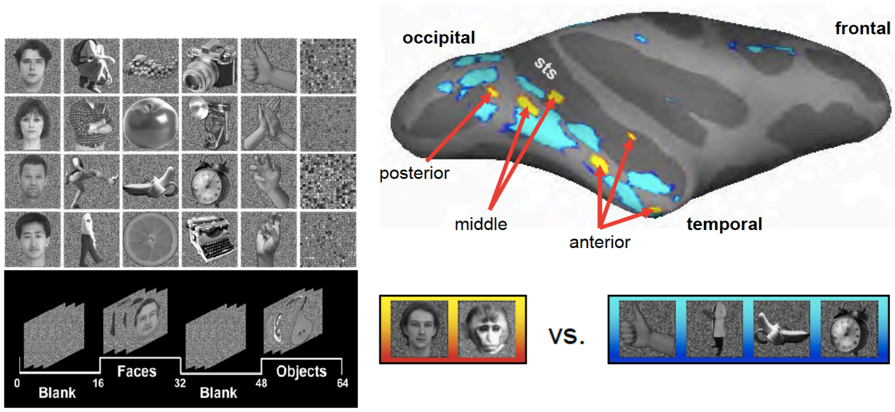
[1]Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, [2]Massachusetts General Hospital NMR Center, Charlestown, Massachusetts 02129, and [3]Department of Psychology, Yale University, New Haven, Connecticut 06520-8205



Nancy Kanwisher



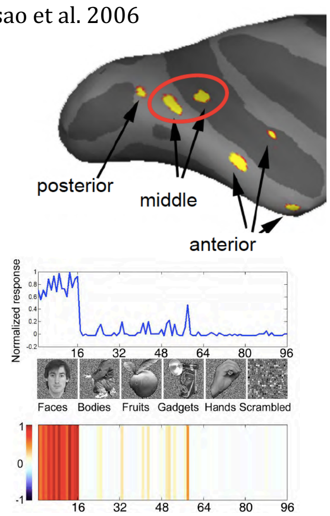
Face patches in macaque IT cortex

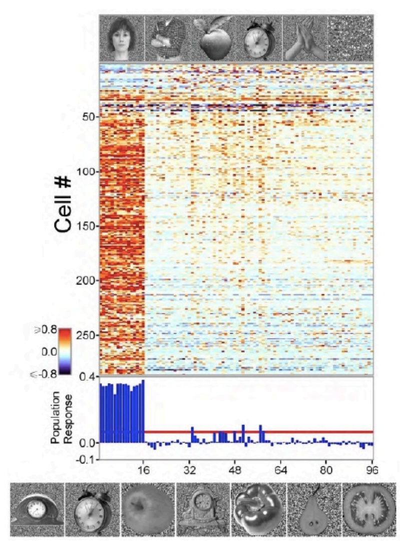


Tsao, Freiwald, Tootell, Livingstone, 2006

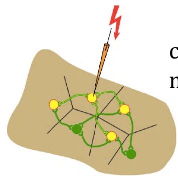
Targeting neurons in middle face patch using single cell recording

Tsao et al. 2006

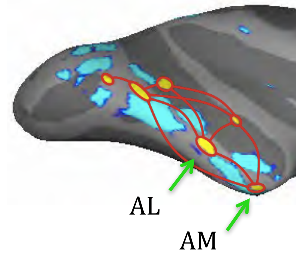




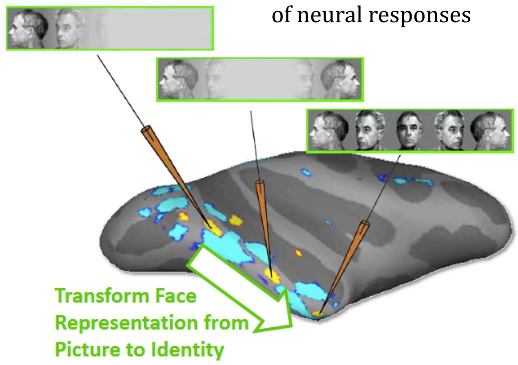
The face patch network

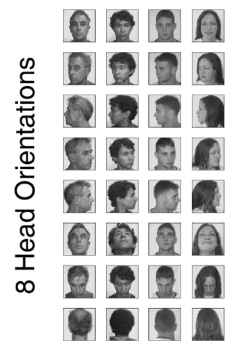


combined microstimulation & fMRI to measure connectivity of face patches



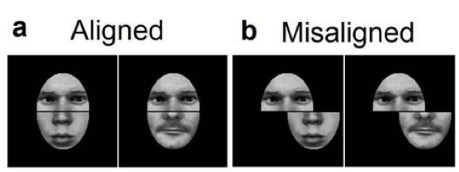
used single cell recording to probe viewpoint dependence of neural responses





Other observations…

- intact faces yield larger neural responses than scrambled or inverted faces
- composite face effect: greater response for aligned vs. misaligned faces

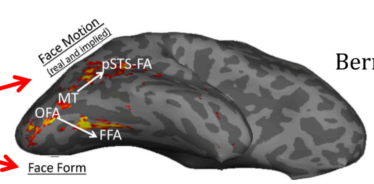


- IT neurons: response to whole face = sum of responses to parts
- some face areas show large increase in neural responses when natural face movements are added, e.g. facial expressions

human fMRI studies

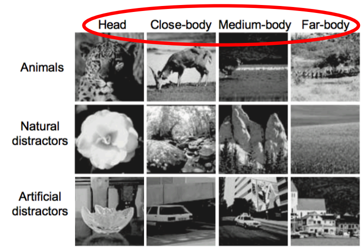
dorsal pathway

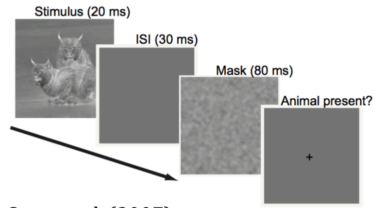
ventral pathway



Bernstein & Yovel, 2015

Rapid object detection/categorization

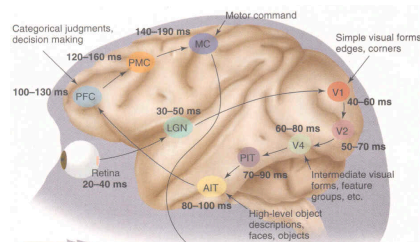




Serre et al. (2007)

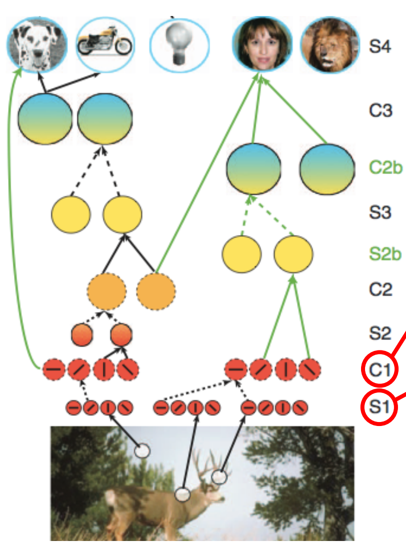
- 1,200 images, half contain animals and half are "distractors"
- respond as quickly as possible: does the image contain an animal or not?
- human subjects were ~80% correct

It takes about 100 ms for visual signals from the eye to reach the first cortical areas engaged in object/face recognition



Thorpe & Fabre-Thorpe (2001)

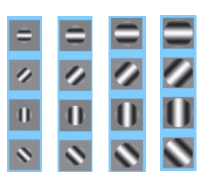
HMAX model of recognition



model of processing in ventral stream for rapid object categorization (100-150ms)

complex cells: more invariant to feature position and size

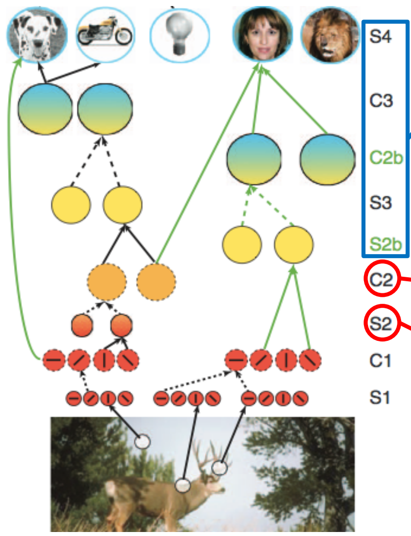
simple cells: detect edges at different positions, orientations, scales



...

early stages are "hard-wired"

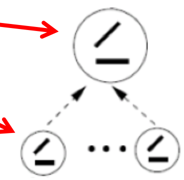
HMAX model of recognition, cont'd



C2b, C3: pool inputs with MAX

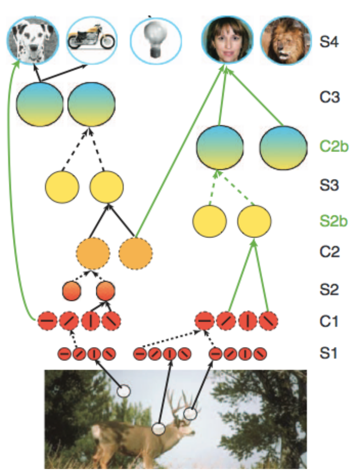
S2b, S3, S4: combine more complex features with *weights that are learned*

C2 units: same selectivity as S2 units with more tolerance to position & size (pool S2 units of same selectivity but different positions and sizes)



S2 units: combinations of C1 units at different orientations with *weights that are learned*

HMAX model of recognition, cont'd



- learning of wiring and weights for top-level object classification by *supervised learning*
- wiring and weights between C and S units at early levels are also learned

  e.g. C1 → S2, S2b   C2 → S3
- *unsupervised learning* of feature combinations that appear most often in natural images
- good match to neural responses, V1 → IT
- "neural tuning size" (number of C1 inputs to each S2 unit) accounts for holistic effects (composite-face, face-inversion, whole-part)
- good performance on natural images