Neurobiology HMS 130/230 Harvard / GSAS 78454

Visual object recognition: From computational and biological mechanisms

Lecturer: Carlos R. Ponce, M.D., Ph.D.

Postdoctoral research fellow
Margaret Livingstone Lab, Harvard Medical School
Center for Brains, Minds and Machines, MIT

crponce@gmail.com

Today's meeting: Early Steps into Inferotemporal Cortex

Agenda

Today's theme: inferotemporal cortex (**IT**), a key locus for visual object recognition

1. What is **IT**?
   - a brief review of the ventral stream and how **IT** fits in it

2. What do **IT** neurons do?
   - selectivity

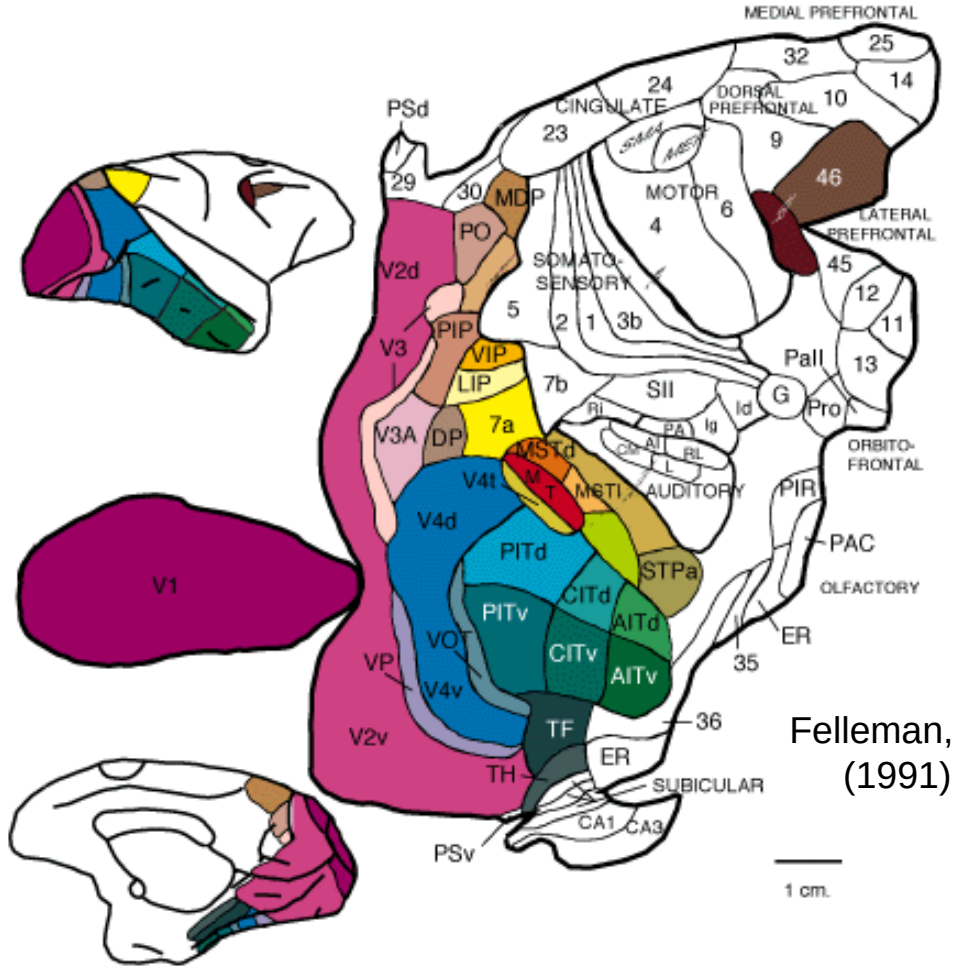
3. How well do **IT** neurons do their job?
   - the problem of invariance

4. Some unresolved questions in **IT**

5. Segue into the paper: how do we understand **IT** neurons at the population level?

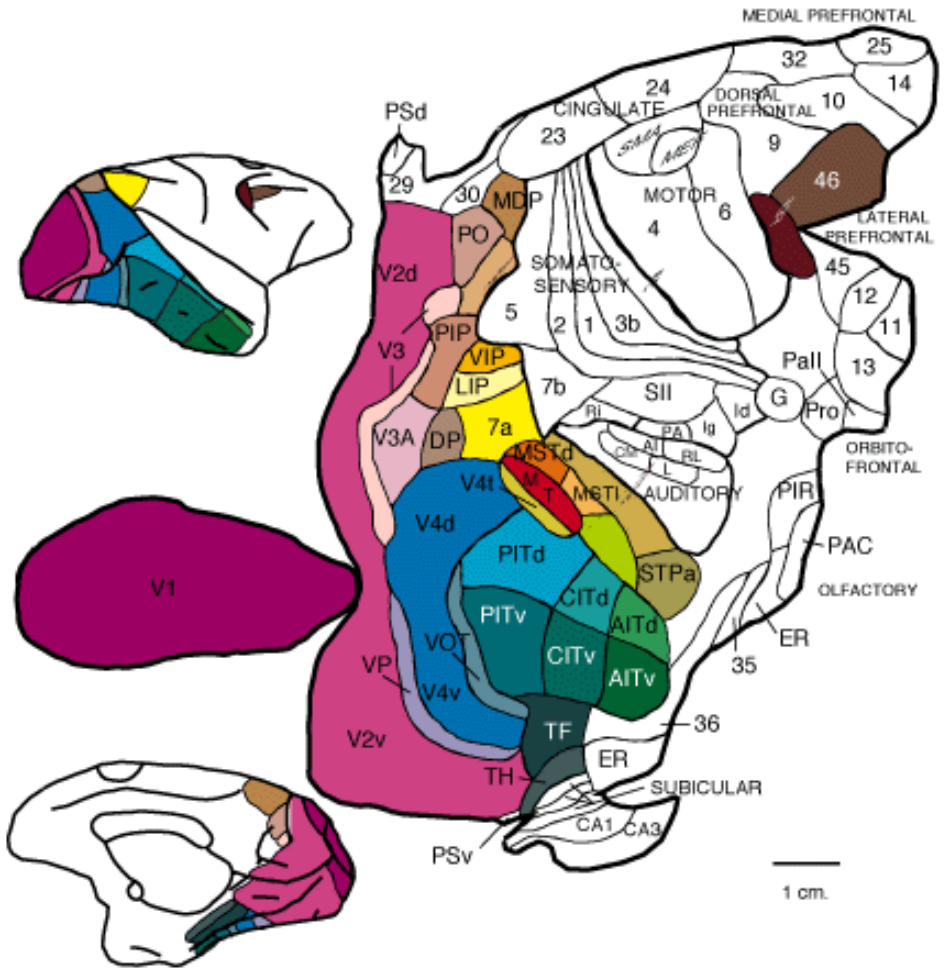
1. What is inferotemporal cortex (IT)?

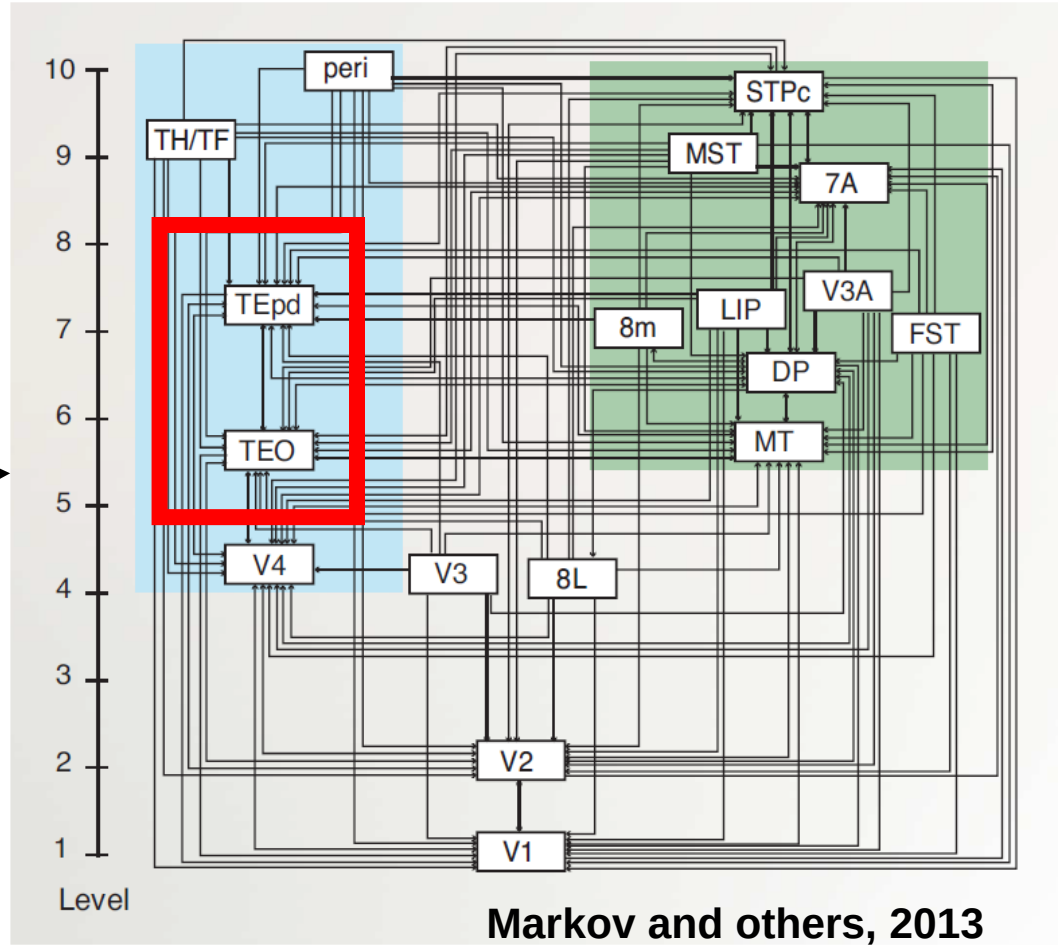
There are over 30 visual areas in the brain of the macaque



Felleman, D. J. and Van Essen, D. C. (1991) *Cerebral Cortex* **1:**1-47.

IT is the **last** exclusively visual area of the **ventral stream**, following areas V2 and V4



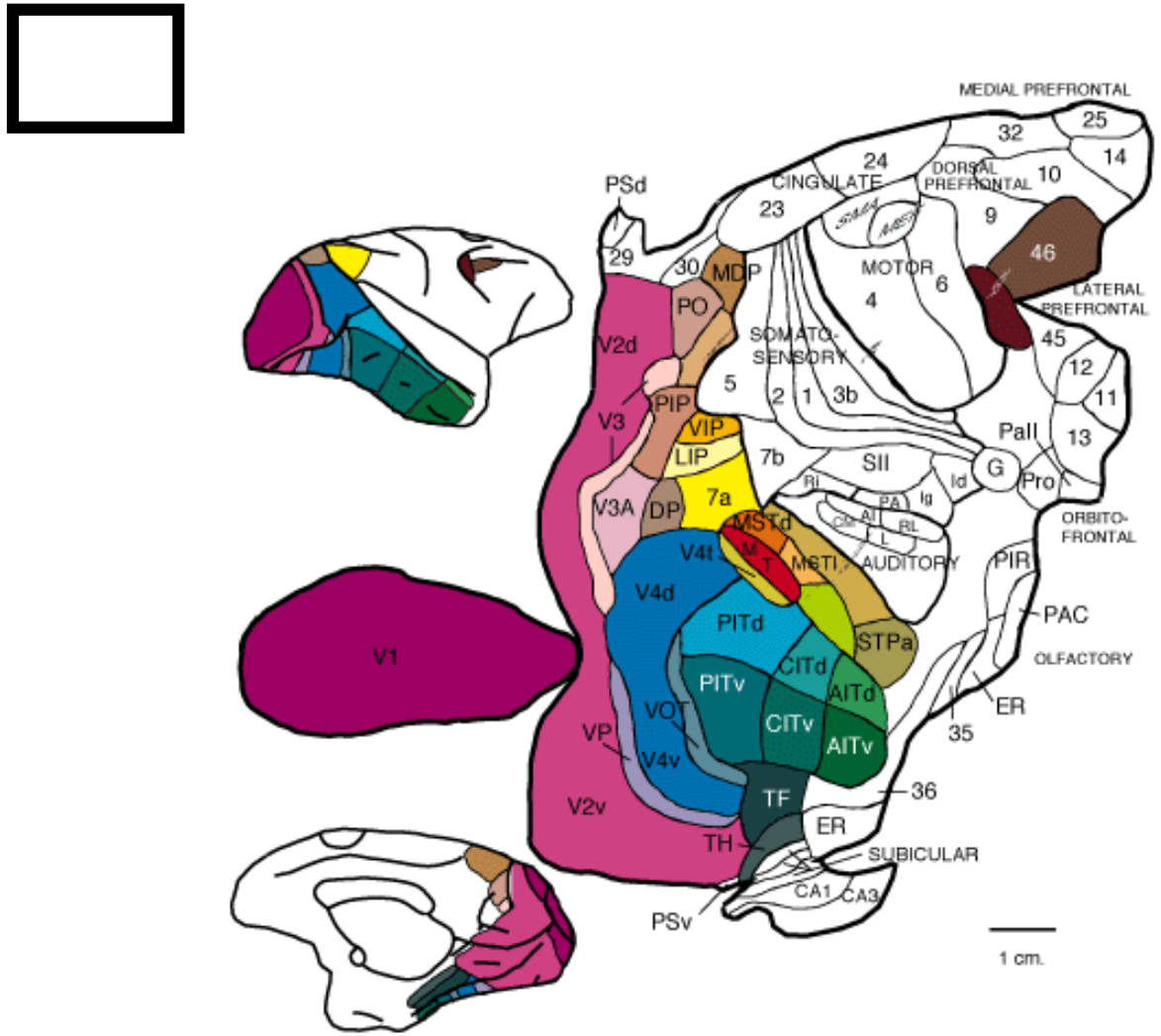


Markov and others, 2013

How do we organize these ventral stream areas into a hierarchy?

We can organize cortical areas through their laminar (layer) connection patterns

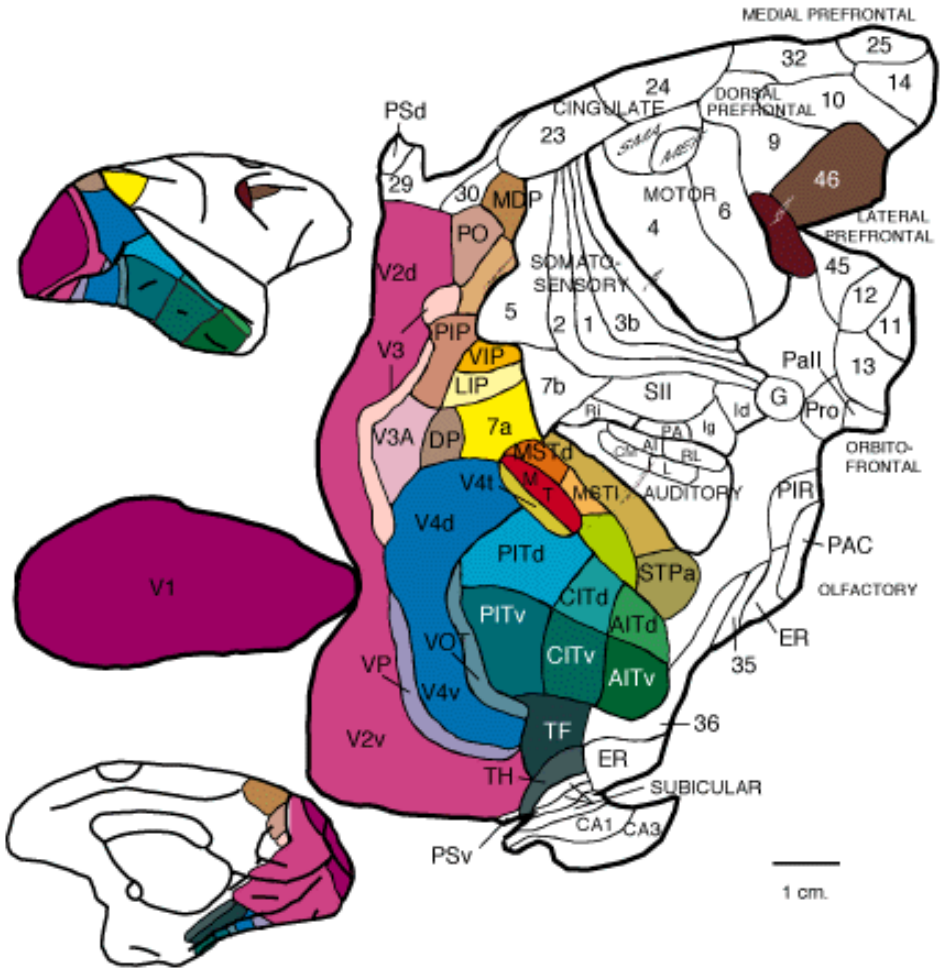
a. Select a cortical area (say, posterior IT)



We can organize cortical areas through their laminar (layer) connection patterns

a. Select a cortical area (say, posterior IT)

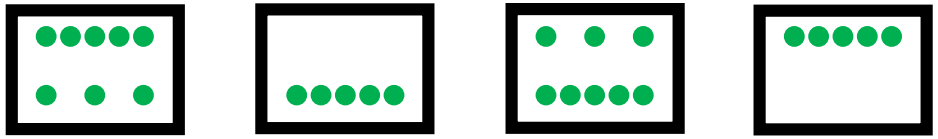
b. Inject a retrograde tracer



We can organize cortical areas through their laminar (layer) connection patterns

a. Select a cortical area (say, posterior IT)

b. Inject a retrograde tracer



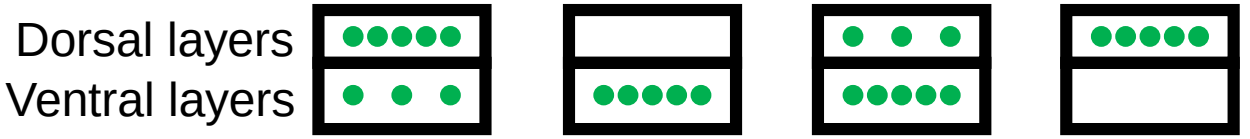
area X area Y area Z area A

Neurons in many areas take up the tracer

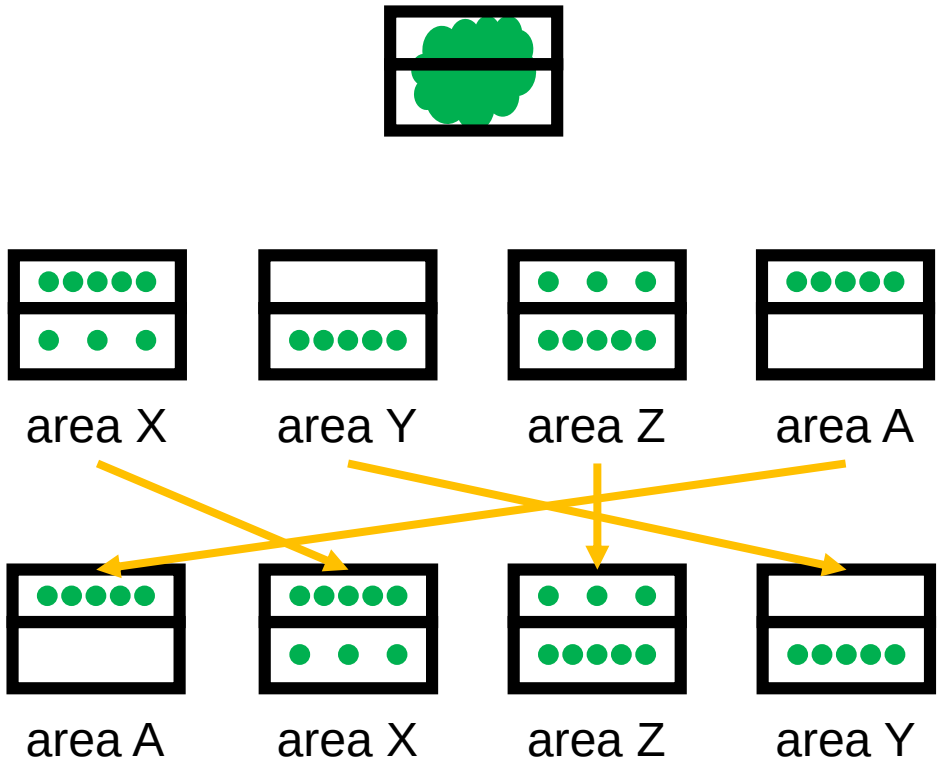
We can organize cortical areas through their laminar (layer) connection patterns

a. Select a cortical area (say, posterior IT)

b. Inject a retrograde tracer

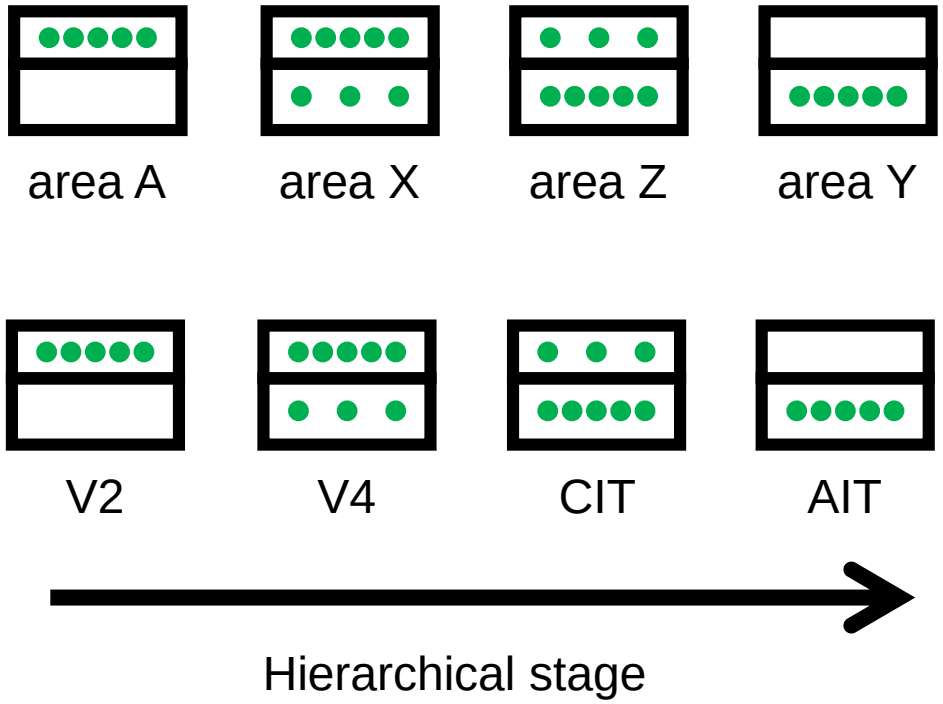


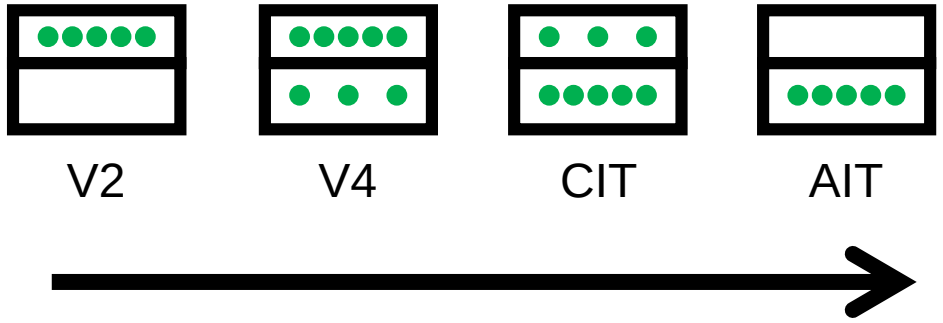
- count the number of labeled cells in the dorsal layers
- count the number of labeled cells in the ventral layers

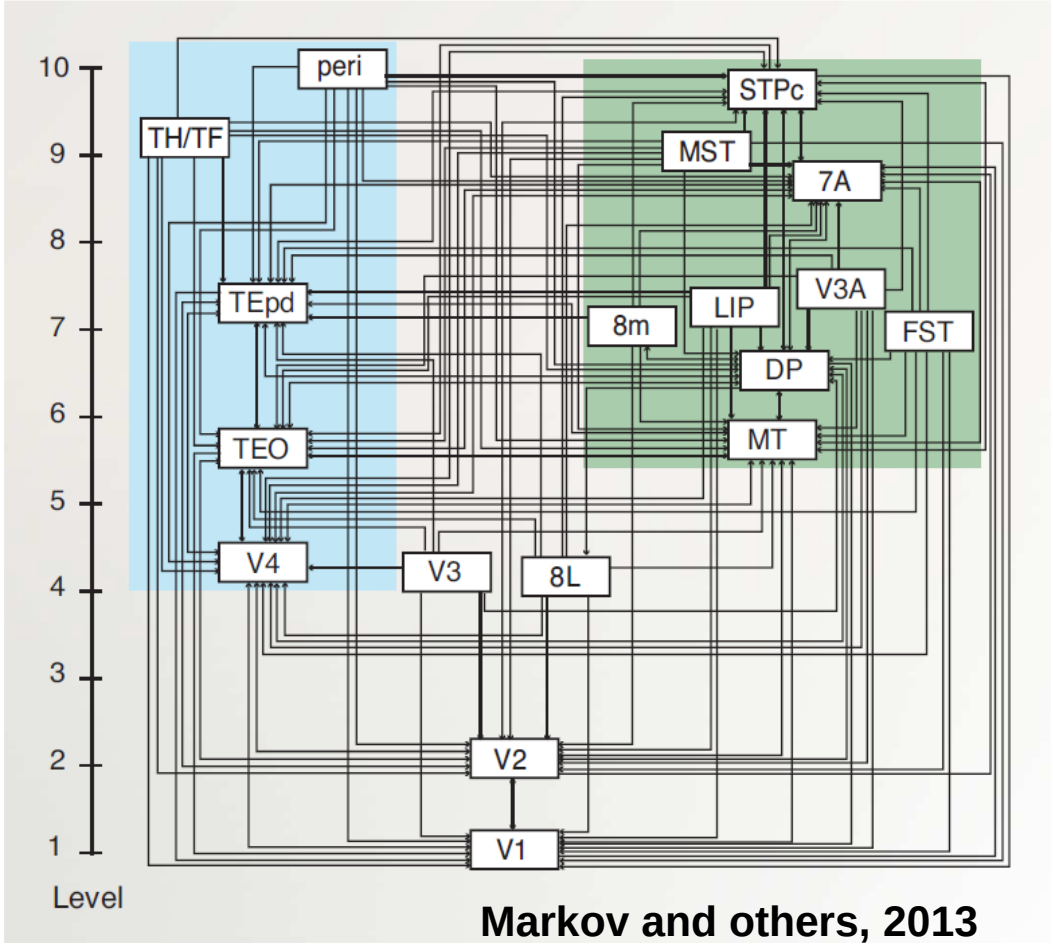


- sort areas by the ratio (# cells in dorsal layers / # cells in ventral layers)

the results in a consistent rank of cortical areas across individuals (and species)

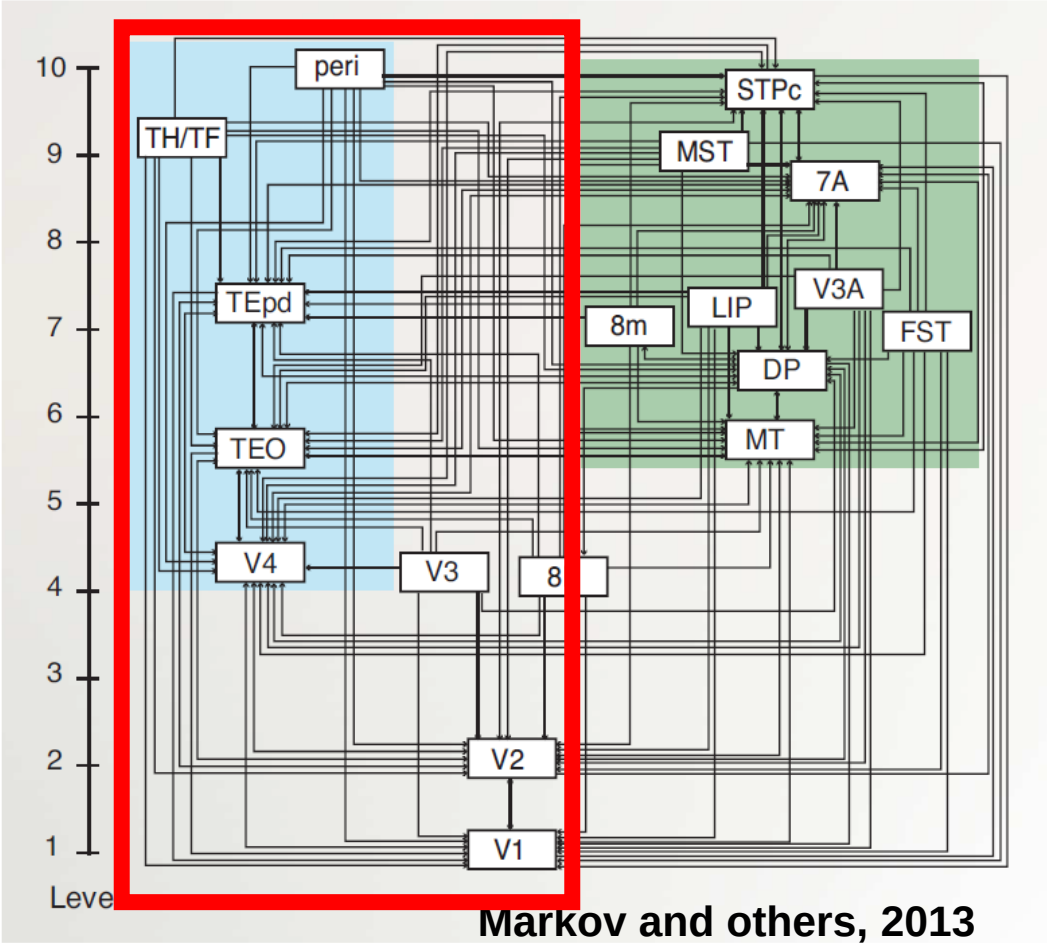






Markov and others, 2013

Historically, this hierarchy has been described as the “ ventral stream” (Ungerleider and Mishkin, 1982)



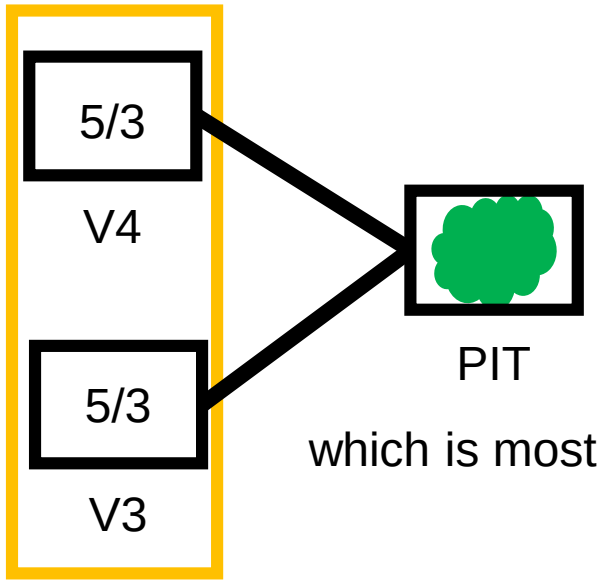
Markov and others, 2013

But if all these areas are so highly interconnected, how are they a “stream?”

IT depends on some regions more than others

how we know

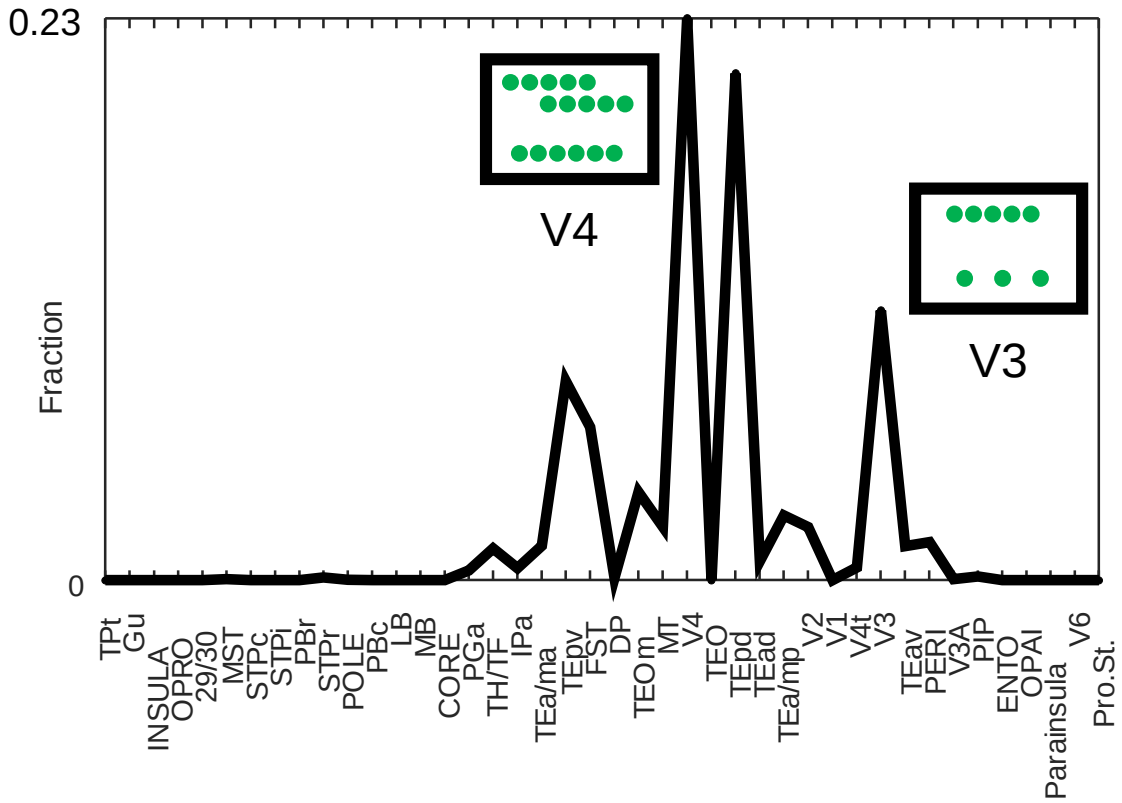
say you find two visual regions at approximately the same hierarchical level



which is most important to PIT?

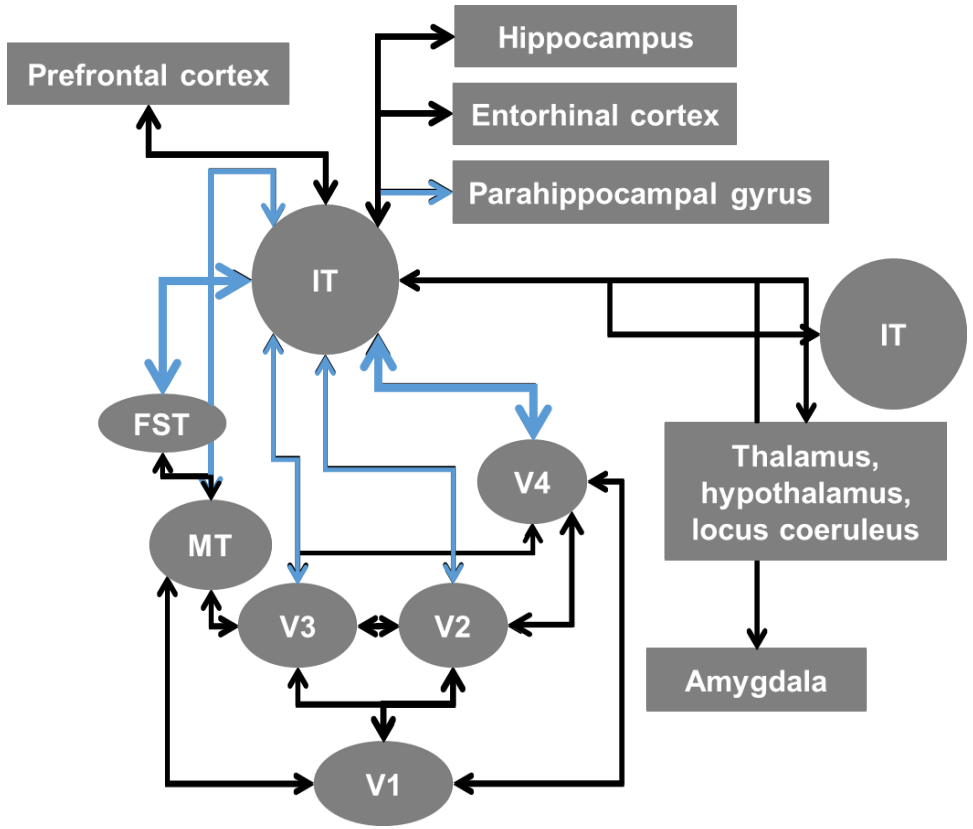
answer: count the total number of cells labeled for every injection!

Markov and others (2013) defined the relative weights from cortical area to cortical area

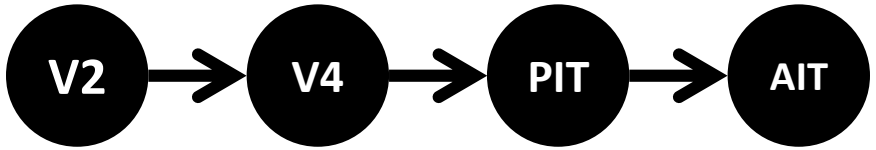


Here's one example: posterior IT

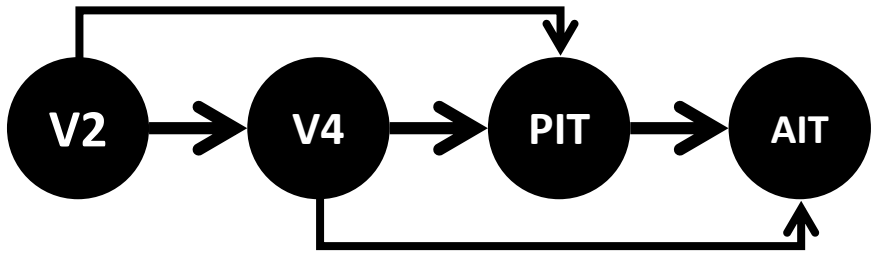
By applying weights to these connections, we can better understand the “chain of command”



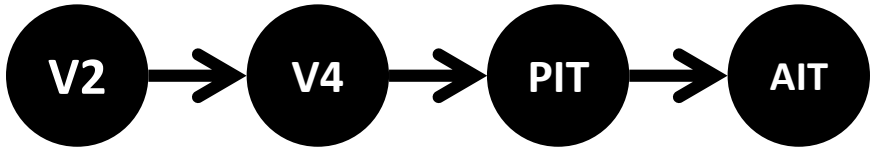
Because IT depends more on V4 than in other regions, we can think of IT as part of a “stream”



Once we get a hold of this primary pathway, we’ll bring in the rest!



depends



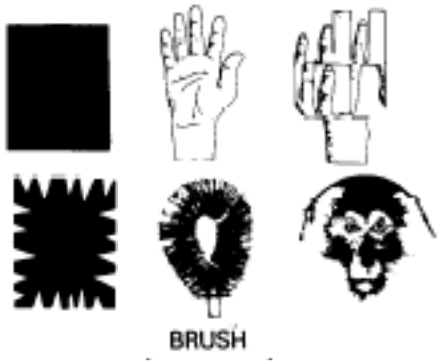
IT “depends” on V4 for what?

2. What do IT neurons do?
- selectivity in IT

IT neurons respond to (“prefer”) **complex** images

Pictures and drawings of natural images

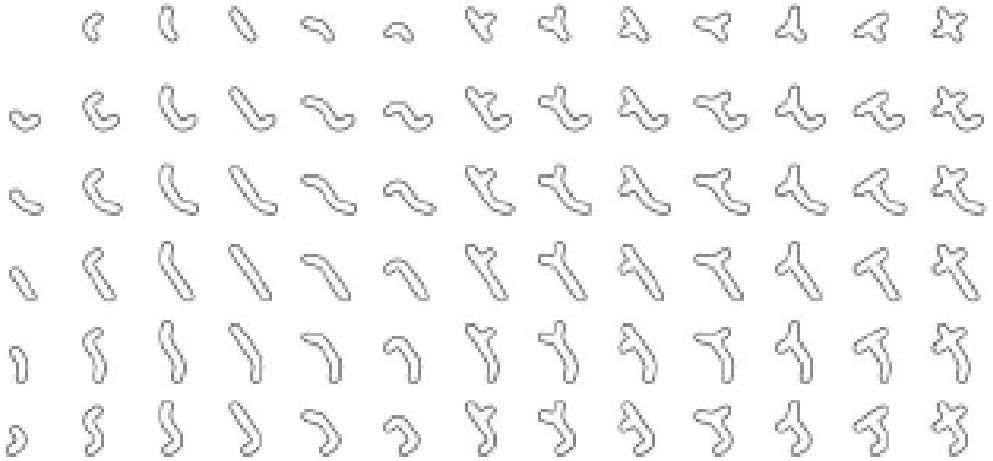
1984: Desimone, Albright, Gross and Bruce



2005 - Hung, Kreiman, Poggio and DiCarlo

2007: Kiani, Esteky, Mirpour and Tanaka

Parametrically defined objects (“curvature”)



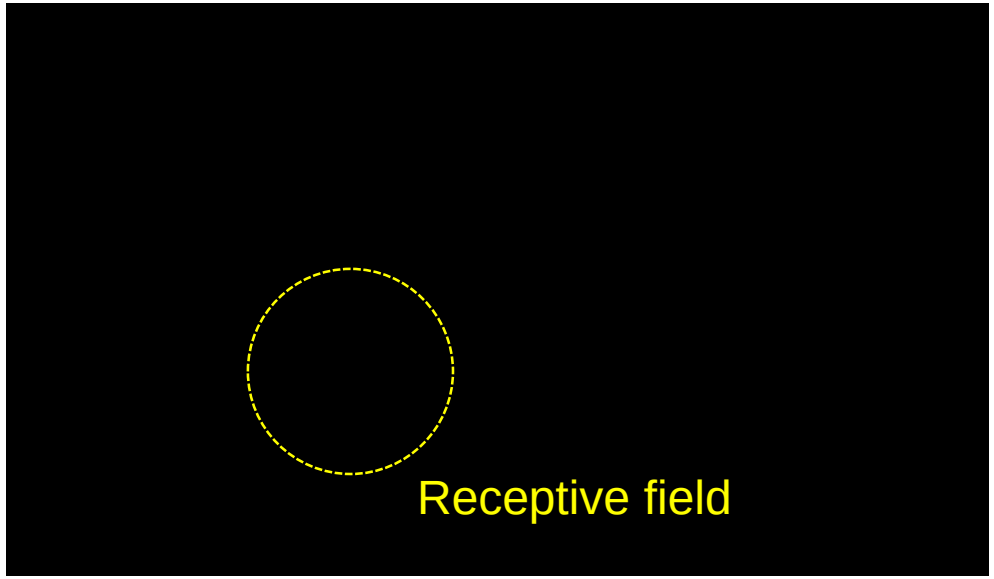
2006: Connor and others

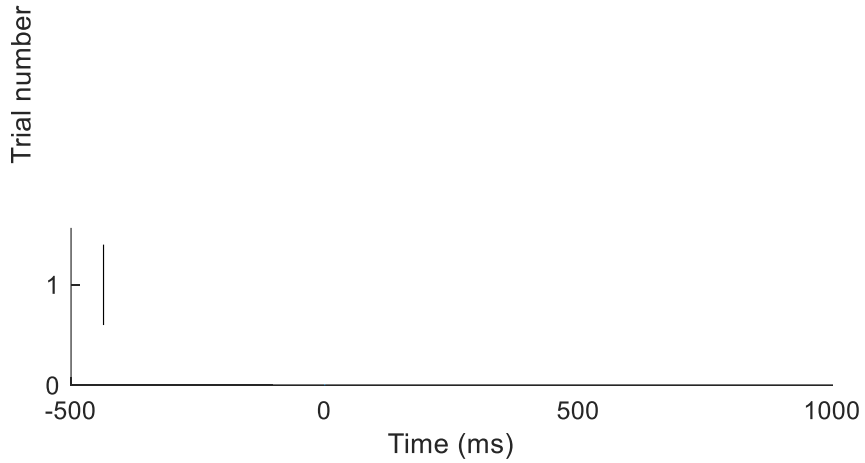
1995: Logothetis, Pauls and Poggio

How do we know what a cell “**prefers**”?

We count spikes.

Imagine we’ve identified an IT neuron’s RF

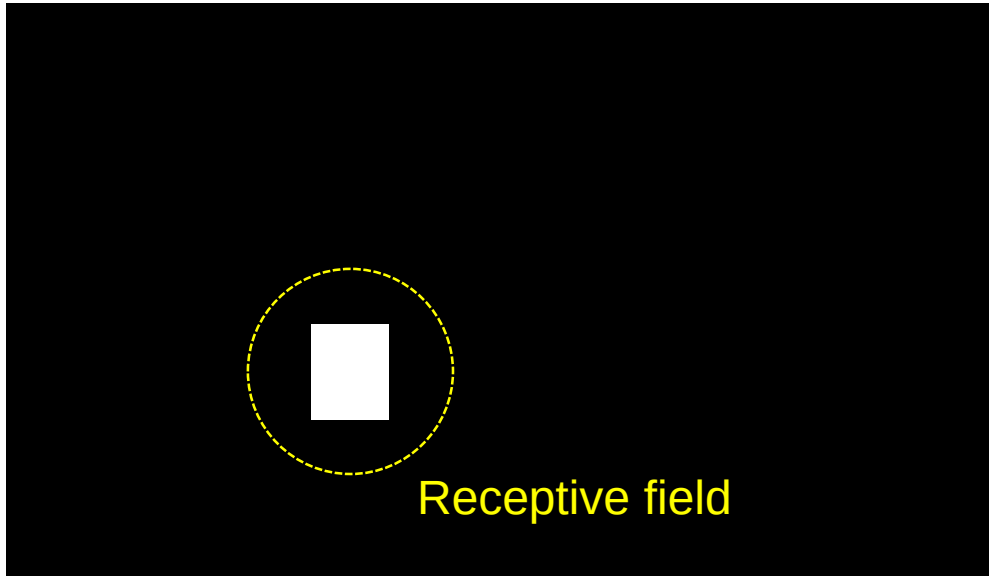


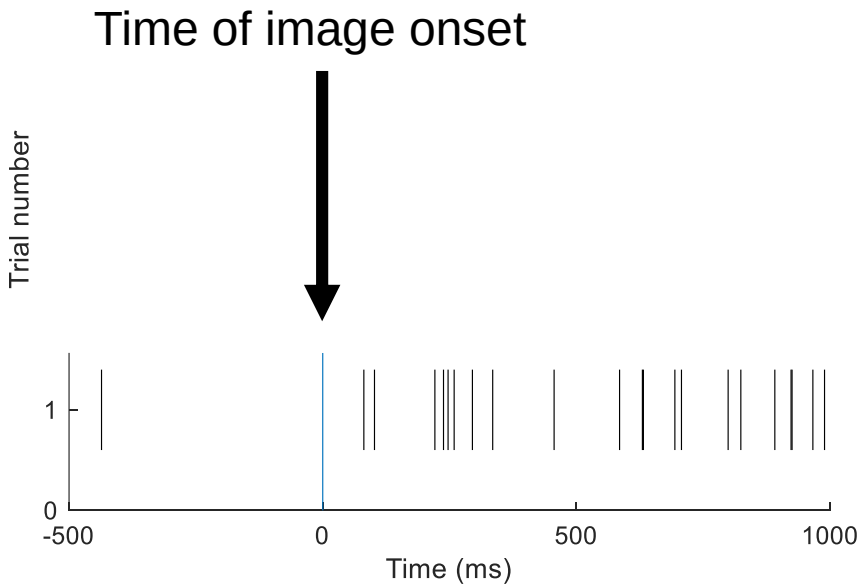


During rest, the unit may fire ~ 6 spikes per s

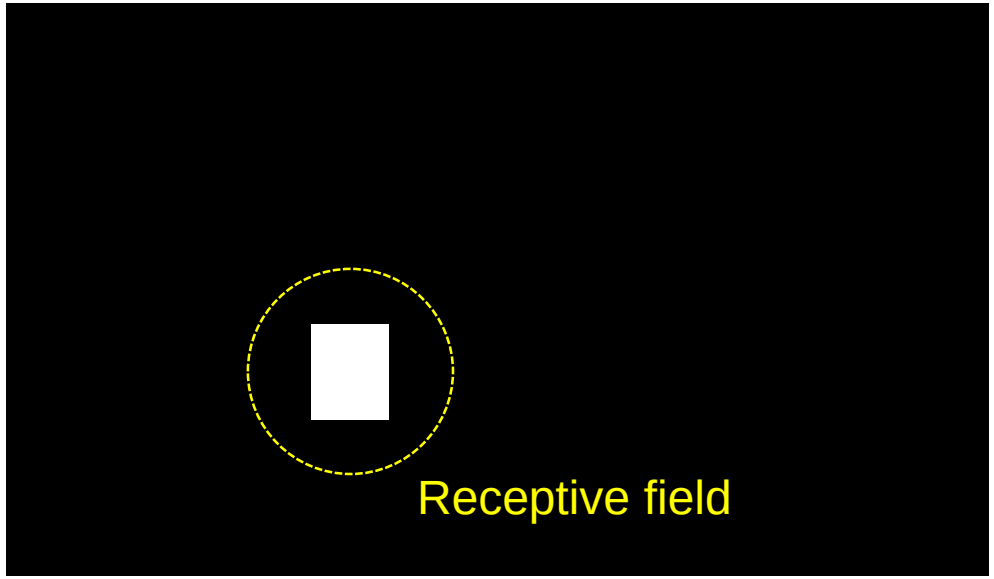
Credit: Praneeth Namburi

When we flash an image in the RF

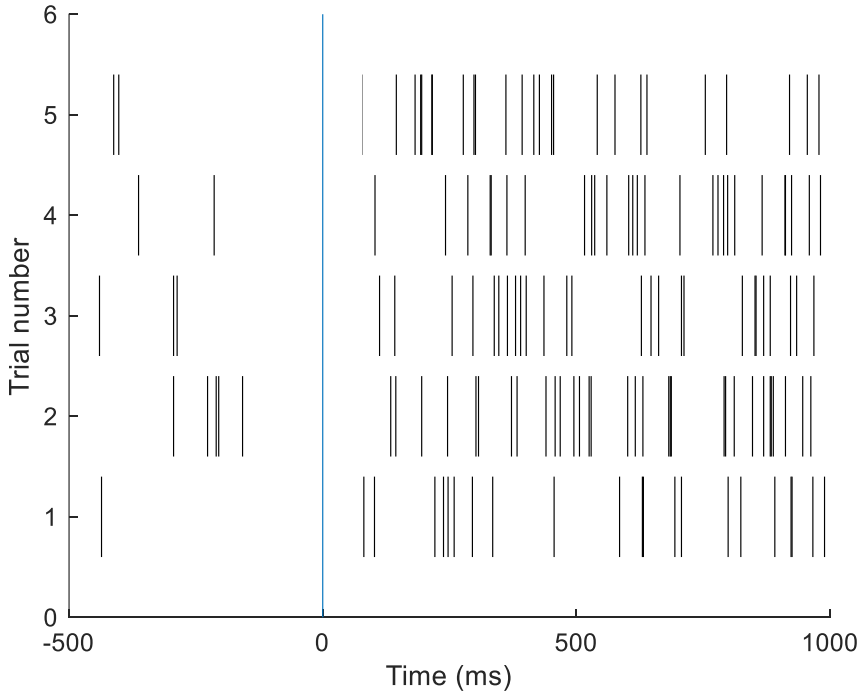




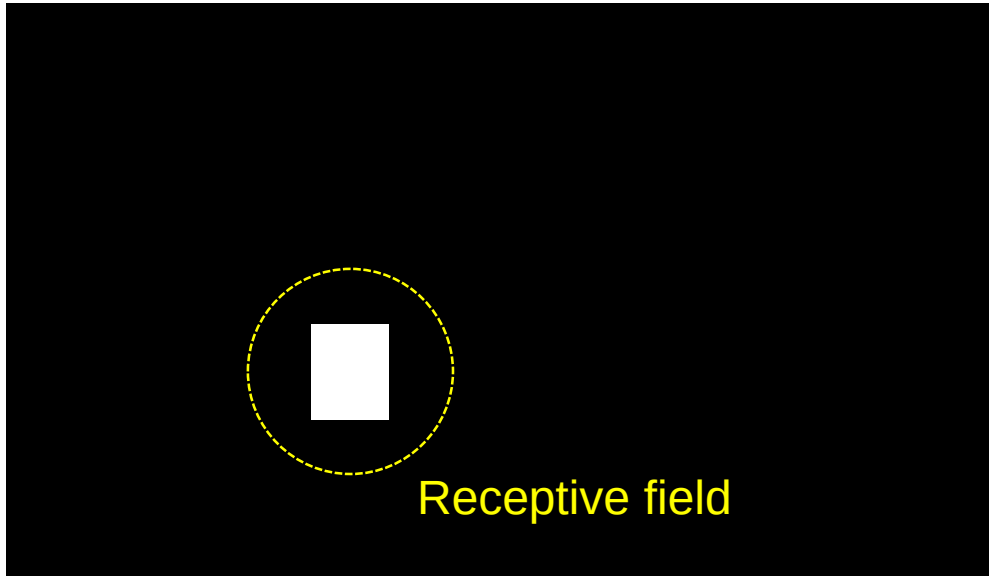
We look for changes in the spike rate

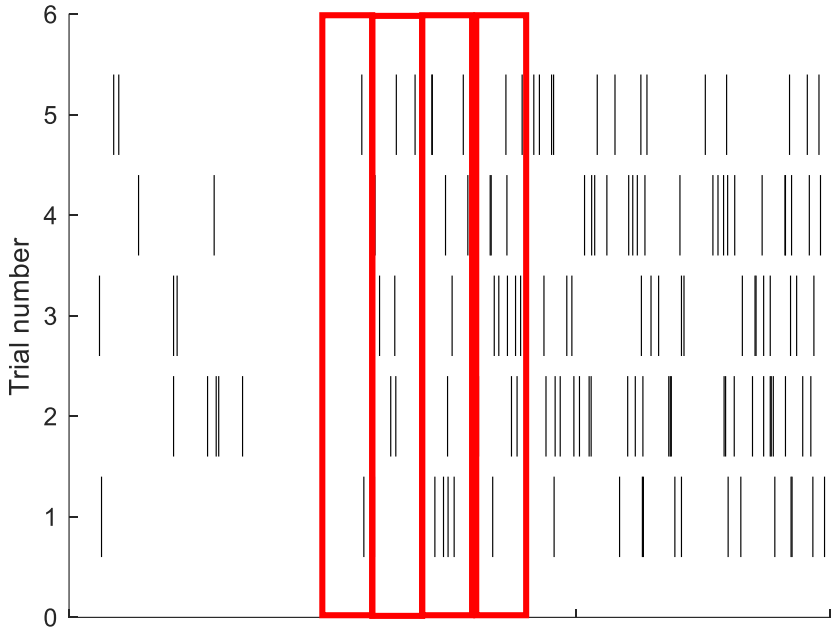


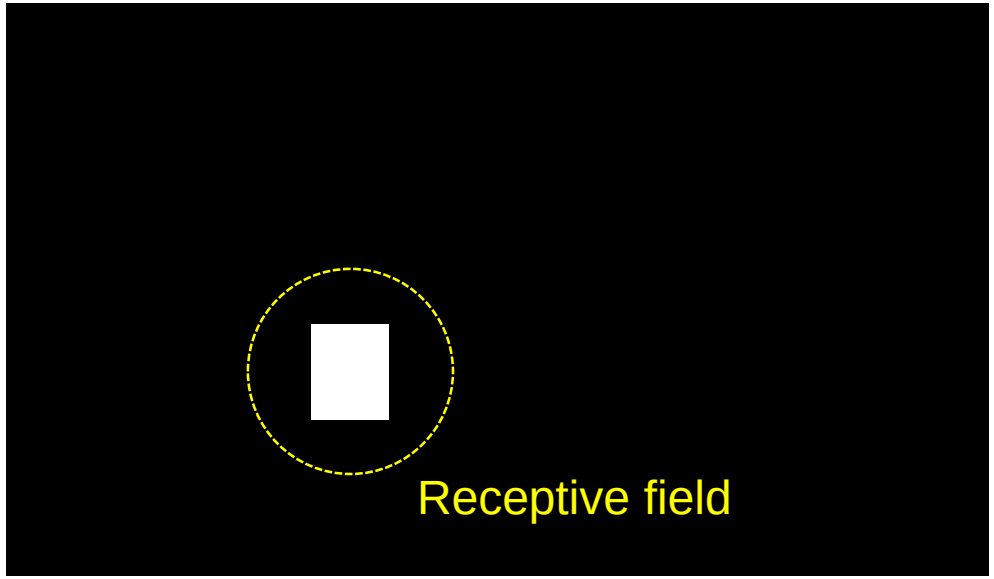
To control for random changes in spike rate, we repeat the presentation multiple times



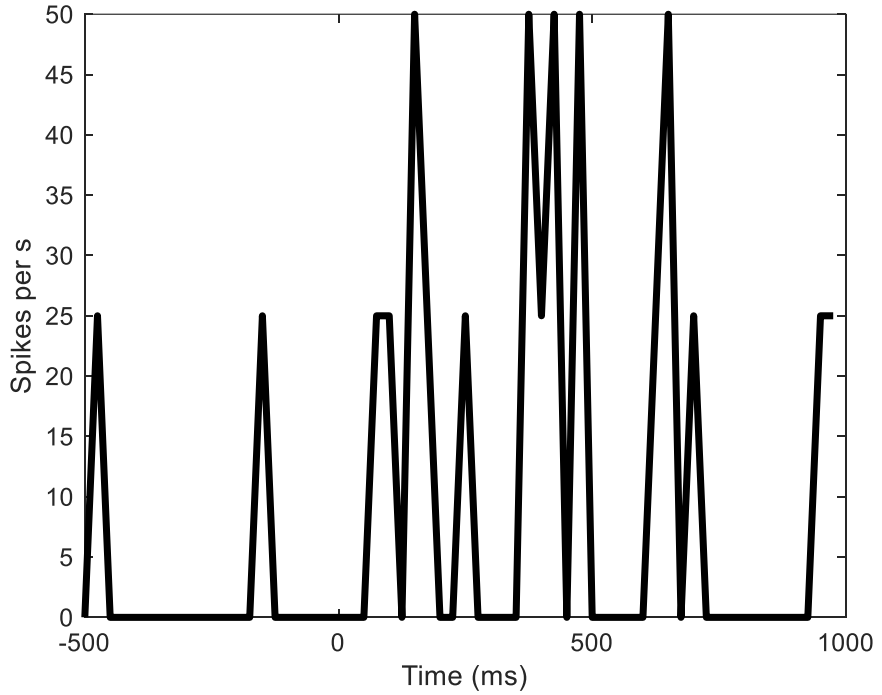
If we count the number of spikes in a time bin (say, 25 ms)



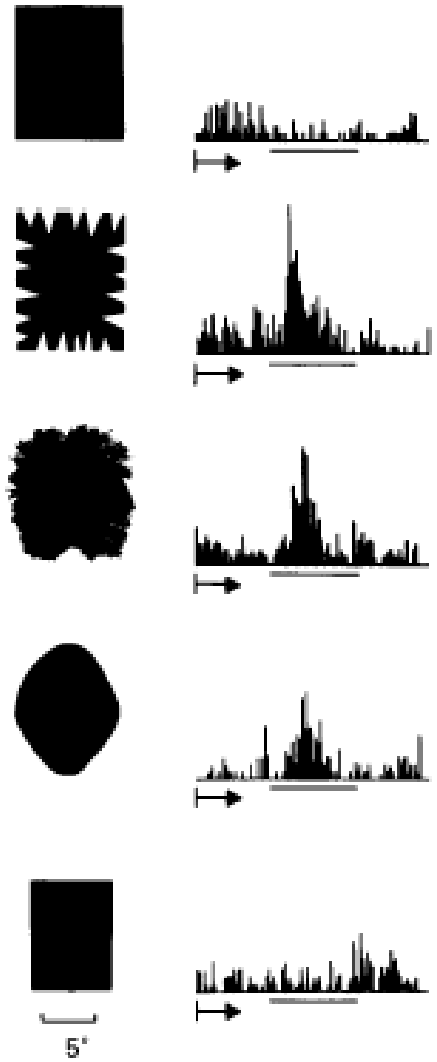




We can derive a peri-stimulus histogram (PSTH)



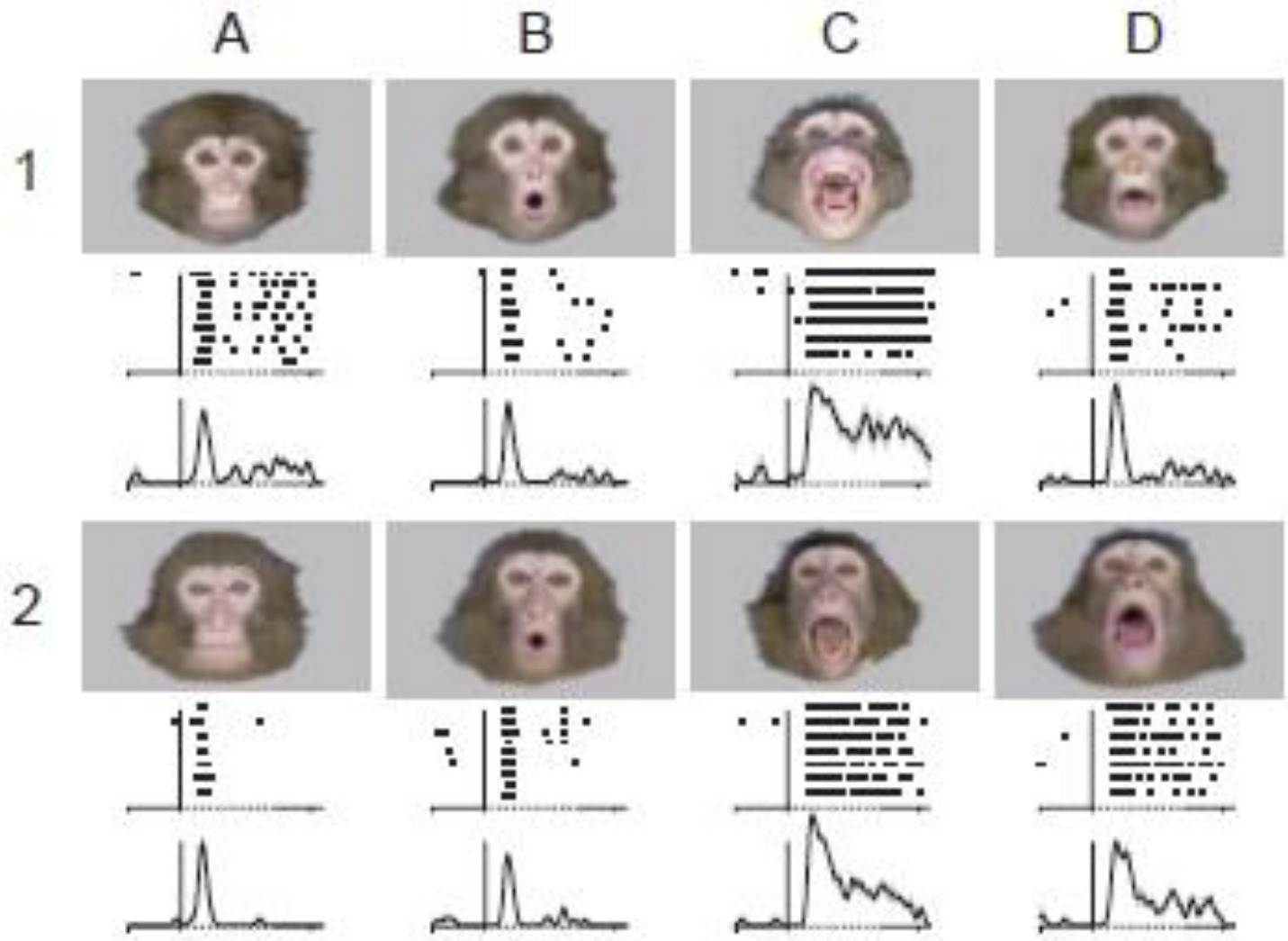
IT cells emit different numbers of spikes and show different PSTH profiles in response to different images...



STIMULUS-SELECTIVE PROPERTIES OF INFERIOR TEMPORAL NEURONS IN THE MACAQUE[1]

ROBERT DESIMONE,*,[2] THOMAS D. ALBRIGHT,‡ CHARLES G. GROSS,‡ AND CHARLES BRUCE§

PSTH shape can show when different types of preferences are expressed by the neuron



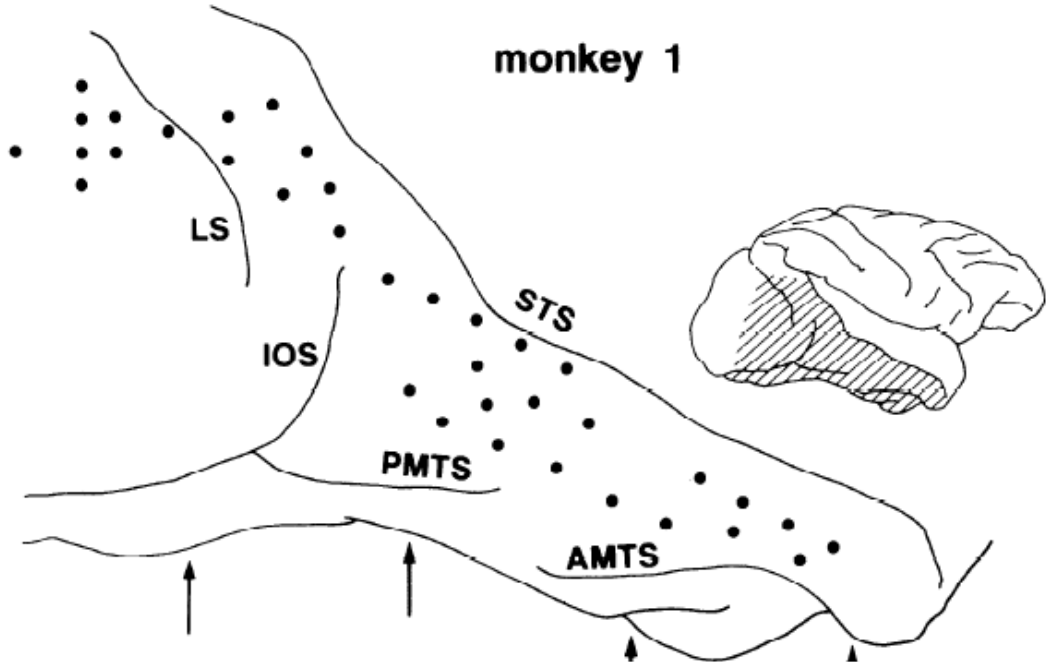
**Global and fine informatic..
coded by single neurons
in the temporal visual cortex**

Yasuko Sugase*†‡§, Shigeru Yamane*, Shoogo Ueno‡
& Kenji Kawano*

PSTHs also show that IT neurons prefer more complex images depending on their position in the temporal lobe

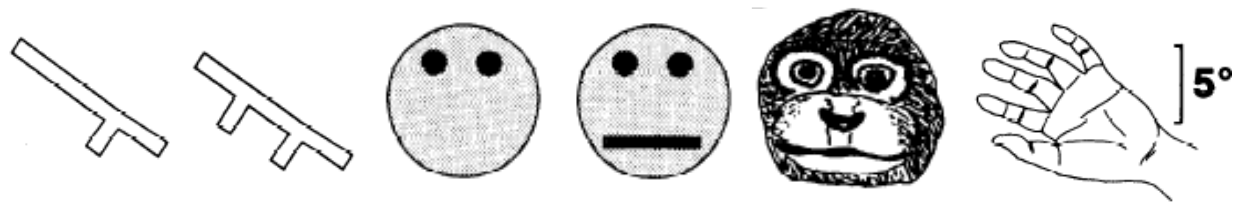
Keiji Tanaka
RIKKEN Institute

Recorded responses from single neurons along the occipito-temporal lobe

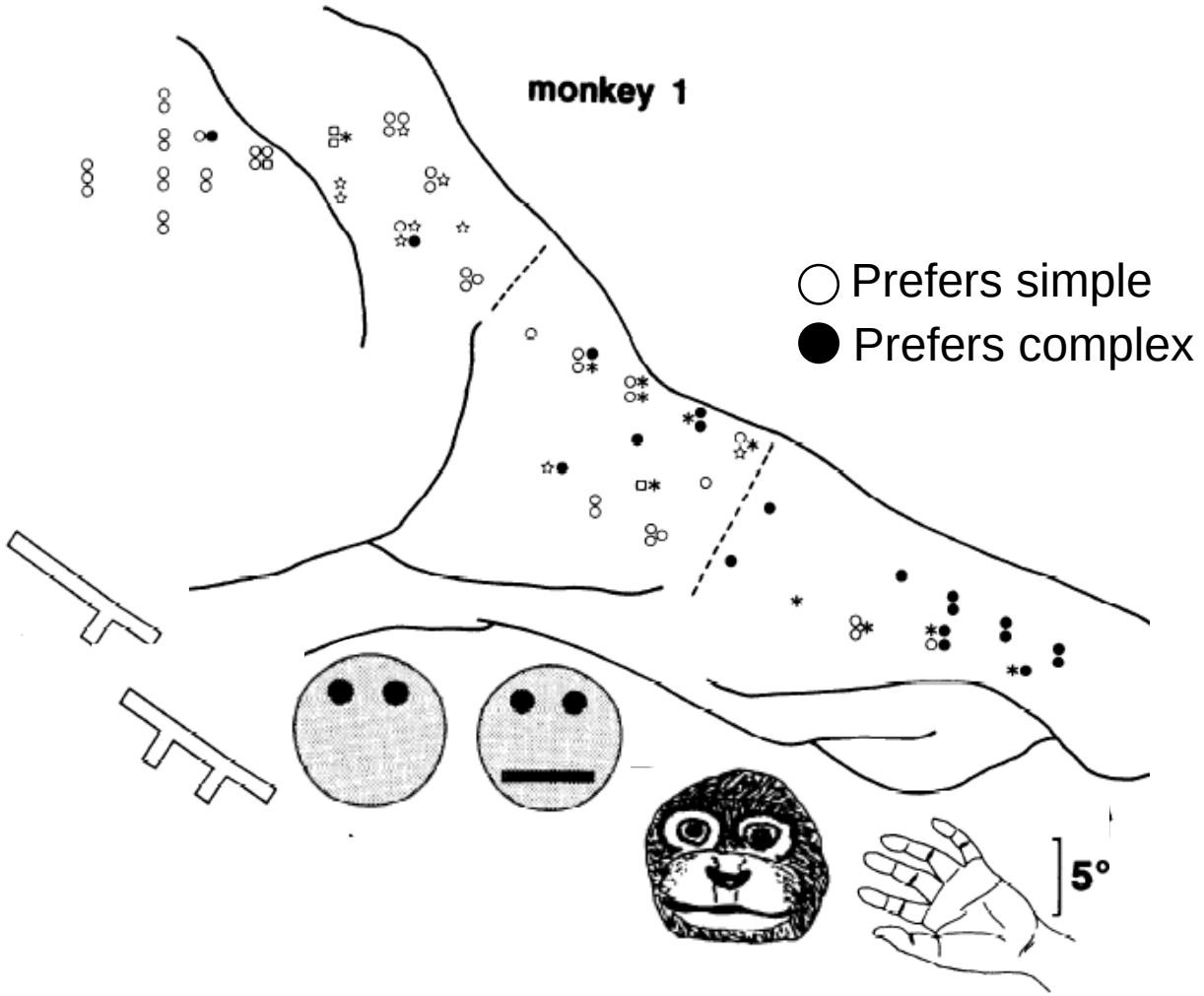


EUCALY KOBATAKE AND KEIJI TANAKA

They stimulated neurons using complex and simple images

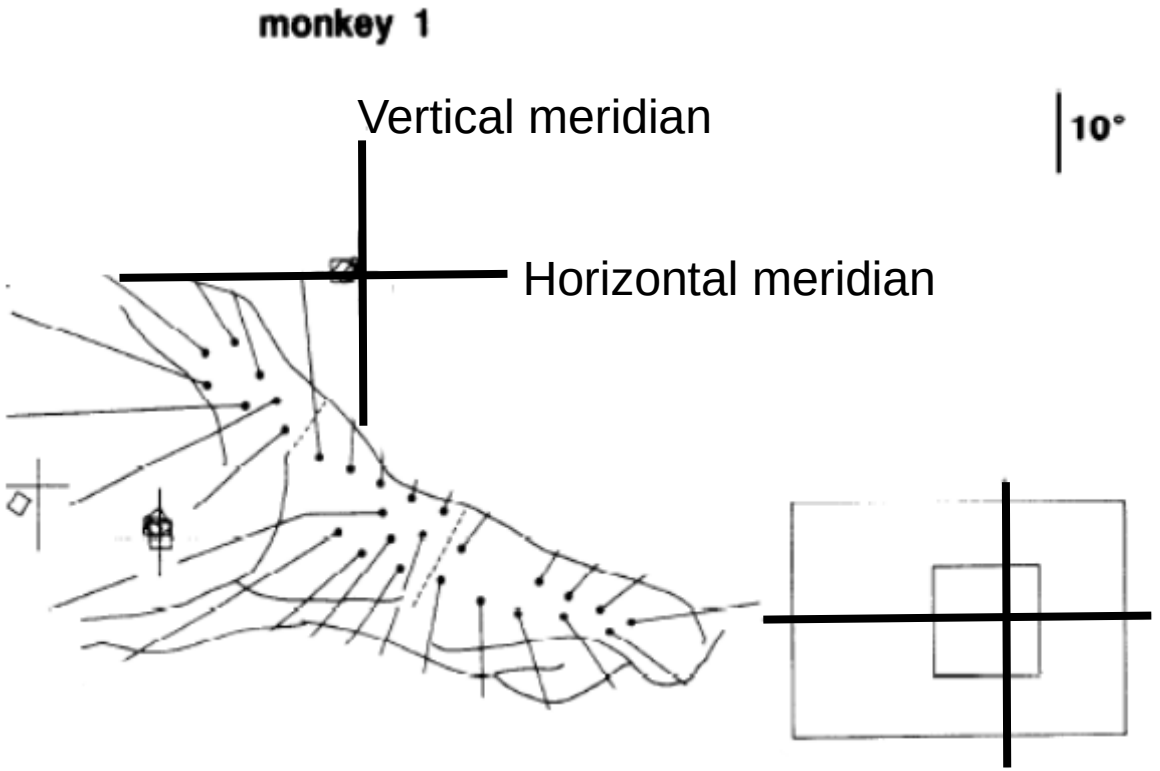


IT cells closer to V1 (more posterior) prefer simpler features.



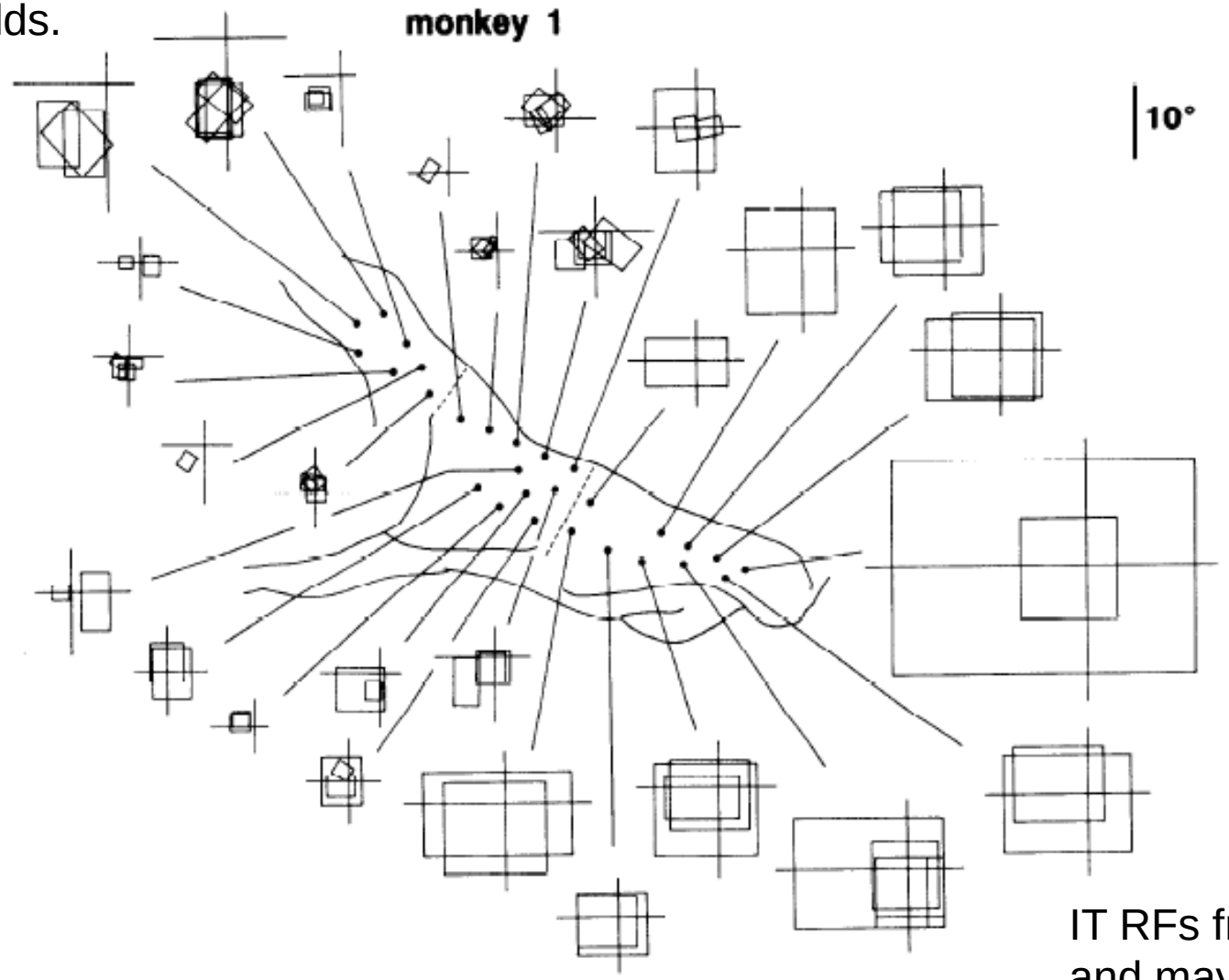
EUCALY KOBATAKE AND KEIJI TANAKA

IT cells closer to V1 (more posterior)
have smaller receptive fields.



EUCALY KOBATAKE AND KEIJI TANAKA

IT cells closer to V1 (more posterior)
have smaller receptive fields.

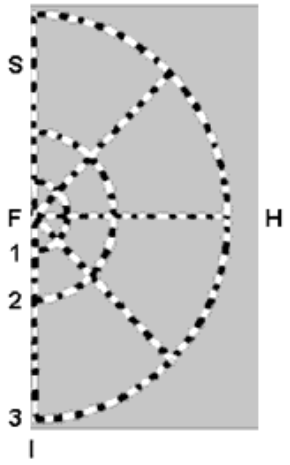


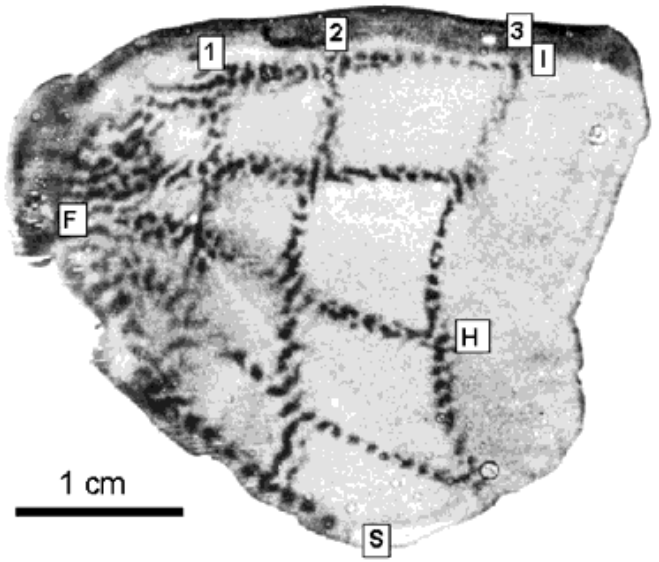
IT RFs frequently include the fovea, and may extend to the contralateral hemifield.

EUCALY KOBATAKE AND KEIJI TANAKA

IT cells also change in their retinotopy

Retinotopy: when cells which are physically near one another in the brain respond to parts of the visual field that are also near each other





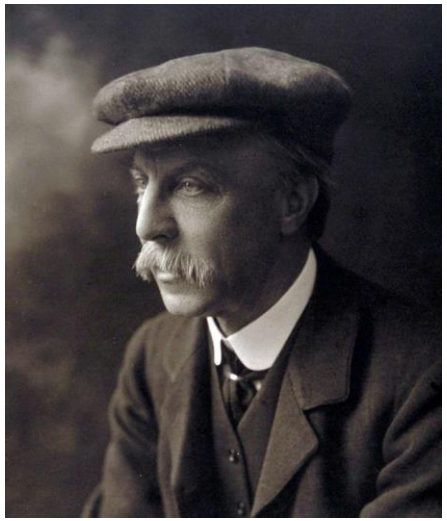
Tootell et al (1988a)

IT cells further from V1 show less and less retinotopy, organizing themselves by feature preference.

Many studies thus established that IT neurons prefer complex shapes

Historically, this idea met with resistance. Let’s review why.

Since the 1800s, it has been known that the brain is divided into functional regions

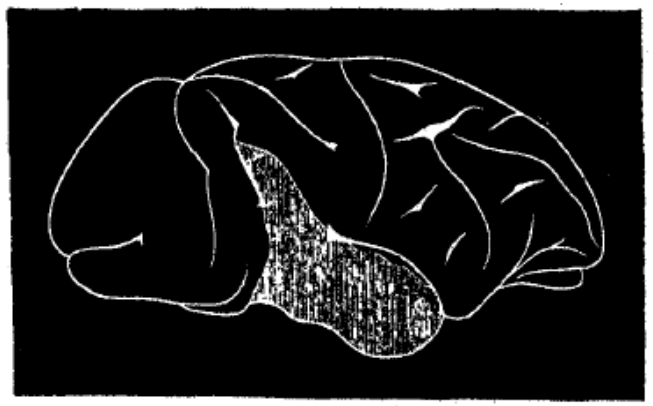


Edward Albert Schafer, 1850-1935
British physiologist

XI. *An Investigation into the Functions of the Occipital and Temporal Lobes of the Monkey's Brain.*

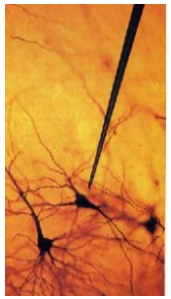
By Sanger Brown, *M.D., and* E. A. Schäfer, *F.R.S., Jodrell Professor of Physiology in University College, London.**

Received November 24,—Read December 15, 1887.



"…the animals, although they received and responded to impressions from all the senses, appeared to understand very imperfectly the meaning of such impressions…even objects most familiar to the animals were carefully examined, felt, smelt and tasted exactly … as an entirely new object…

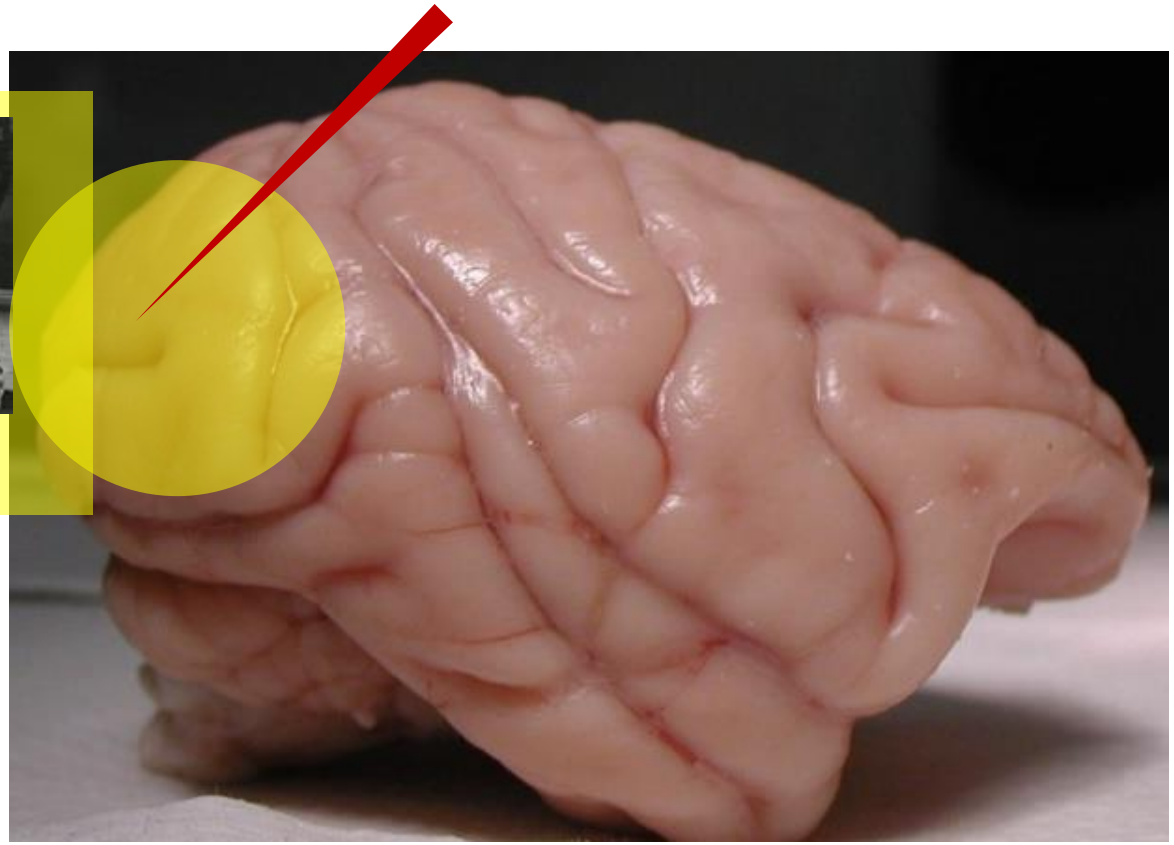
For decades thereafter, investigators performed many lesions experiments to correlate brain locations with behavioral changes.

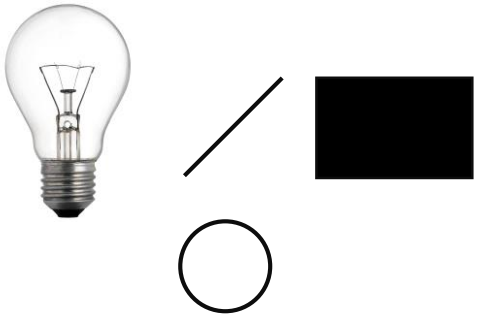


But they started using electrophysiology as their primary tool for mapping, we learned much more.

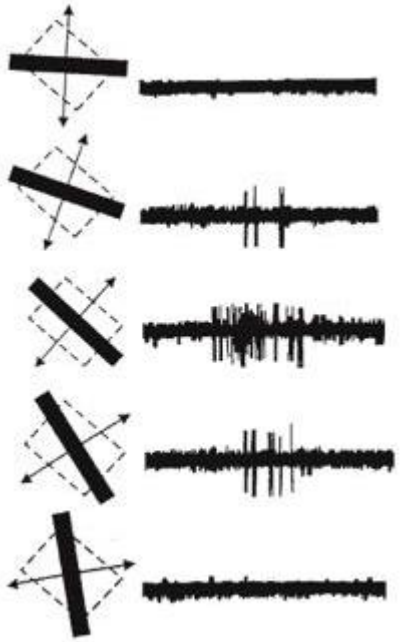
Hubel and Wiesel first showed us that cells in V1 responded differently to the orientation of edges

1962

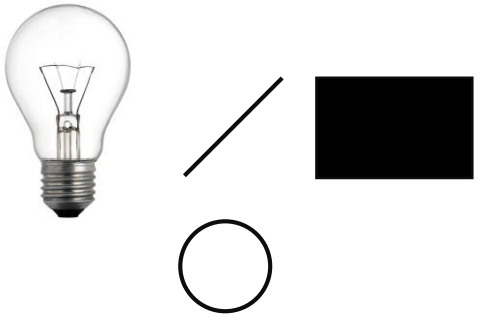




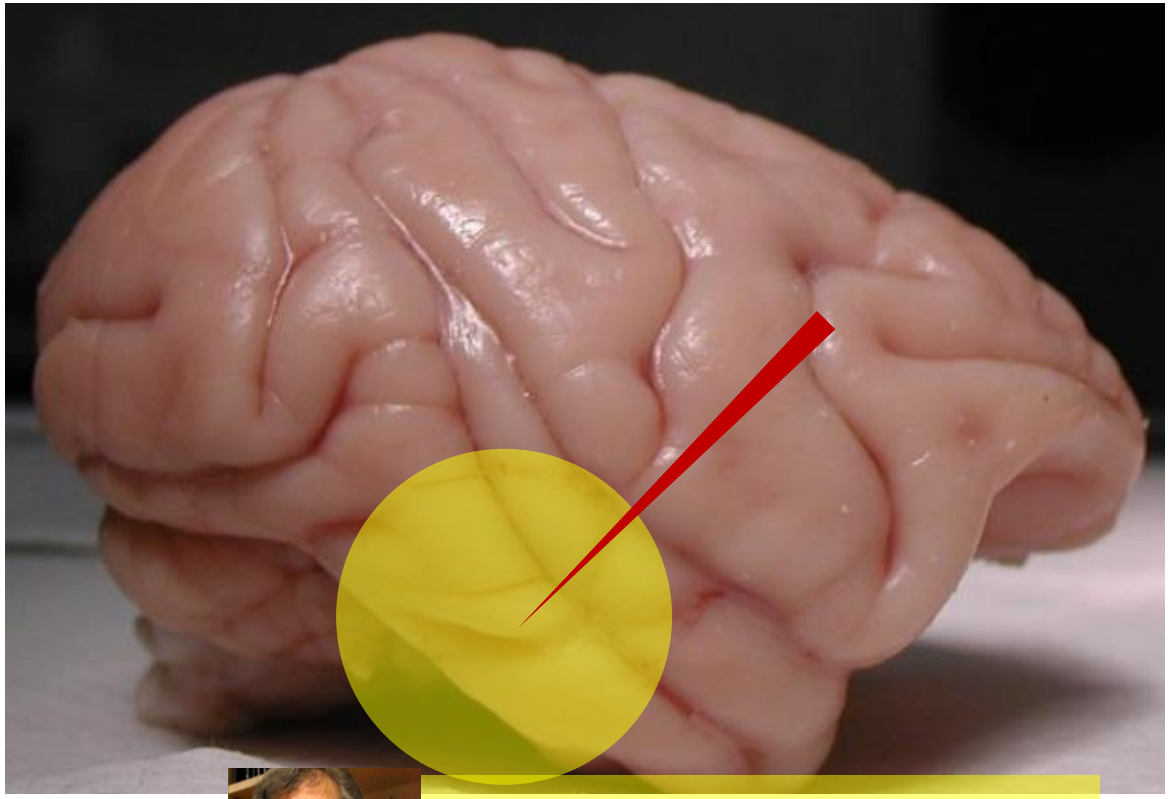
Diffuse light, edges, other simple geometric images



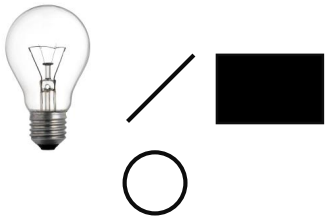
In early days, neurons in other parts of the brain were stimulated with similar images



Diffuse light, edges, other simple geometric images



Charlie Gross, Peter Schiller



No great responses. No receptive fields.
Either this is a very different brain area compared to V1, or the right stimuli weren't used….

They went back to look for effects of attention…

“We set up a board in front of the monkeys with little windows or "peep holes" to which we could apply our eye or present such objects as a finger, a burning Q-tip, or a bottle brush. Most of the units responded vigorously…”

(1969)

Visual Receptive Fields of Neurons in Inferotemporal Cortex of the Monkey

C. G. Gross, D. B. Bender and C. E. Rocha-Miranda

ence. The first is that by largely confining the stimuli to bars, edges, rectangles, and circles we may never have found the “best” sitmulus for each unit. There were several units that responded most strongly to more complicated figures. For example, one unit that responded to dark rectangles responded much more strongly to a cut-out of a monkey hand, and the more the stimulus looked like a hand, the more strongly the unit responded to it.

“When we wrote the first draft...we did not have the nerve to include the ‘hand’ cell until [department head] Teuber urged us to do so.”

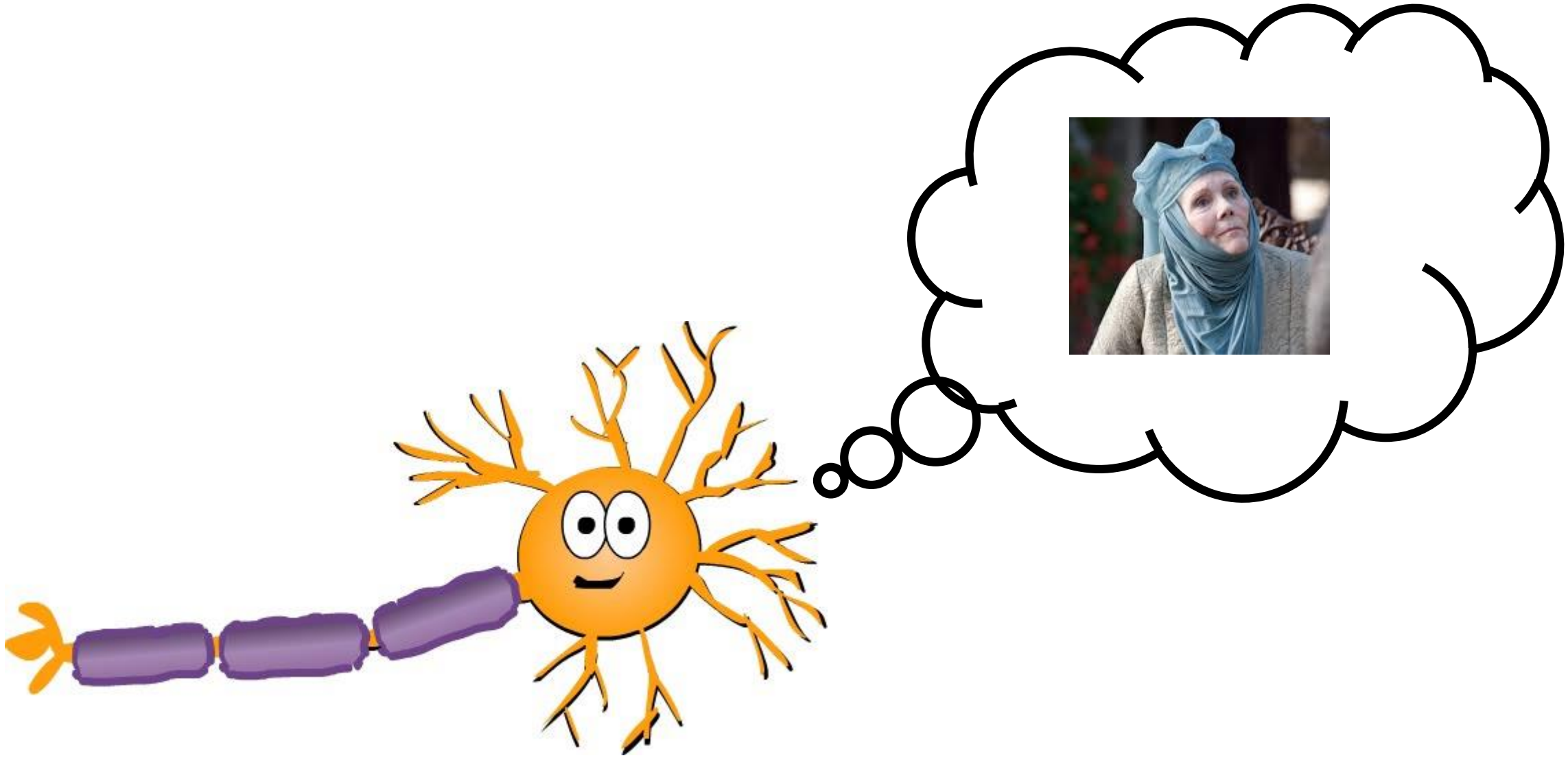
They did not publish the existence of face cells until 1981.

Jerzy Konorski (1967) had recently proposed “gnostic” units – cells that represented “unitary perceptions.” Suggested that they live in IT.

Charles G. Gross

How Inferior Temporal Cortex Became a Visual Area

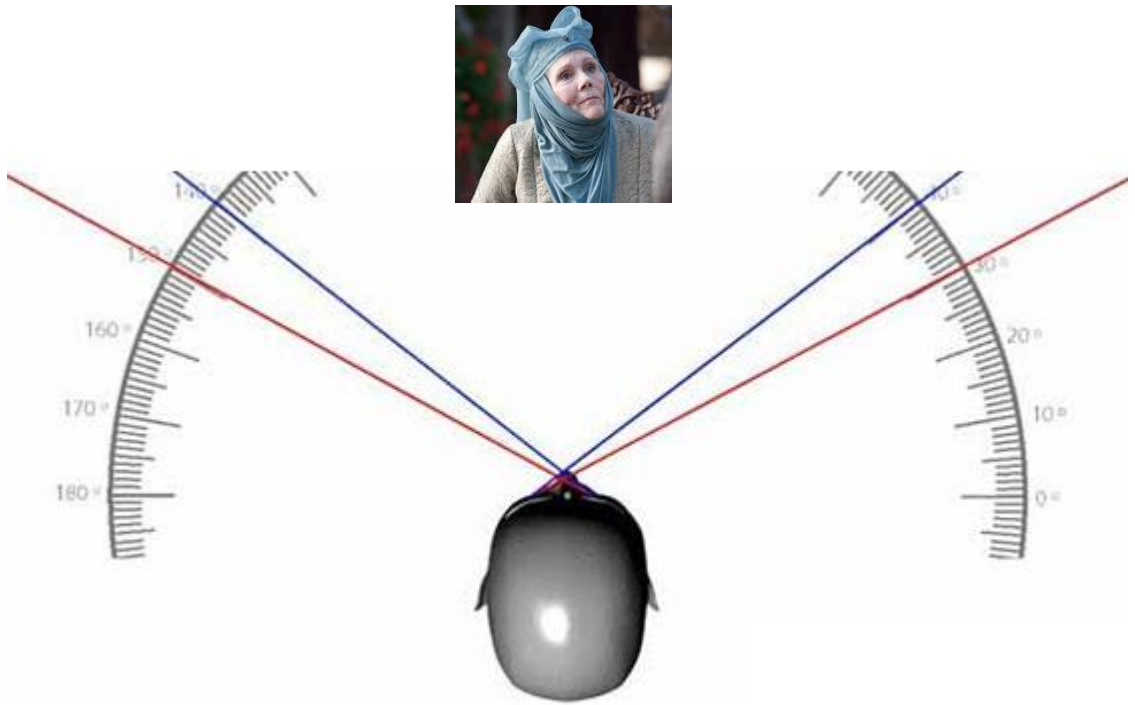
The grandmother cell hypothesis



Over the years, dozens of teams have confirmed that IT neurons do prefer complex images

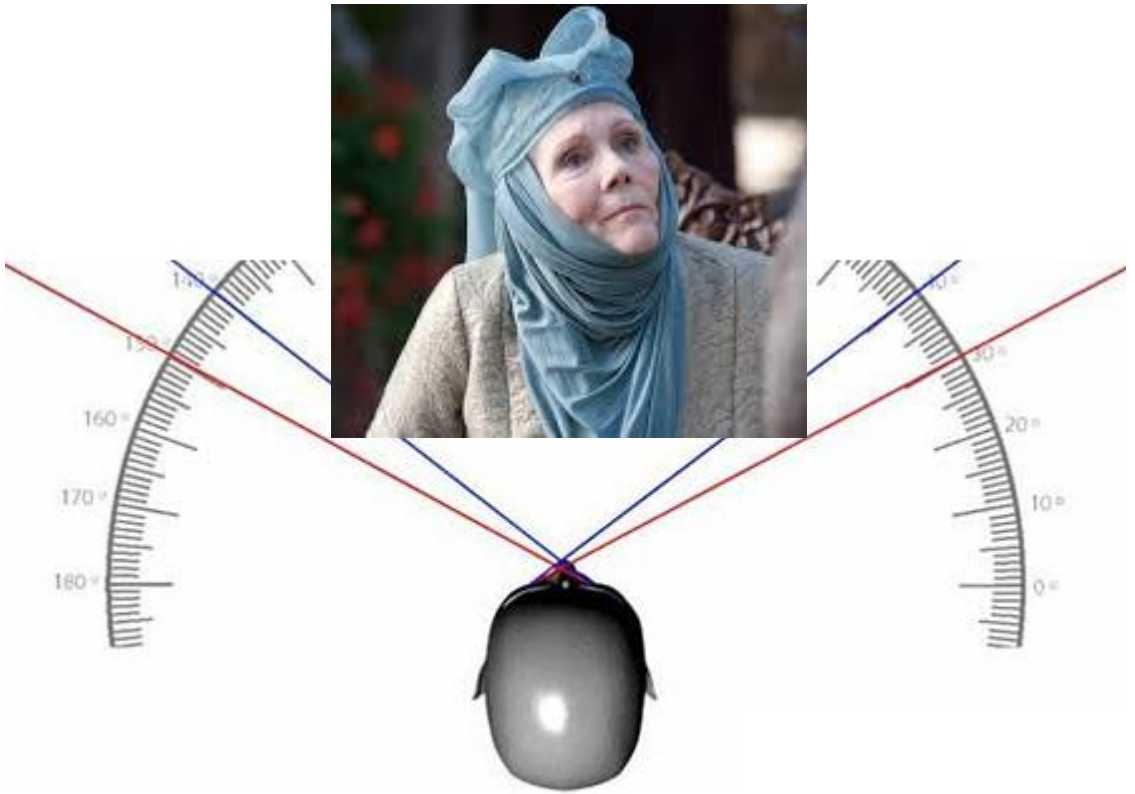
So are these grandmother cells…?

When *we* perceive grandma, we can recognize her even if her image on our retina…



When we perceive grandma, we can recognize her even if her image on our retina…

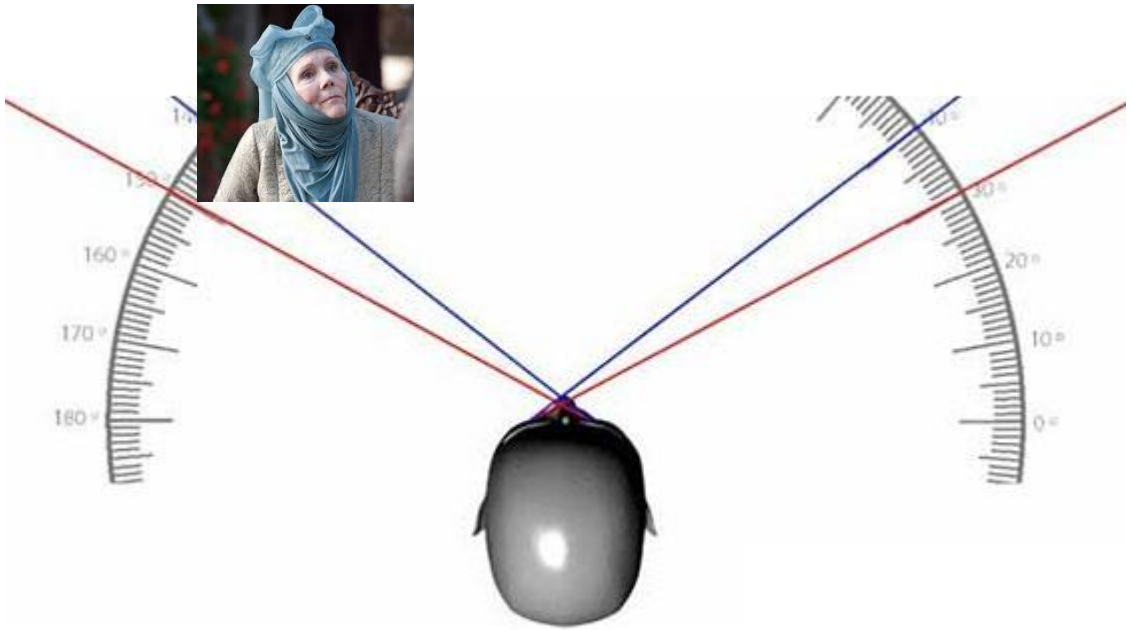
- changes size



When we perceive grandma, we can recognize her even if her image on our retina…

- changes size

- moves to a different place

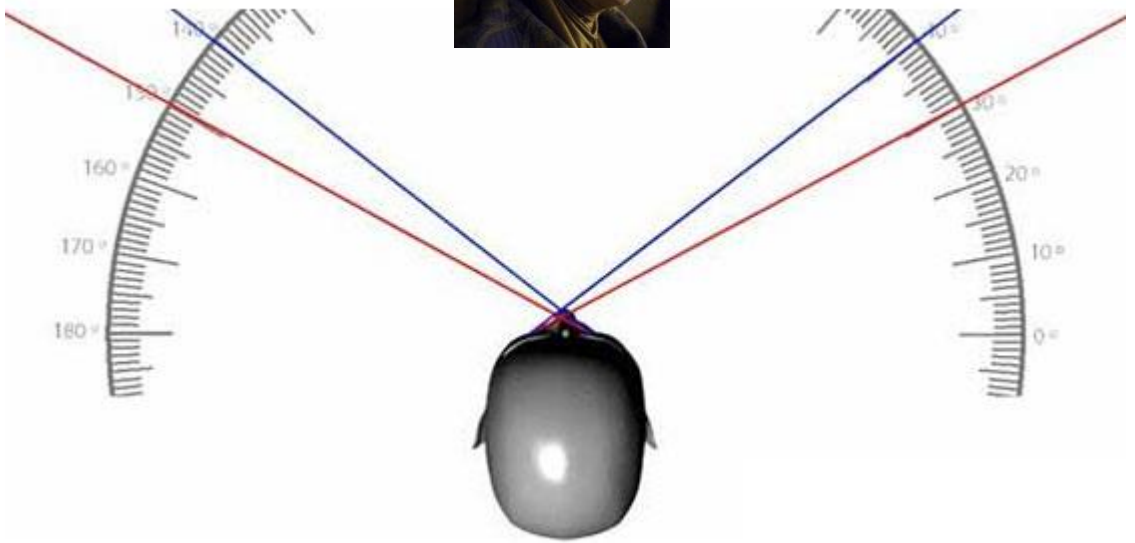


When we perceive grandma, we can recognize her even if her image on our retina…

- changes size

- moves to a different place

- rotates in 3-D (viewpoint position)



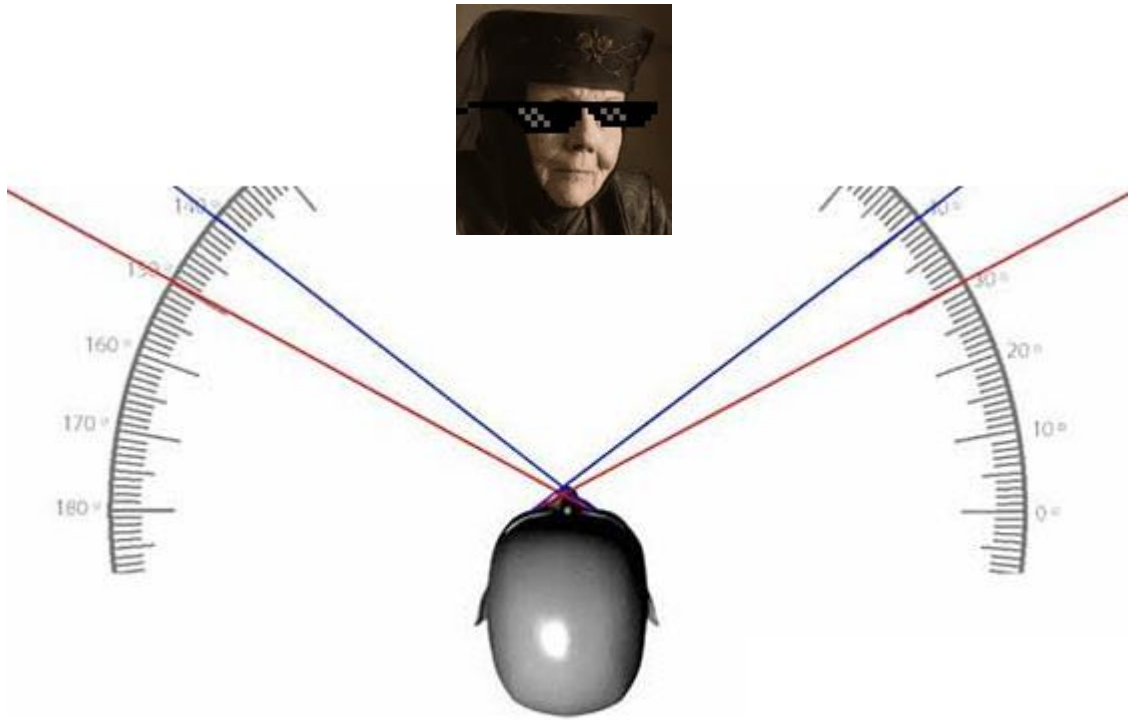
When we perceive grandma, we can recognize her even if her image on our retina…

- changes size

- moves to a different place

- rotates in 3-D (viewpoint position)

- is occluded by an object



3. How well do IT neurons tolerate these changes?
- the problem of achieving invariance

Tomaso Poggio, MIT

One compelling summary of the goal of the ventral stream:

To compute object representations that are invariant to different transformations

(selectivity is much, much easier then!)

most experiments on IT have characterized

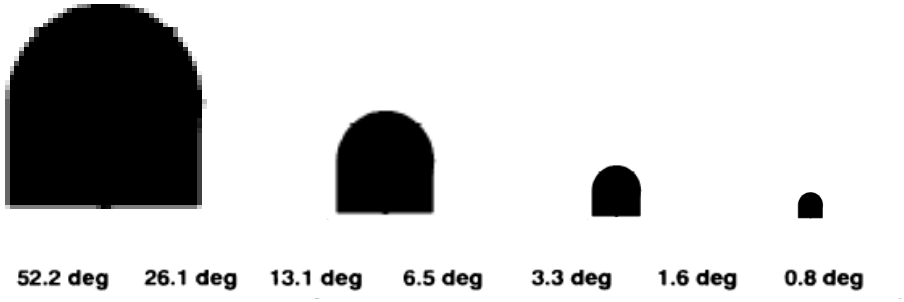
their ability to respond to their preferred stimulus

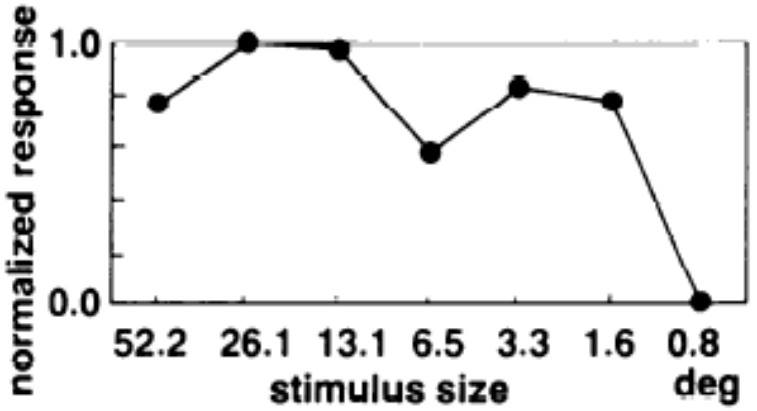
regardless of “nuisance” variables (e.g. position, size, rotation, lighting, occlusion, texture…)

how well do IT neurons respond to their preferred image when it changes **size**?

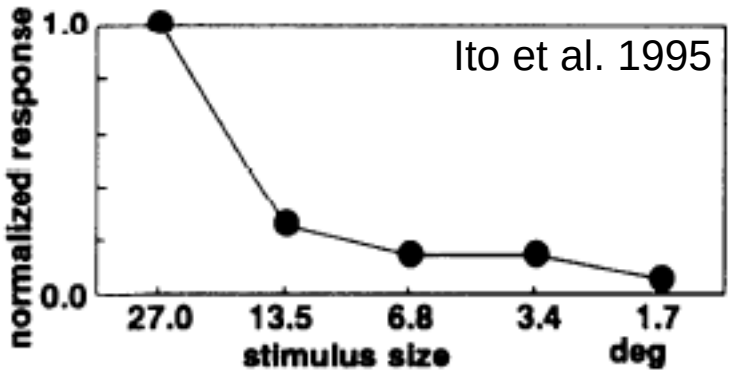
One way to test size invariance: present the same image at different sizes. Does the firing rate change?

Ito et al. 1995 presented different images to IT neurons at different sizes





Sometimes, cells can show little variation in their spike responses to different sizes.



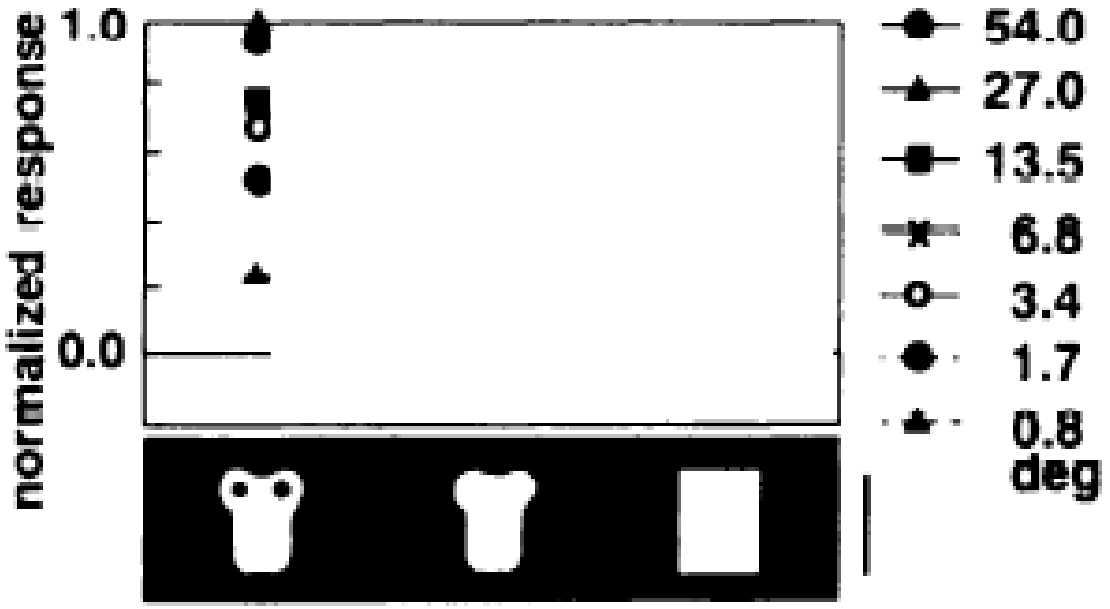
Most of the time, they vary their responses.

More commonly, size tolerance means that neurons keep their ranked image preferences across size changes.

Definition: if a neuron likes image X more than image Y when X and Y are small…

and it also likes image X more than image Y when X and Y are big,

then it is size-invariant



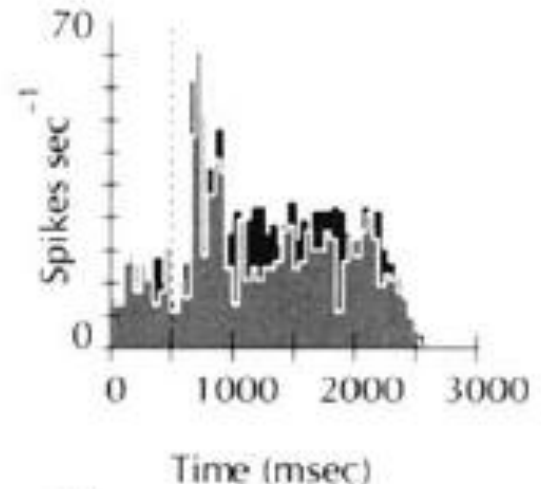
This neuron shows the same relative preference *despite* size changes.

Ito et al. 1995

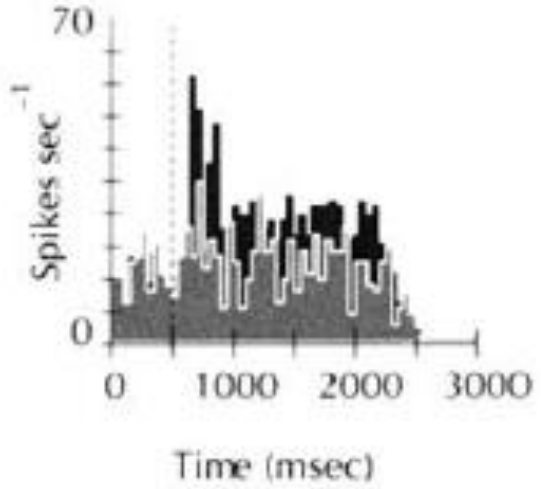
how well do IT neurons respond to their preferred image when it changes **position**?

Logothetis et al. (1995) presented the same object at different positions inside a neuron's RF

Position #1

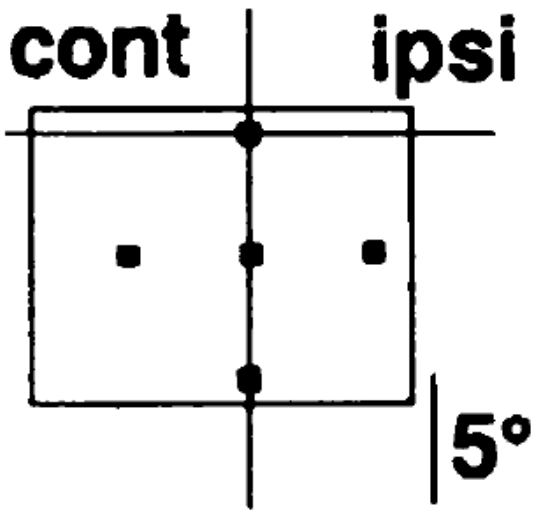


Position #2

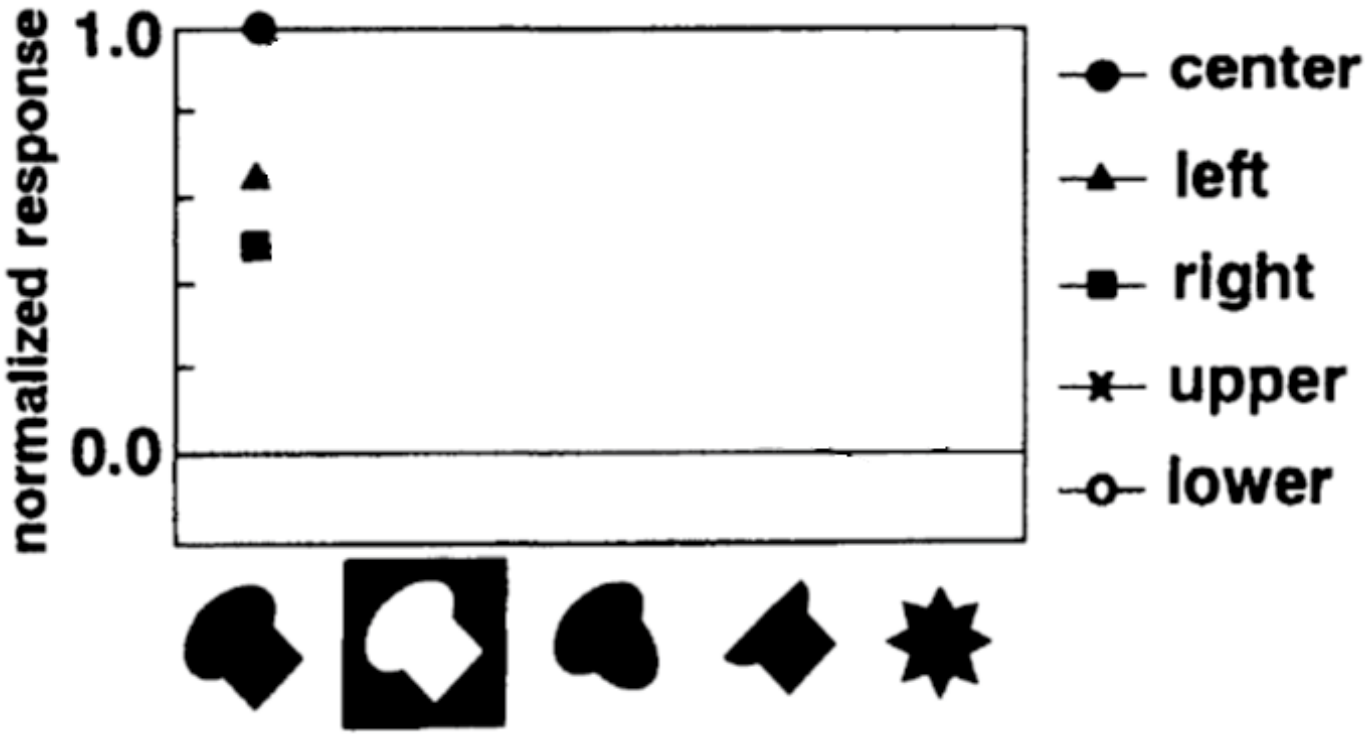


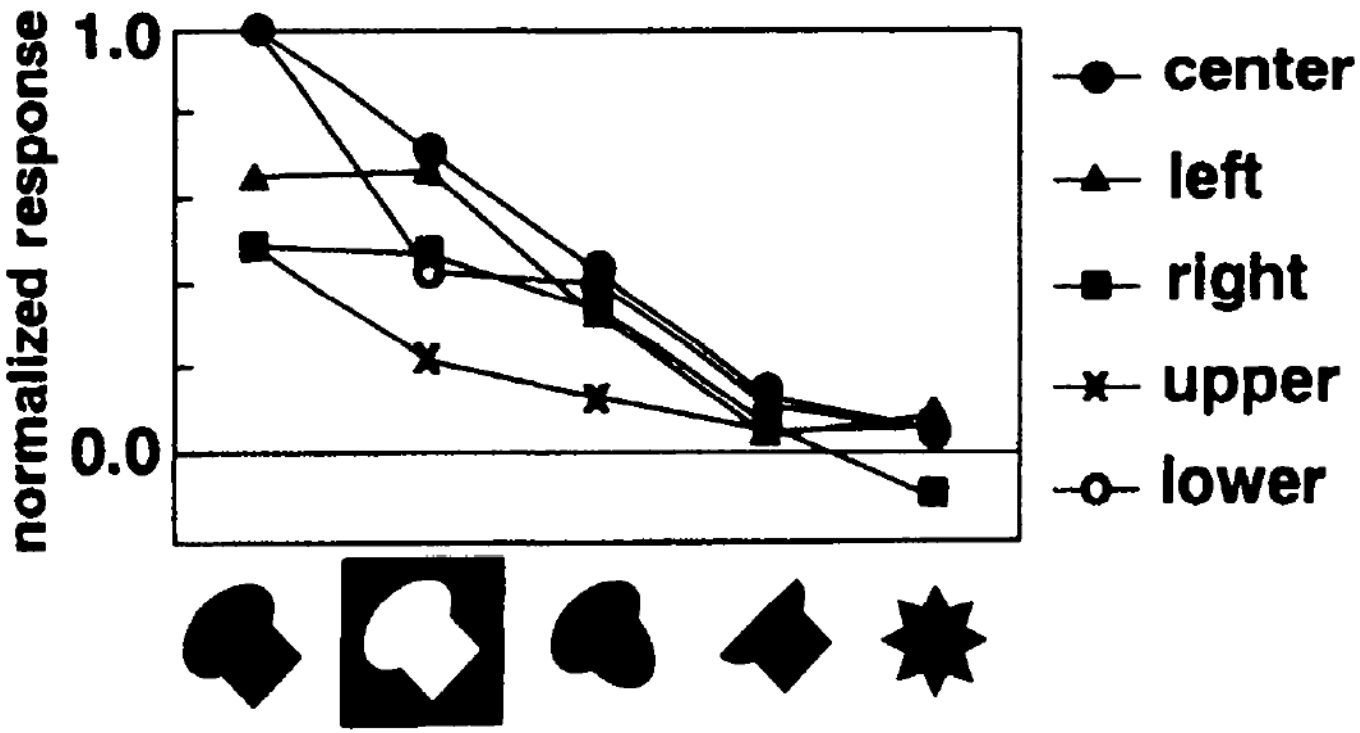
This neuron shows the same firing rate activity AND relative preference despite position changes.

Ito et al. (1995) presented images in five positions inside a neuron's RF



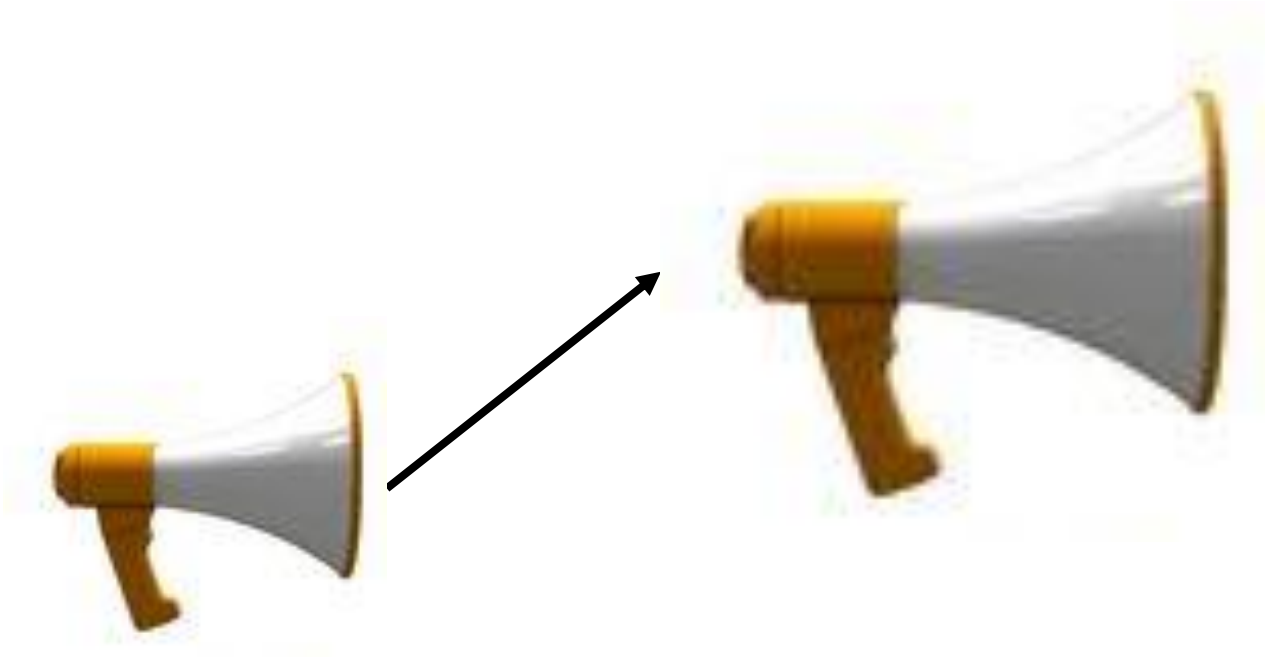
This neuron shows different firing rates as a function of position for a given image





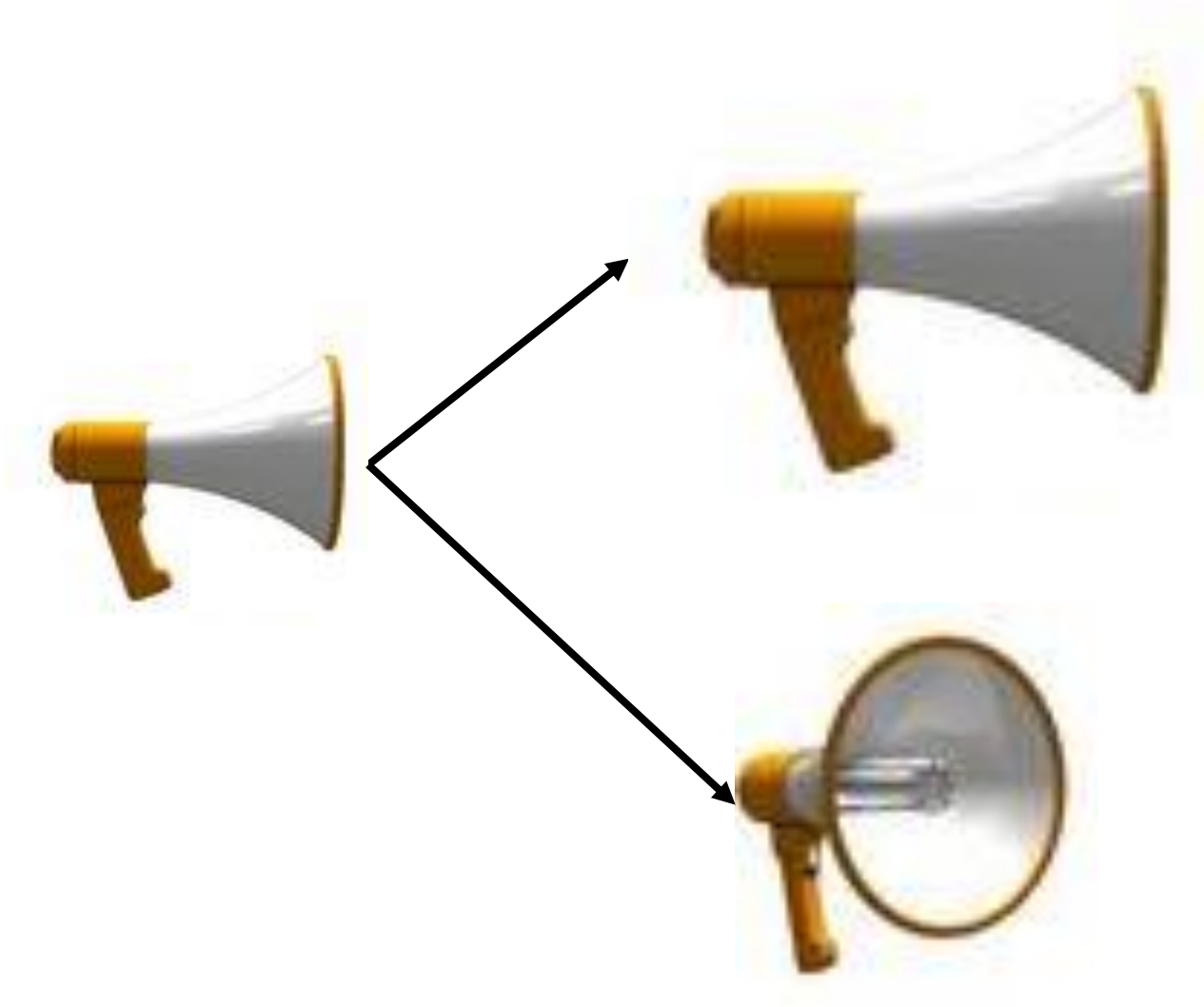
But they can also show the same relative preference for objects *despite* position changes.

Some image transformations are more problematic than others



When an object changes *size* or *position*, it is possible to match the images because all key points are the same

Some image transformations are more difficult than others

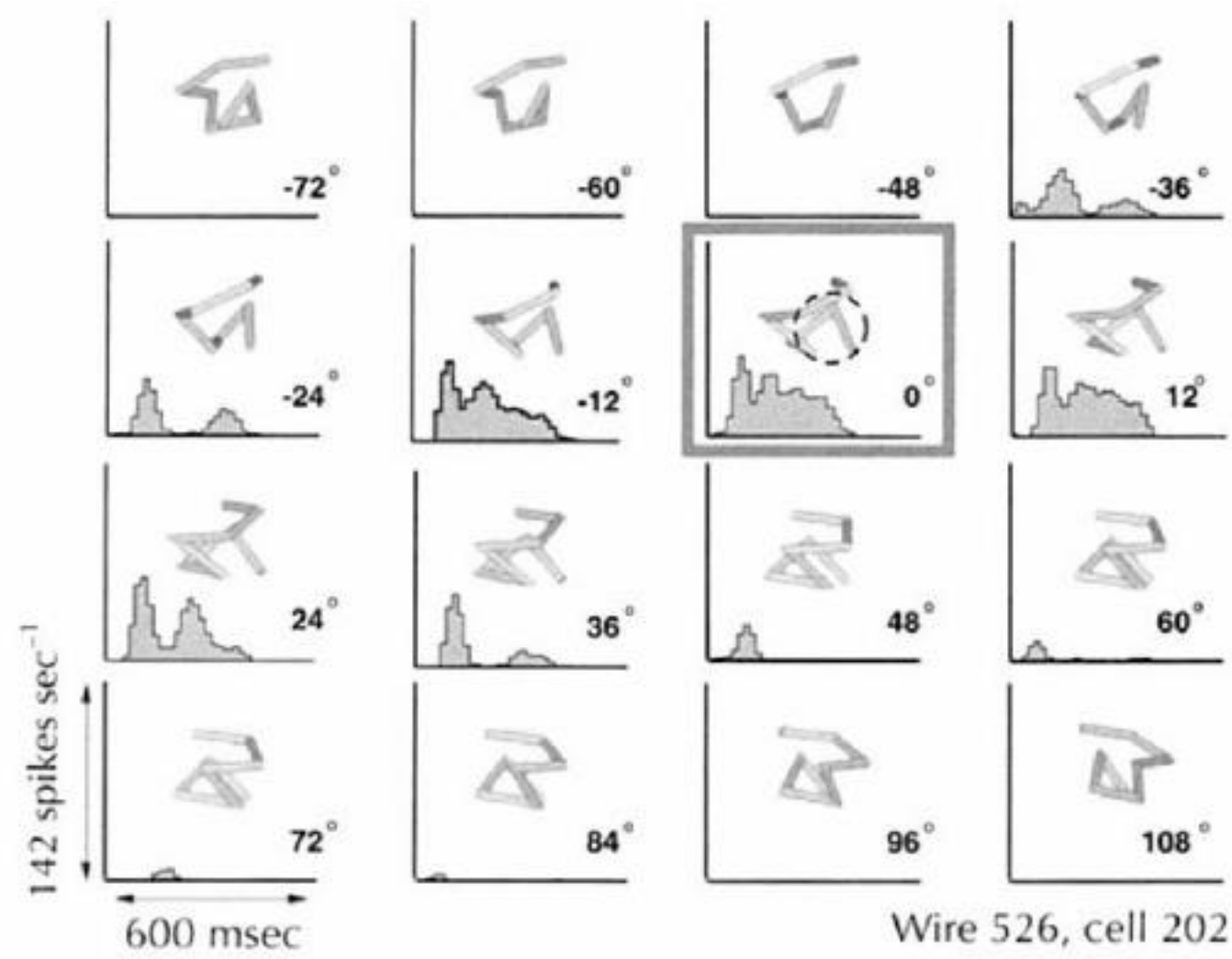


When an object changes size or position, it is possible to match the images because all interest features are the same

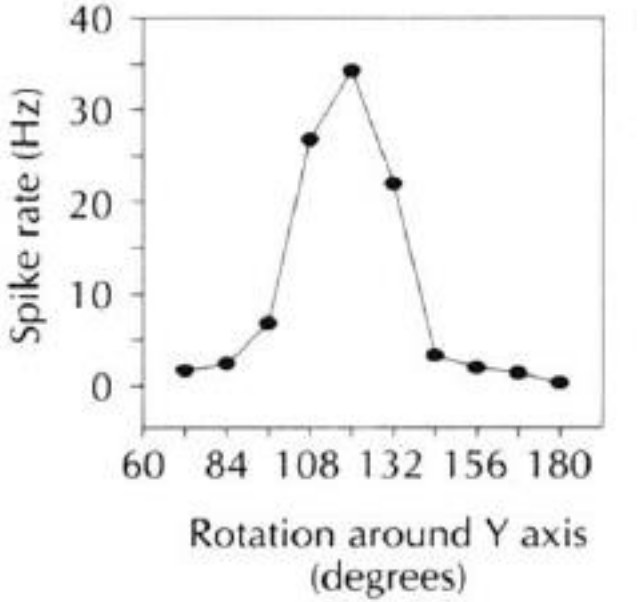
When an object rotates in 3-D space, entirely new parts may emerge

how well do IT neurons respond to their preferred image when it changes **viewpoint**?

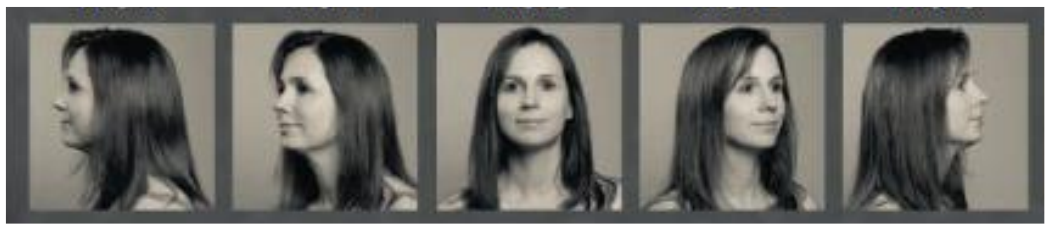
Logothetis and others (1995) showed *paperclip*-like images to IT neurons and measured their “view tuning curves”



IT neurons view tuning curves have widths of ~ 30° rotation



Can individual IT cells tolerate viewpoint changes in more complex images (e.g. faces)?

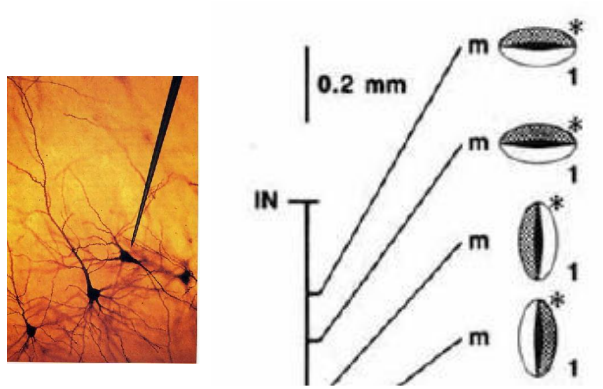


Yes, but it takes lots of work in the form of **patches**!

Current investigations in IT: patches (domains)

Cells with similar preferences cluster together at different scales

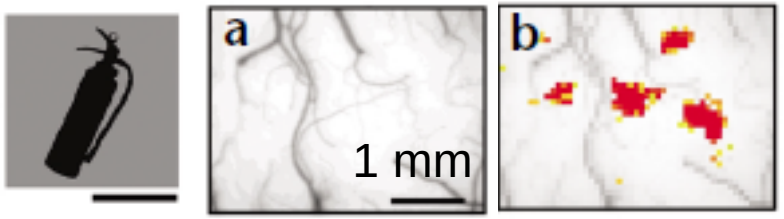
Individual neurons, tens of micrometers apart, tend to share preferences



Fujita et al 1992

(evident with electrophysiology)

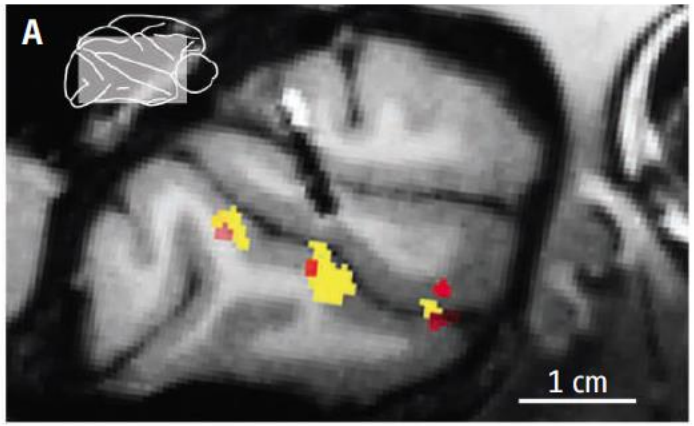
...groups of neurons at scales of <1 mm...



Tsunoda et al 2001

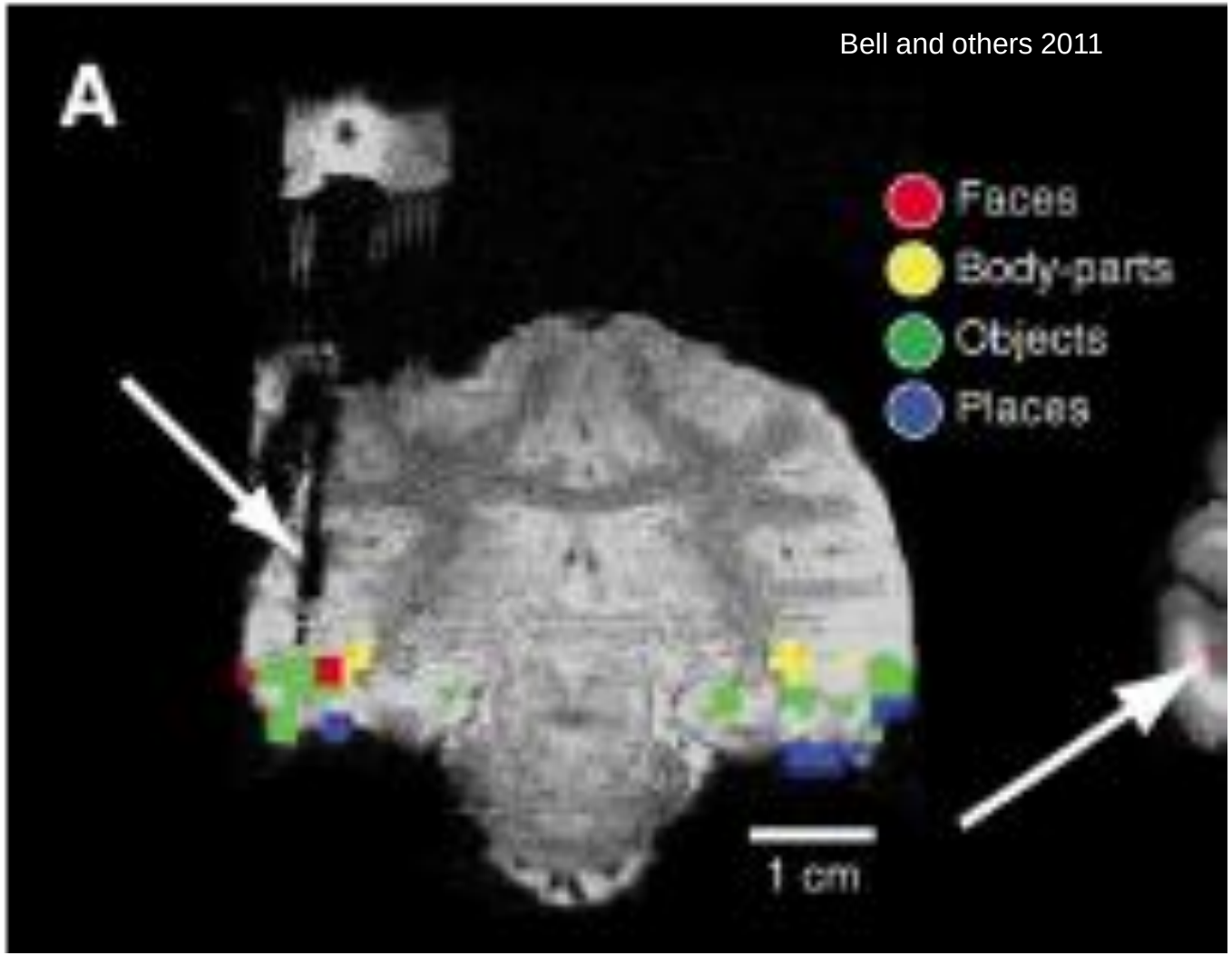
(visible with intrinsic imaging techniques)

Interestingly, also for clusters measuring up to several mm...



(visible in fMRI)

Some of these categories are abstract, and well-summarized by our vocabulary:



Thus we have “face patches,” “body part patches…”

The best-studied patches are selective for faces.
They were first characterized in humans by Sergent and Kanwisher (imaging)

And in monkeys, by Tsao, Freiwald and Livingstone (electrophysiologically)

These patches are present in virtually every monkey and human:

Why are patches necessary? Are they genetically encoded or developed purely through experience?

- We know it is computationally possible to get face recognition WITHOUT patches (as you will see in the neural networks talk)

The face network develops viewpoint invariance along its domains.

Patch ML neurons respond to similar viewpoints, despite person identity

Patch AL neurons respond to some viewpoints and their mirror images.

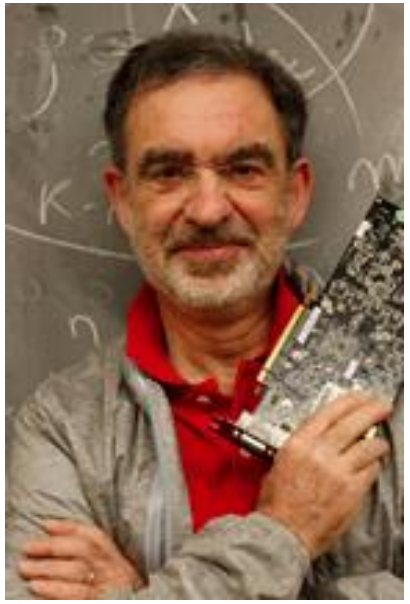
Patch AM neurons respond to identity despite viewpoint.

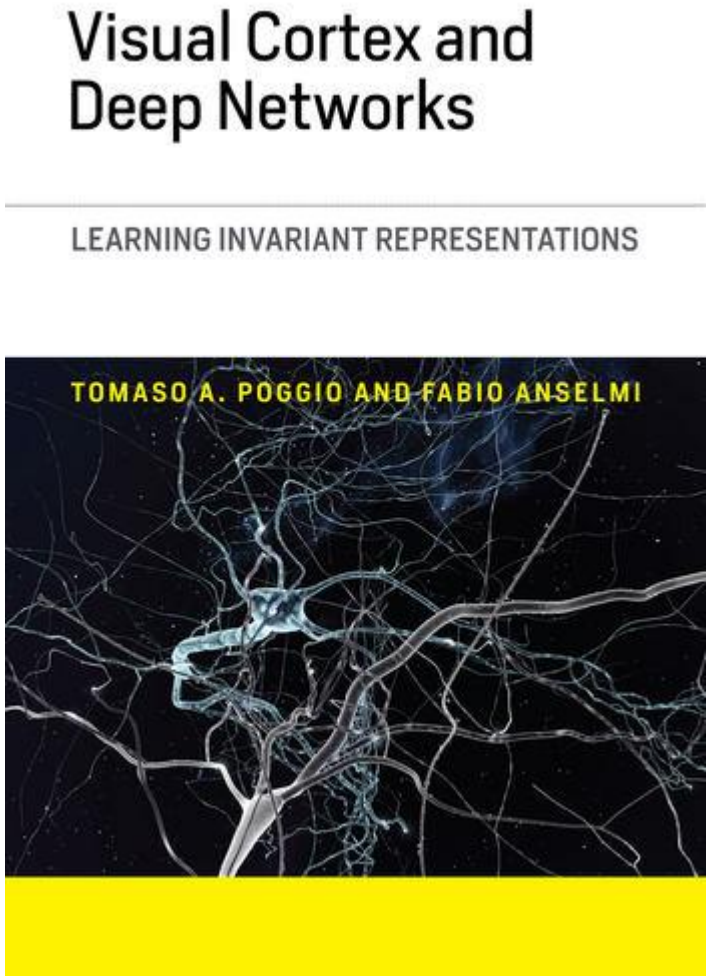
Freiwald and Tsao 2010

Figure from Charles Connor, 2010

Poggio and Anselmi have developed a general theory that proposes that viewpoint invariance is the key reason for the development of patches

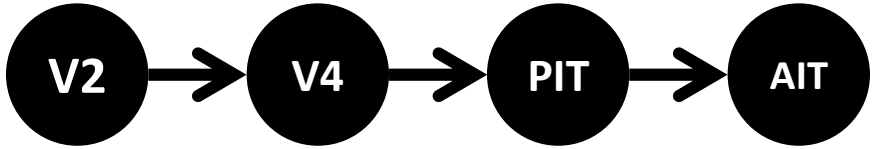
Tomaso Poggio, MIT

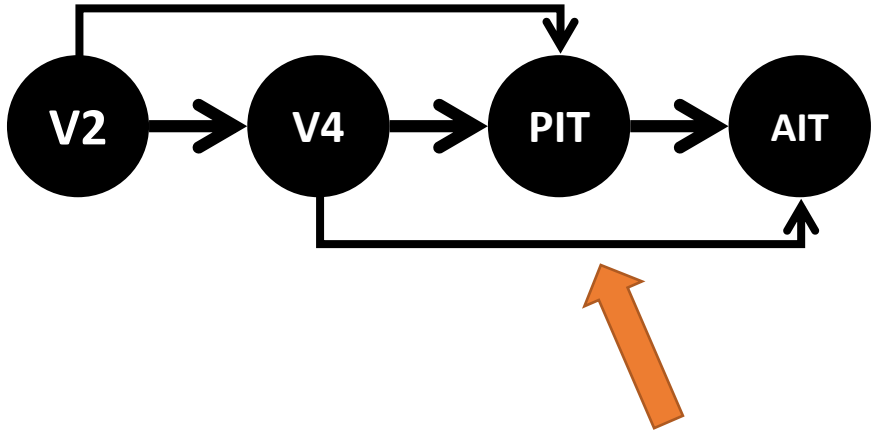




Current investigations in IT (2): bypass pathways and feedback

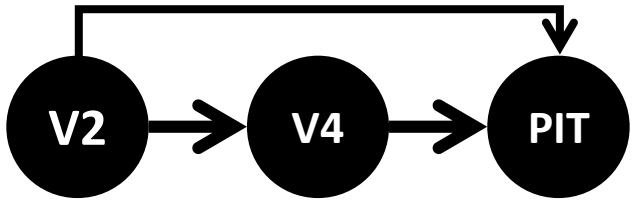
Because IT depends more on V4 than in other regions, we can think of IT as part of a “stream”

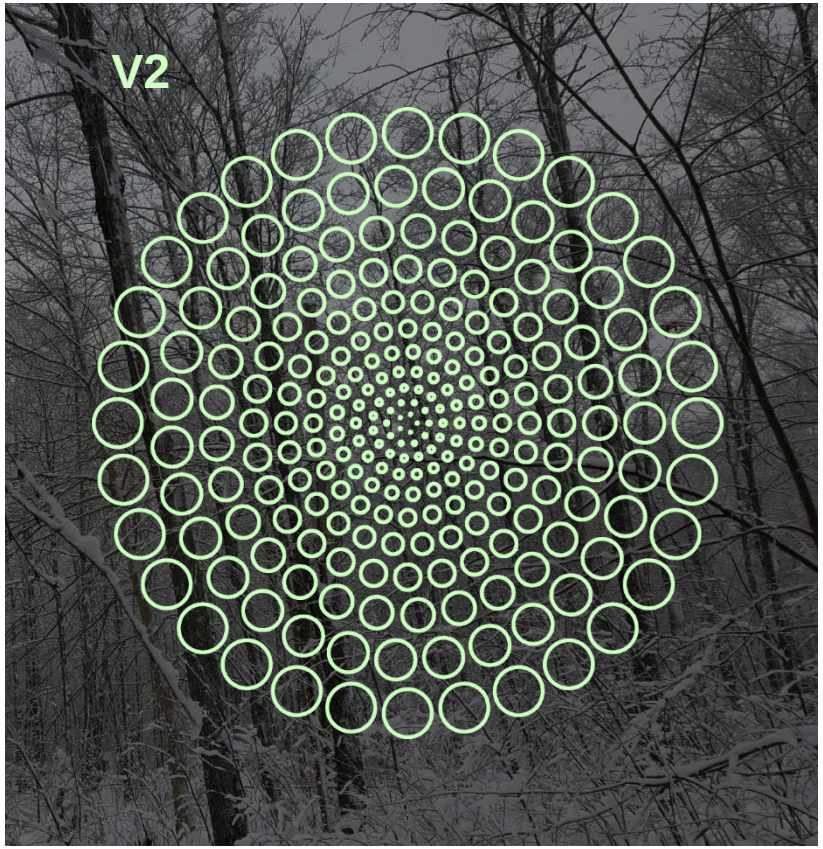


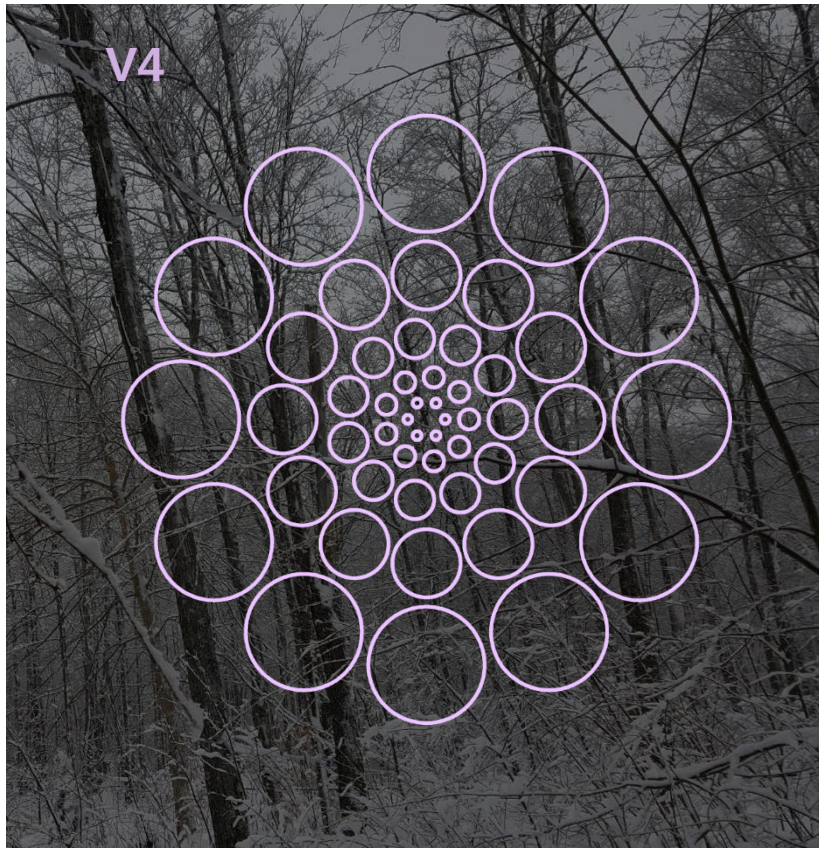


What are these guys doing?

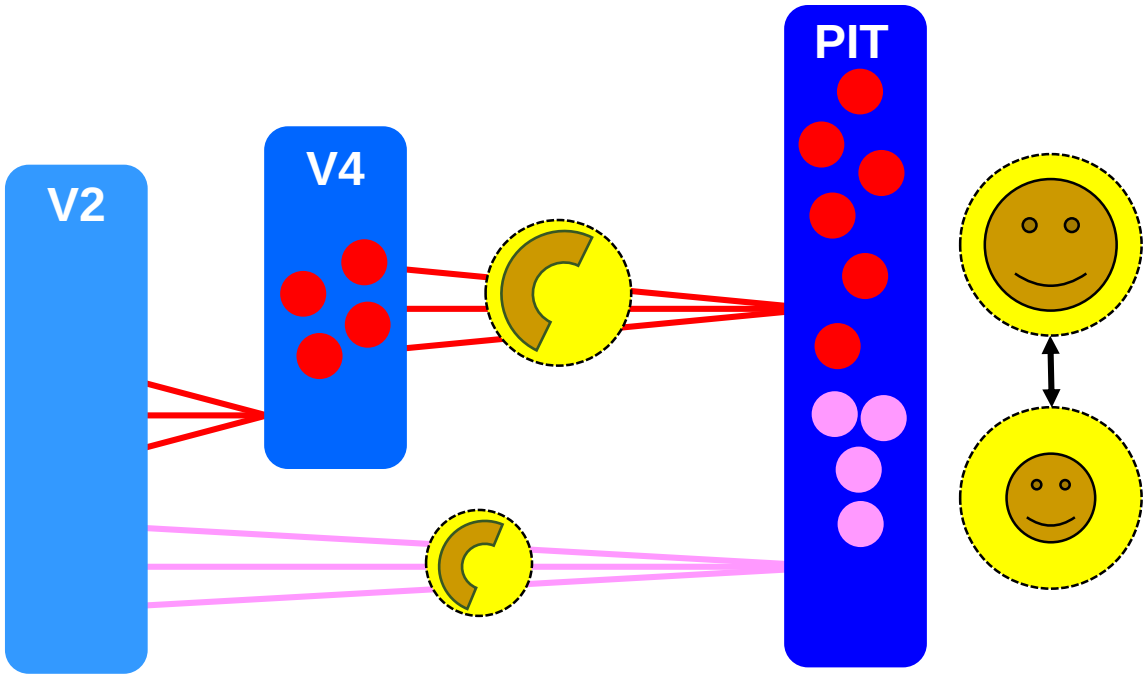
What is the most prominent difference between V2 and V4?







modified from Freeman and Simoncelli, 2011 (based on Gattass, Gross and Sandell, 1981)



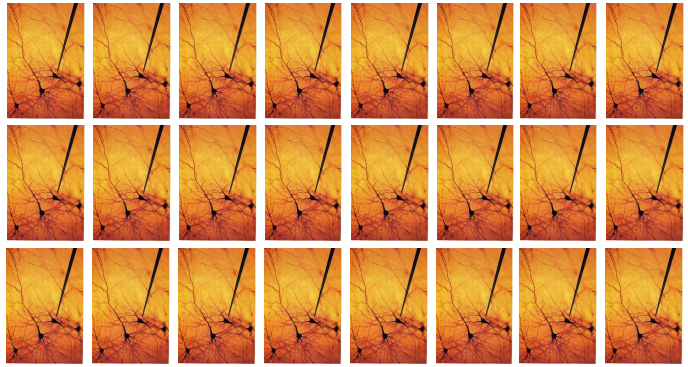
IT sites may use parallel pathways to keep their preferences across different scales (size invariance!)

To be determined!

Current investigations in IT (3): How do IT neurons encode information at the population level?

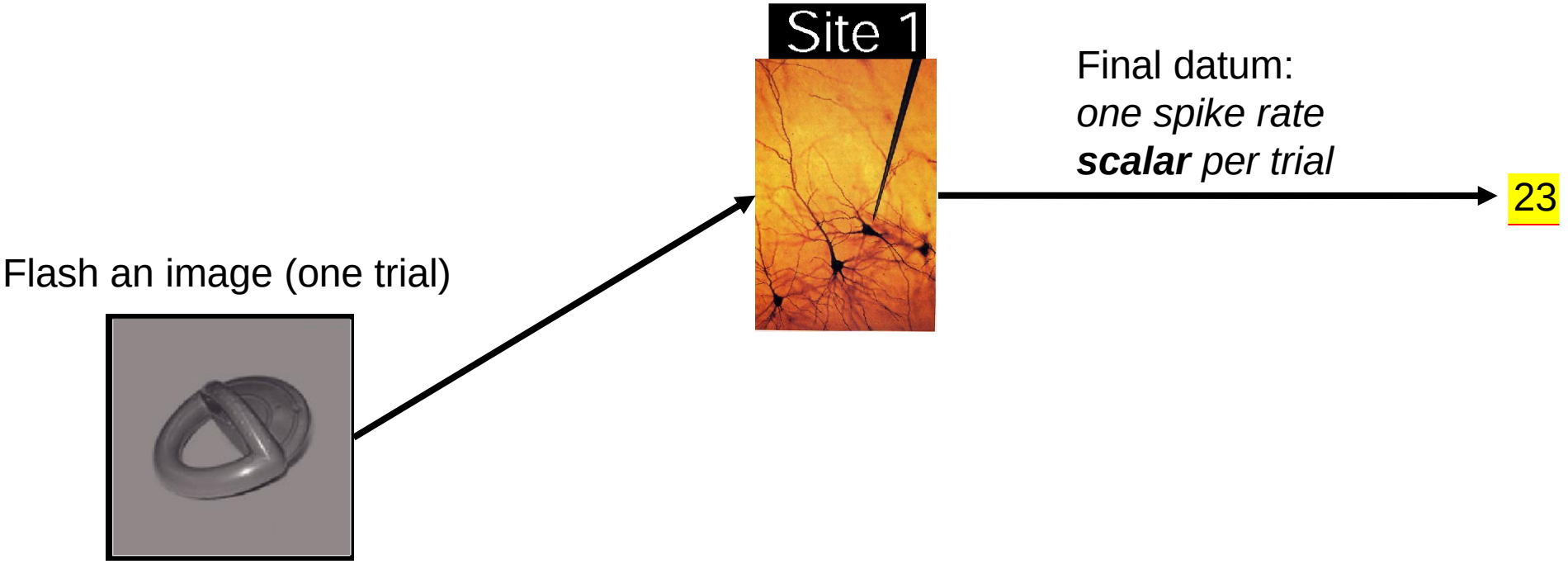
Intro to the paper discussion

Virtually all studies above were conducted using single-electrode experiments



What do we do when we have many, many electrodes?

In single-cell electrophysiology…



In single-cell electrophysiology…

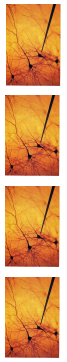
Flash an image (one trial)

Site 1

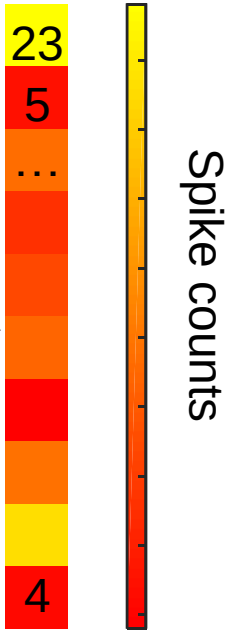
Final datum:
one spike rate
scalar *per trial*

23

Site 1
Site 2
Site 3
⋮

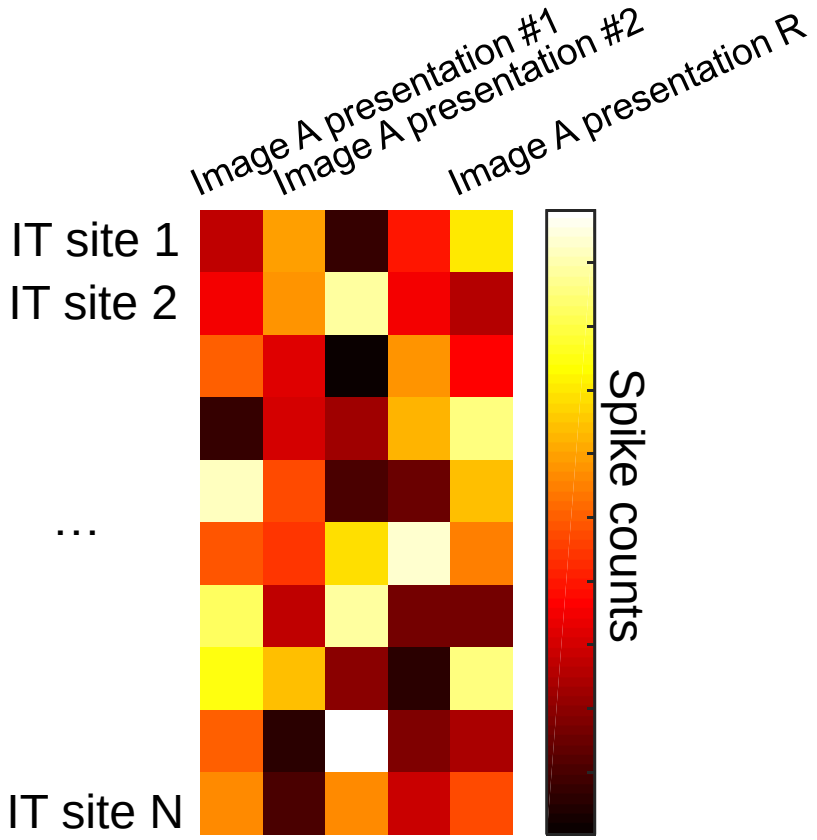


Final datum:
one spike rate
vector *per trial*

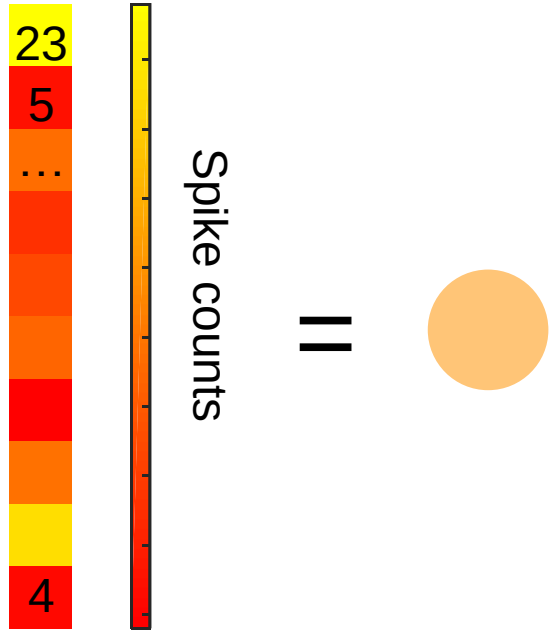


Spike counts

There are as many vectors as there are image flashes (presentations).

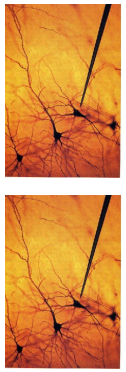


Think of each vector as a *point* in a coordinate space

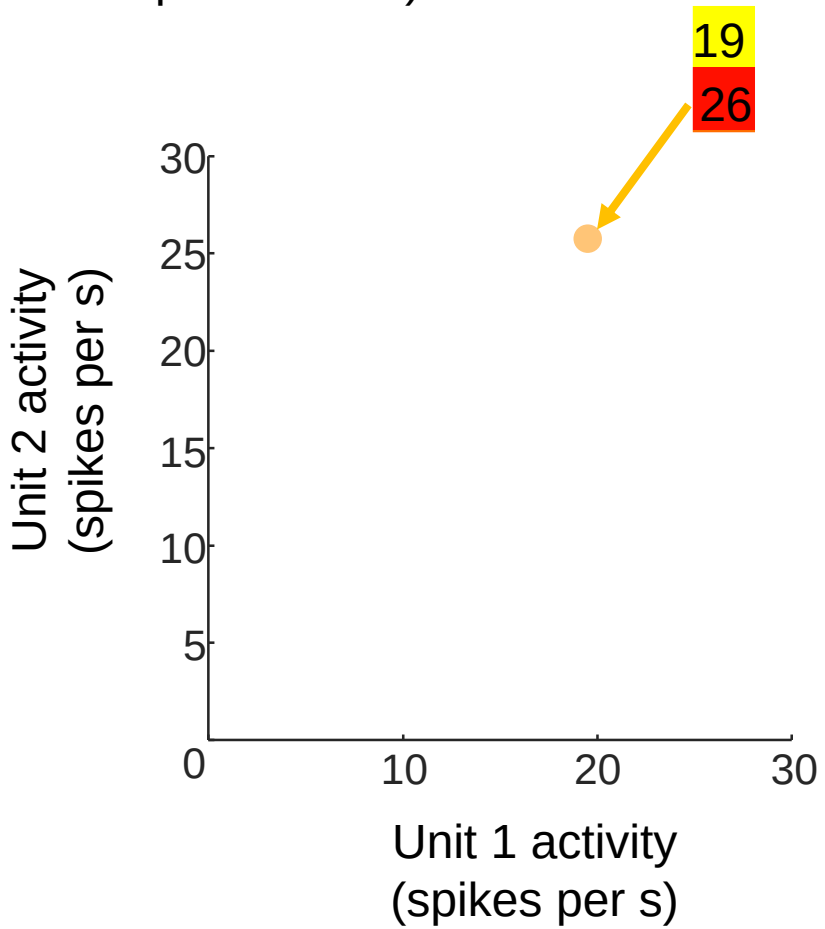


Imagine you have flashed image X

while recording from two cells concurrently

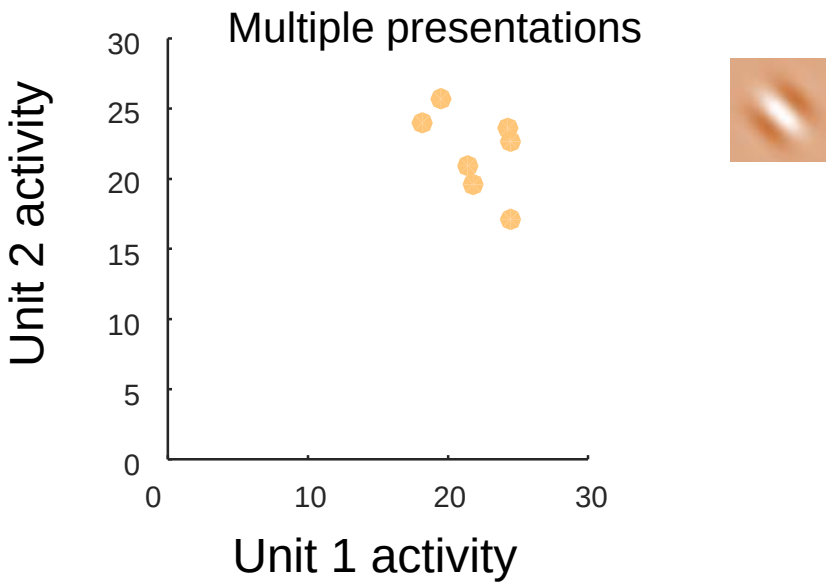


this results in response vector comprising two elements
(spike rate #1 and spike rate #2)

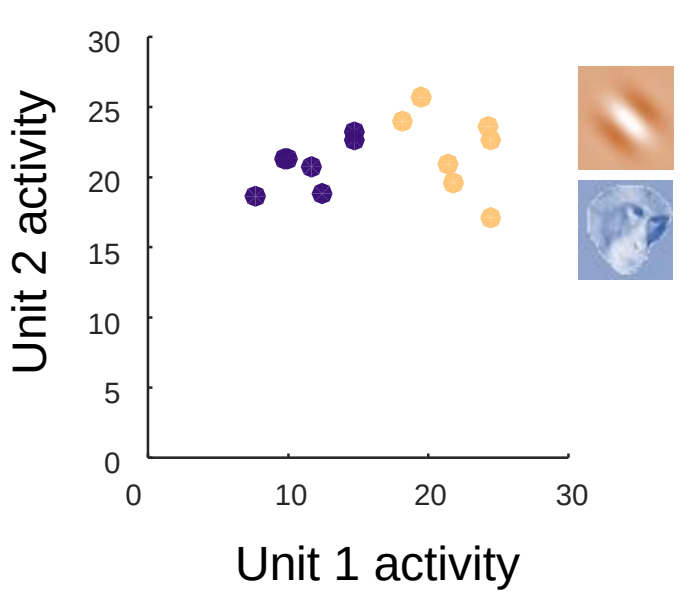


Response cloud for image 1

Multiple presentations

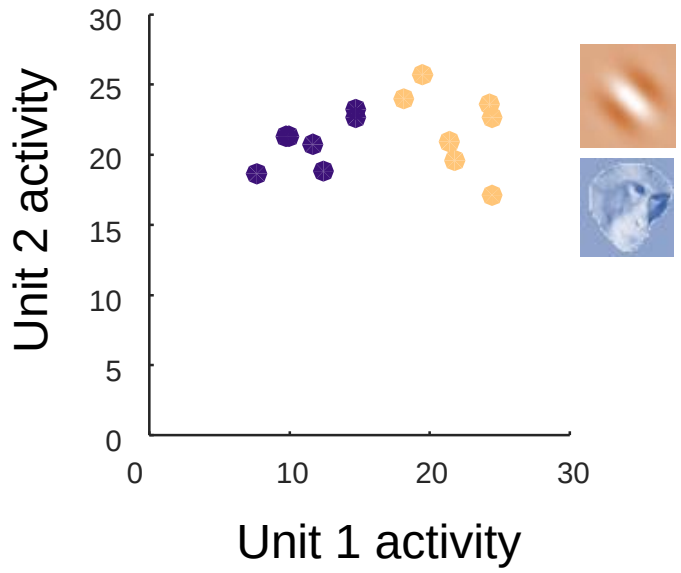


Response clouds for images 1 and 2

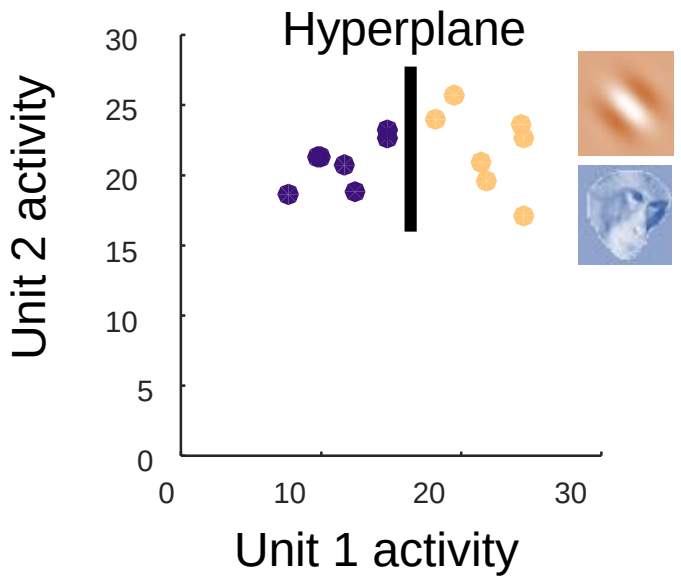


Different coordinate positions suggest separate representations in neural space

We need a statistic to tell us how separable these response clouds are in multi-dimensional space



One method to determine the separability of each cluster: statistical classifiers



Statistical classifier: a function that returns a binary value (“0” or “1”). These include rule-based classifiers, probabilistic classifiers, and geometric classifiers.

One example:

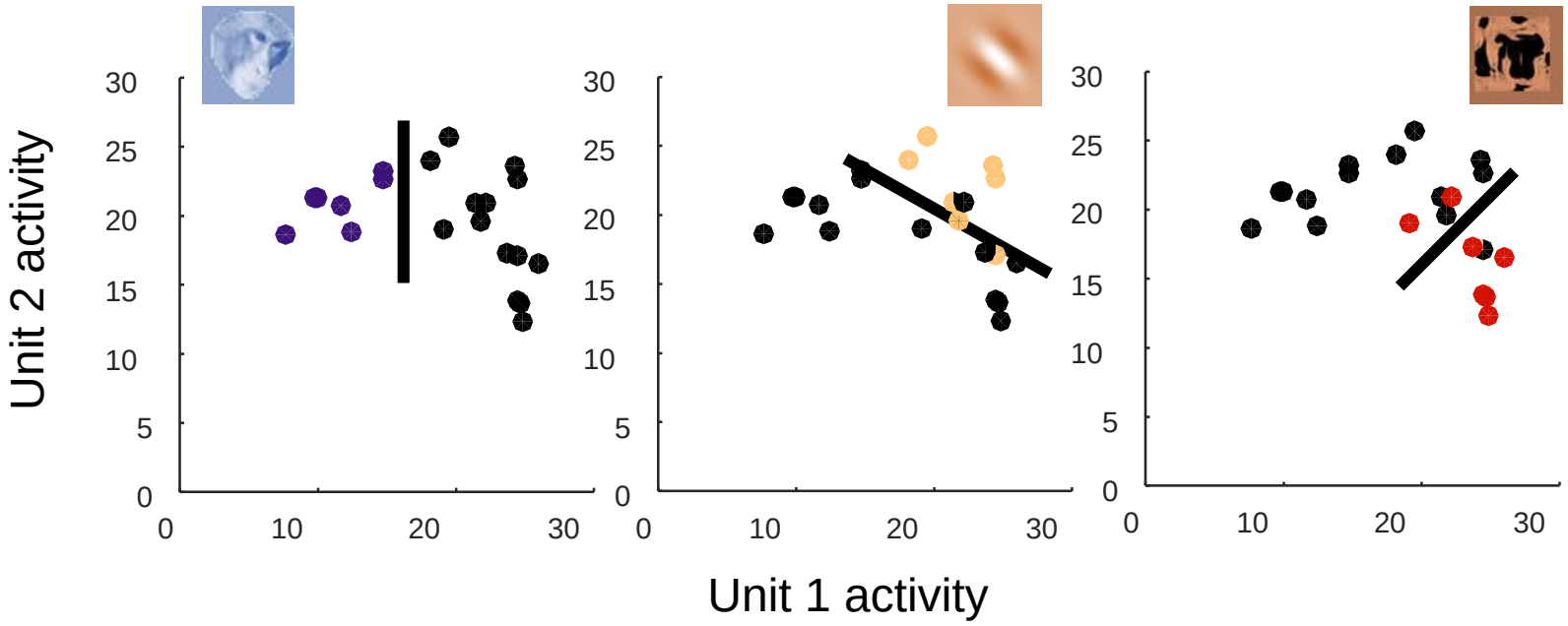
Support vector machines

-linear kernel

For a binary task, accuracy usually ranges between 50 and 100%

For multi-class classification, we can use a one-vs-all (aka one vs. rest) approach.

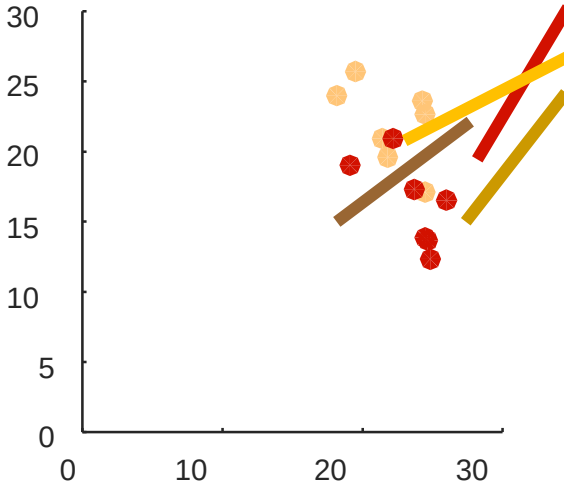
Label one category as positive, everything else as negative



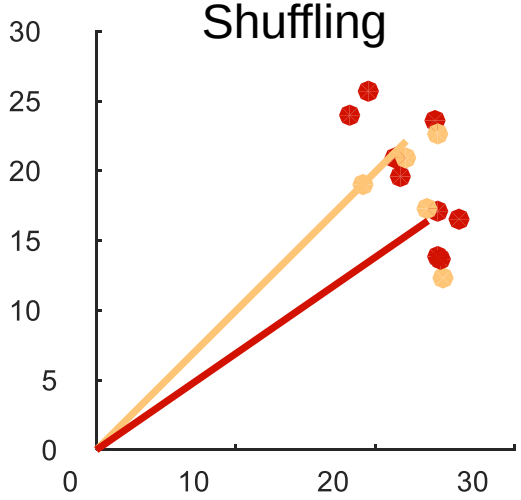
Test a new set of points, and identify which classifier gives the highest activation.

How do we define the statistical reliability of classification accuracy?

cross-validation



Shuffling



Randomly partition the data into subsets (90% for training, 10% for testing)

Repeat the procedure shuffling the class labels to check for accuracy bias.

Accuracy (correct labeling)
vs.
accuracy (shuffled labeling)

Now we have all we need to dig into the paper

Fast Readout of Object Identity from Macaque Inferior Temporal Cortex

Chou P. Hung,[1,2,4*†] Gabriel Kreiman,[1,2,3,4*] Tomaso Poggio,[1,2,3,4]
James J. DiCarlo[1,2,4]

Some of the papers mentioned in this lecture

1984 - Desimone, Albright, Gross and Bruce, Stimulus selective properties of IT neurons, JNeurosci

1992 - Sergent, Ohta and MacDonald, Functional neuroanatomy of face and object processing, Brain

1993 - Sary, Vogels and Orban, Cue invariant shape selectivity of macaque IT, Science

1994 - Kobatake and Tanaka, Neuronal selectivities to complex object features, J Neurophysiol

1995 - Ito, M., Tamura, H., Fujita, I., & Tanaka, K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol, 73(1), 218-226.

1995 - Logothetis, N. K., Pauls, J., & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. Current Biology, 5(5), 552-563.

1996 - Tanaka, K. Inferotemporal cortex and object vision. Annual Review of Neuroscience, 19, 109-139.

1996 - Logothetis, N. K., & Sheinberg, D. L. Visual object recognition. Annual Review of Neuroscience, 19, 577-621.

1997 – Kanwisher et al, The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception, JNeurosci.

1999 - Sugase et al. Global and fine information coded by single neurons in IT, Nature

2001 - Tsunoda et al. Complex objects represented in IT by the combination of feature columns, NN.pdf

2005 - Hung, C., Kreiman, G., Poggio, T., & DiCarlo, J. Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. Science, 310, 863-866

2005 - Quiroga, Reddy, Kreiman and Fried, Invariant visual representation by single neurons in the human brain, Nature

2006 - Brincat and Connor Dynamic shape synthesis in posterior IT, Neuron, Supp

2006 - Tsao et al. A cortical region consisting entirely of face-selective cells, Science

2007 - Kiani_Esteky_Mirpour_Tanaka, Object Category Structure IT with Supp

2009 - Liu H, Agam Y, Madsen J, Kreiman G. Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62:281-290

2010 - Freiwald and Tsao, Functional Compartmentalization and Viewpoint, Science

2012 - Markov et al, A weighted and directed interareal connectivity matrix for macaque cerebral cortex, Cerebral Cortex

2013 - Markov et al, Cortical high-density counterstream architectures, Science