Neurobiology HMS 130/230 Harvard / GSAS 78454

Visual object recognition: From computational and biological mechanisms

Today's meeting: Early Steps into Inferotemporal Cortex

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

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

crponce@gmail.com

Agenda

A brief recap: what you have seen so far in the course.

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

Lecture parts:

The anatomy of IT
What do IT cells encode? ("selectivity")
How good are they when contextual noise is introduced? ("invariance")
How do we use machine learning techniques to decode information in IT responses?

Paper discussion

A brief recap: tell us about one important fact you learned in…

Lecture 1: 09/12/16. Why is vision difficult? Natural image statistics and the retina

Lecture 2: 09/19/16. Lesions and neurological examination of extrastriate visual cortex.

Lecture 3: 09/26/16. Psychophysical studies of visual object recognition. (Olson)

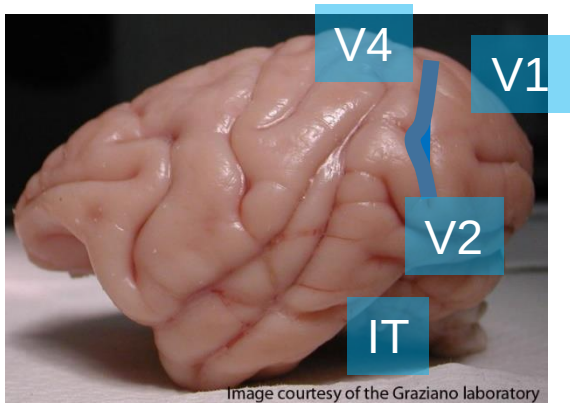
Lecture 4: 10/03/16. Primary visual cortex. (Gomez-Laberge)

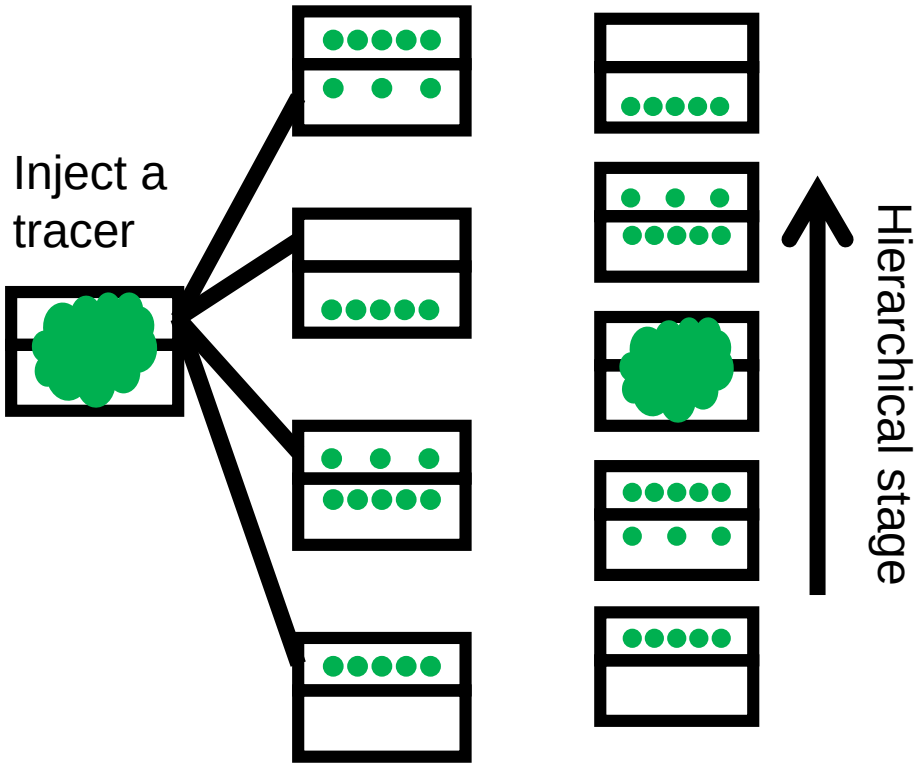
Lecture 5: 10/17/16. Adventures into terra incognita: probing the neurophysiological responses along the ventral visual stream. (Kim)

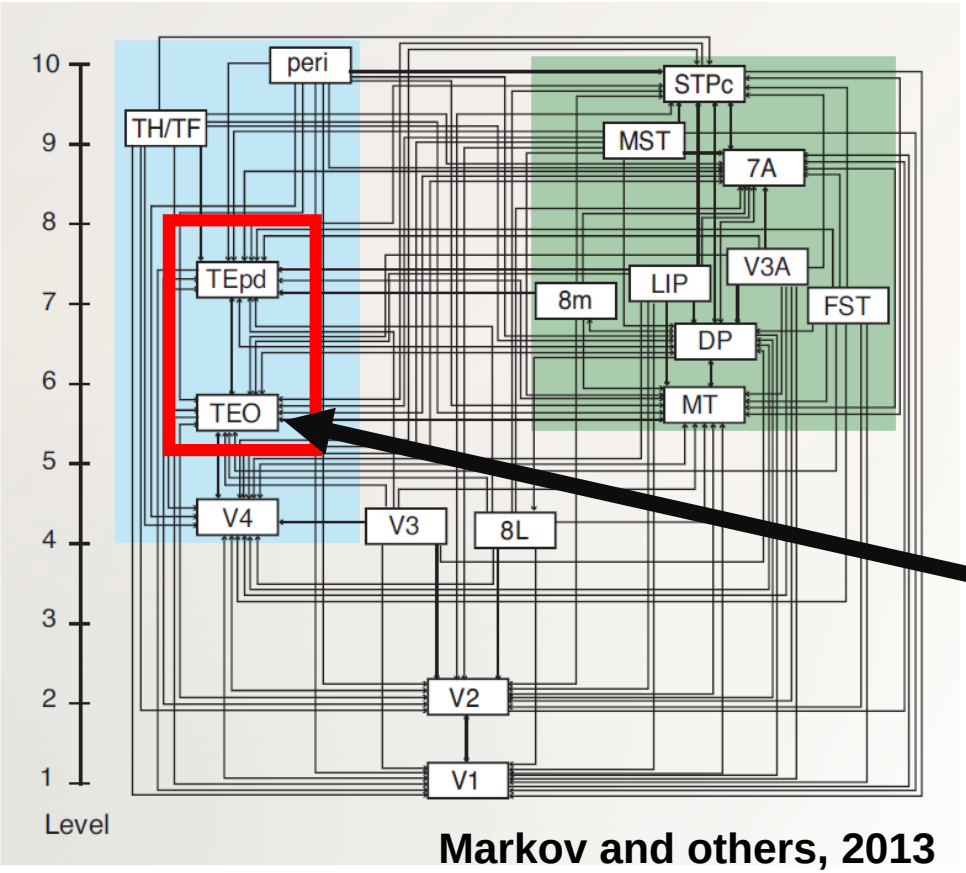
Review of key fact (from last lecture): The visual system is hierarchical

We know this because 1) neurons respond with different latencies to the onset of a flash (LGN cells respond faster than V1, V1 than V2, and so on)

2) Cortical areas show laminar patterns that suggest directionality.





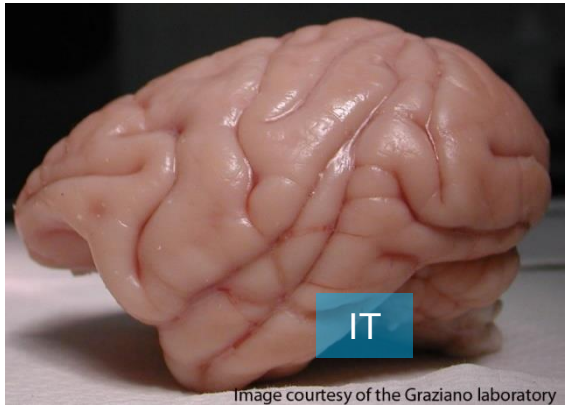


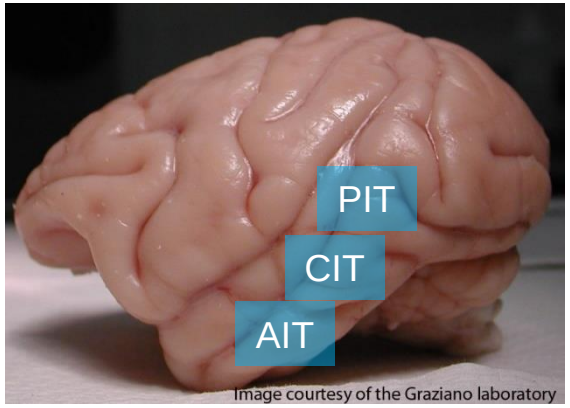
"TEO?"

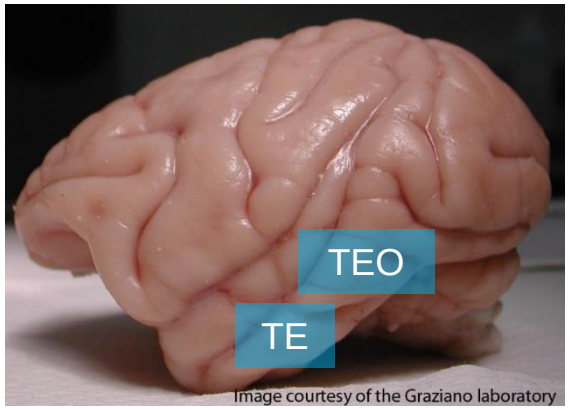
Markov and others, 2013

The anatomy of inferotemporal cortex: input projections

IT goes by many names

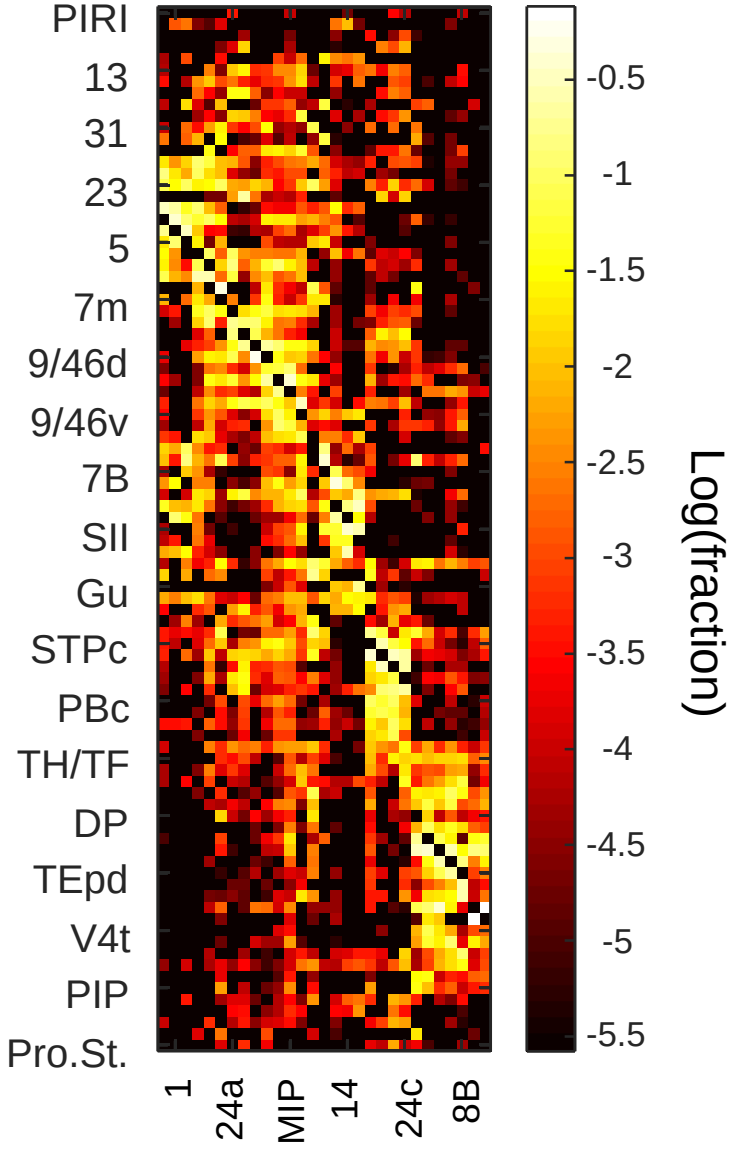






What other brain areas talk to IT?

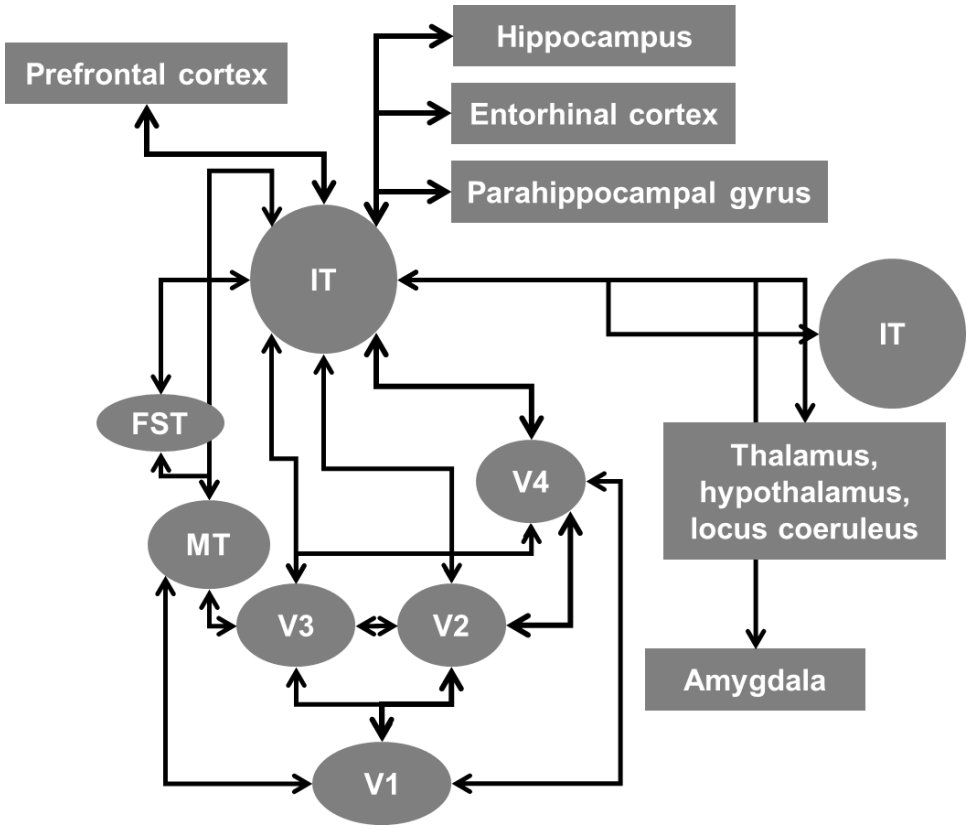
There are weight maps showing the number of cells that project from each area to another.



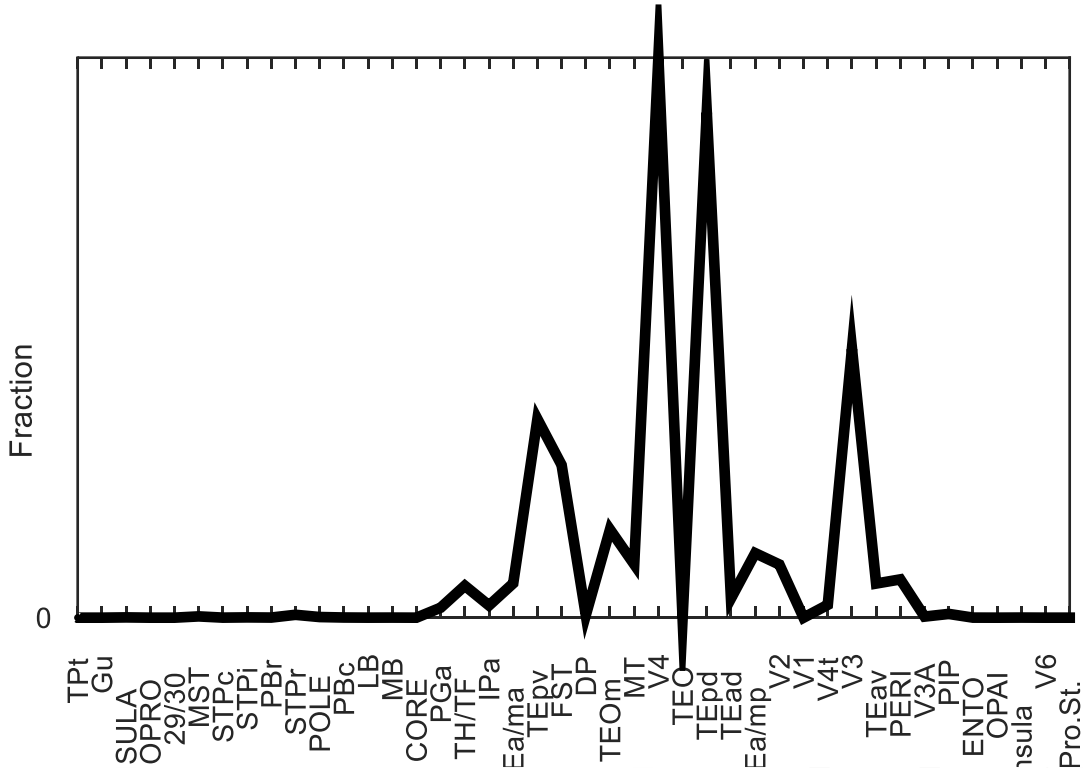
Adapted from Markov et al 2012

The anatomy of inferotemporal cortex: projections

Many areas project to IT.

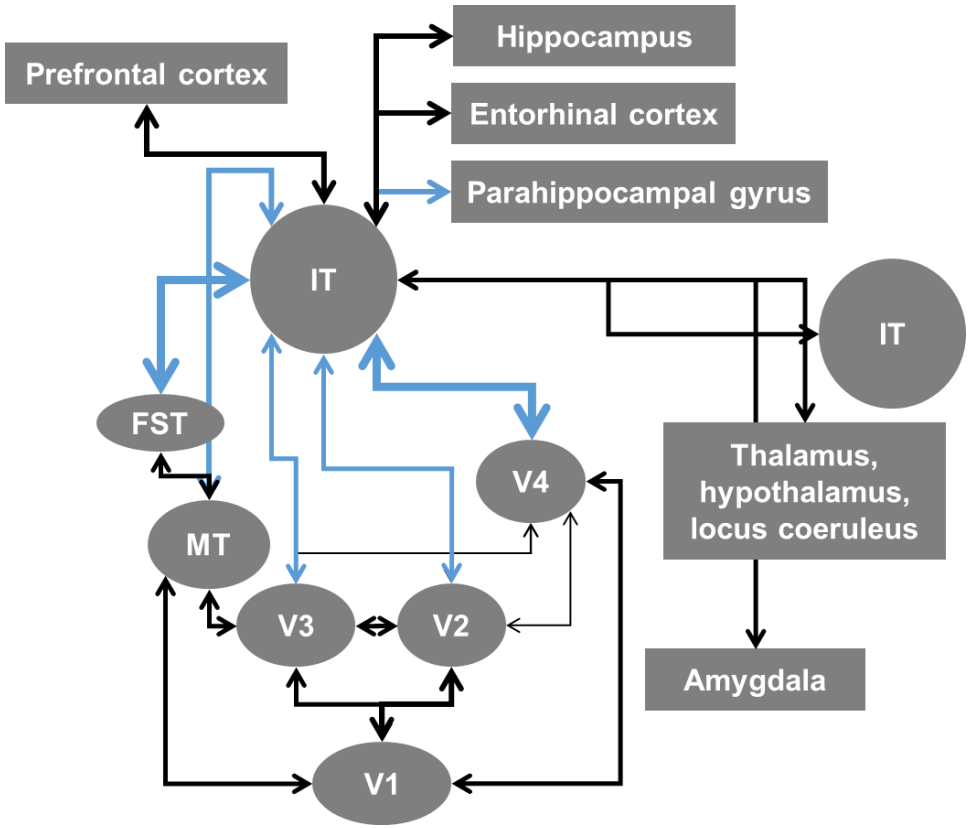


Relative weights of posterior IT inputs

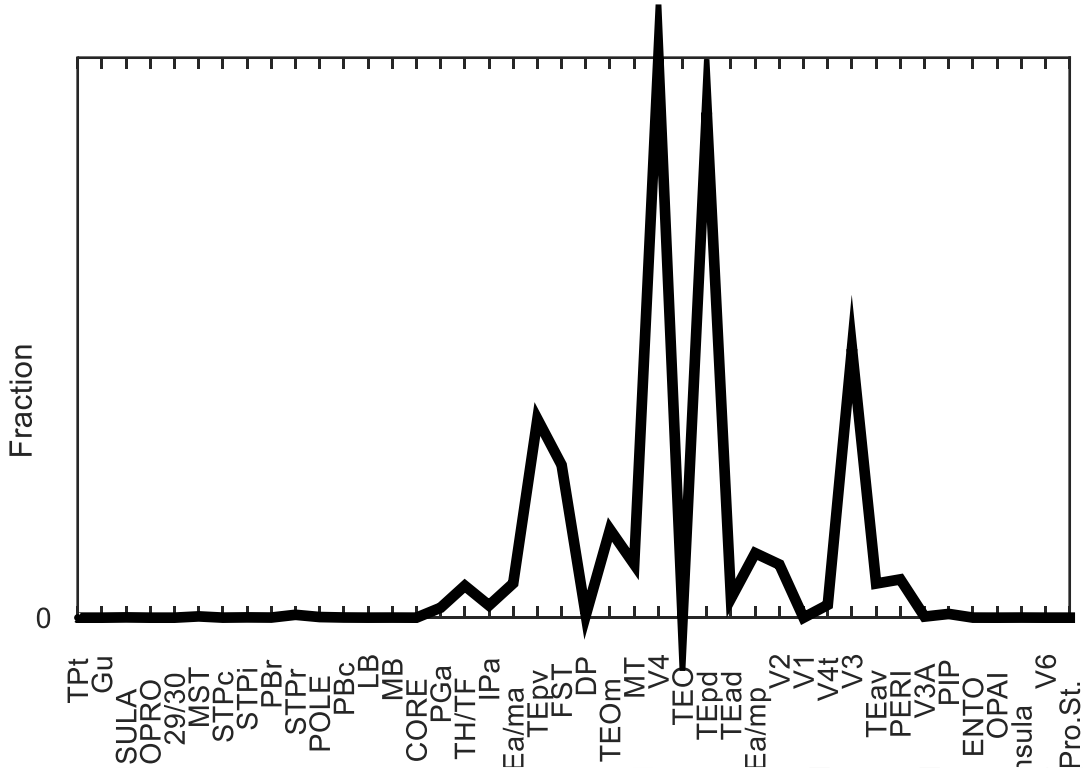


Markov and others, 2013

The anatomy of inferotemporal cortex: projections

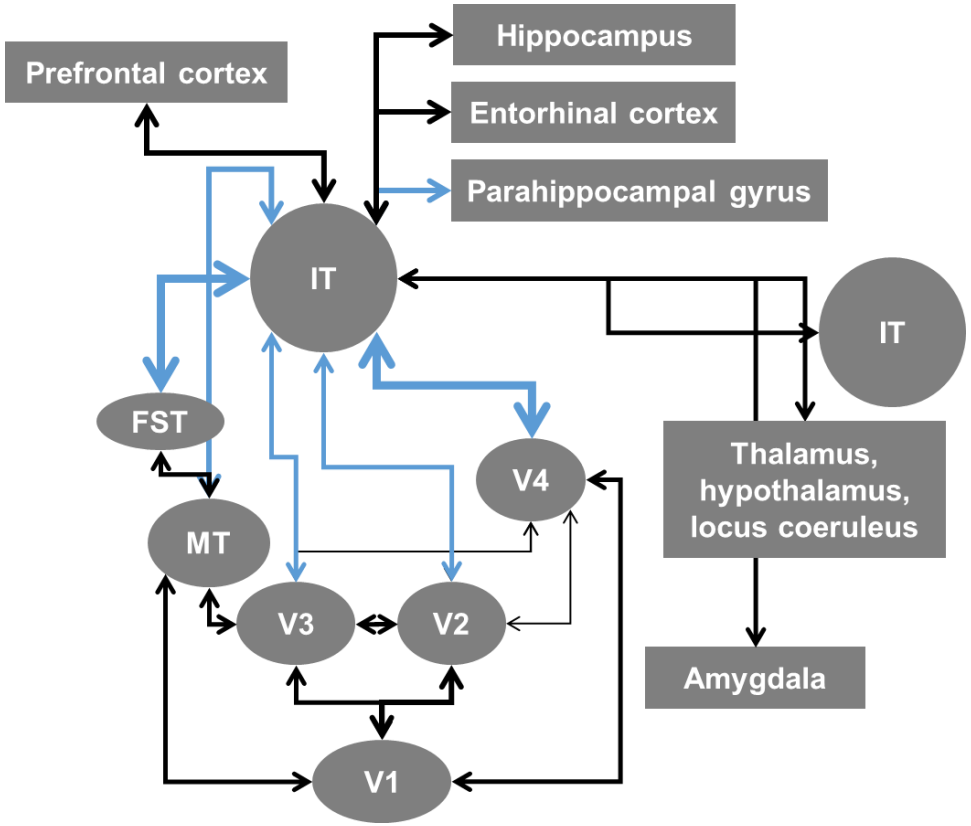


Relative weights of posterior IT inputs



The anatomy of inferotemporal cortex: subdivisions

Visual information about objects continues to be transmitted to other parts of the brain



IT is interesting because it is the last exclusively visual area in the hierarchy

Some investigators have subdivided IT into many subareas.

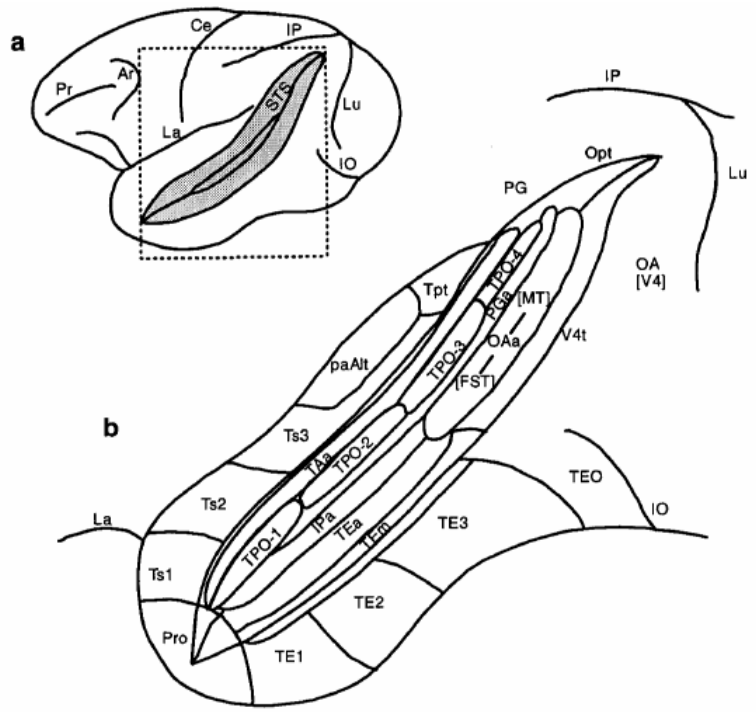


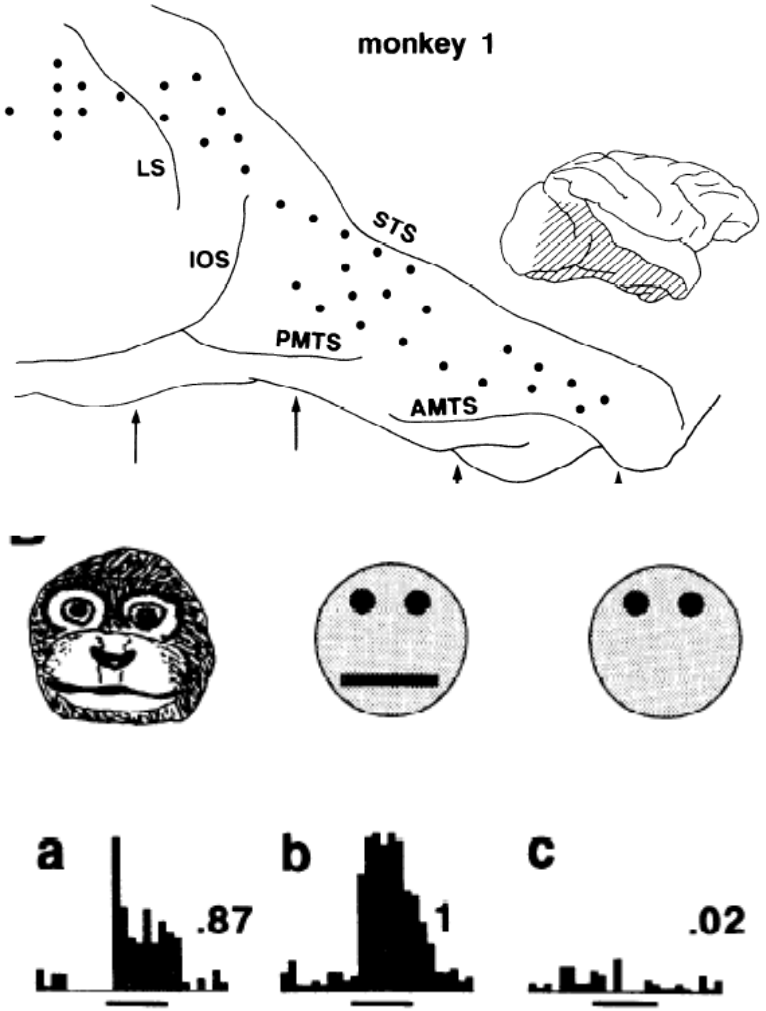
Figure 3 Subdivision of monkey inferior temporal lobe centered around the superior temporal sulcus (STS). (*a*) Lateral view of the cortical surface with major visible sulci labeled: inferior occipital (IO), lunate (Lu), intraparietal (IP), central (Ce), lateral (Sylvian) fissure (La), arcuate (Ar), and principal (Pr). (*b*) Expanded view of the inferior temporal areas surrounding the STS. [Adapted from Seltzer & Pandya (1994).]

Logothetis and Sheinberg. Annual Review of Neuroscience 1996

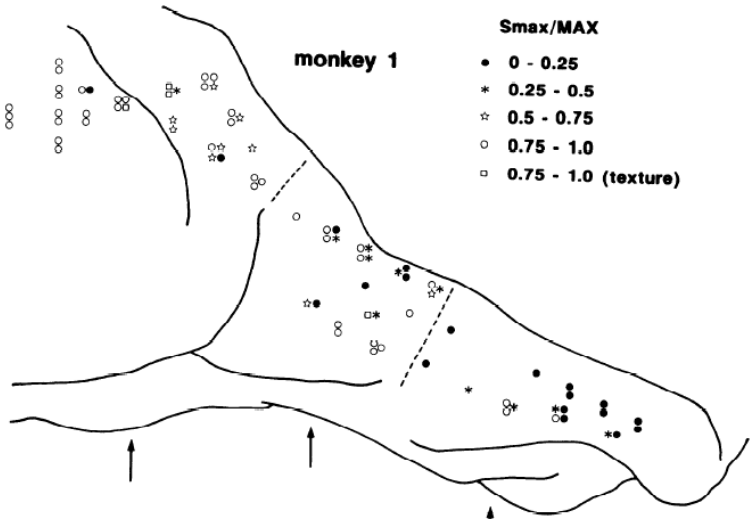
In practice, most of these subdivisions have no specific theoretical roles.

We can think of **IT** as a stream

EUCALY KOBATAKE AND KEIJI TANAKA



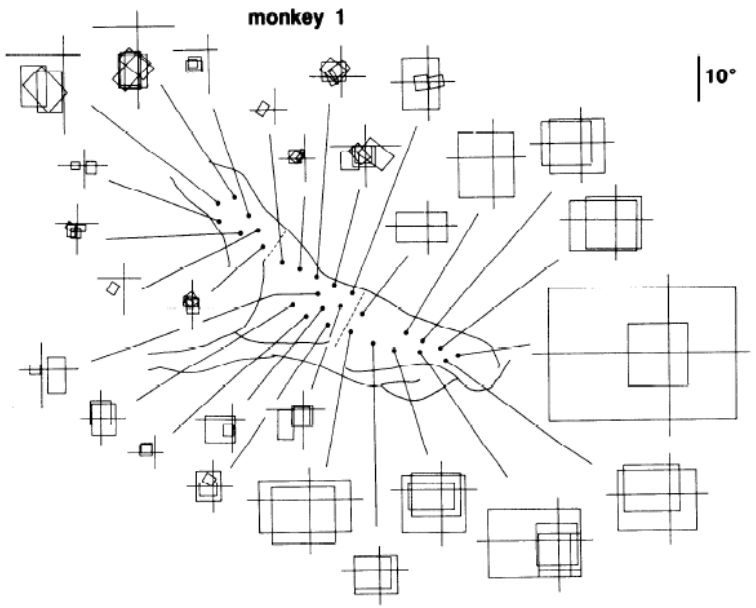
At each site, they measured the number of spikes emitted to individual features vs. combinations of multiple features



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

We can think of **IT** as a stream

EUCALY KOBATAKE AND KEIJI TANAKA

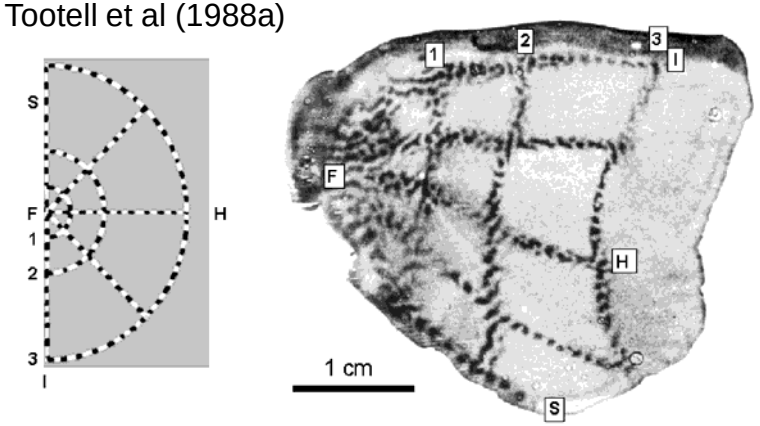


IT cells closer to V1 (more posterior) have smaller receptive fields.

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

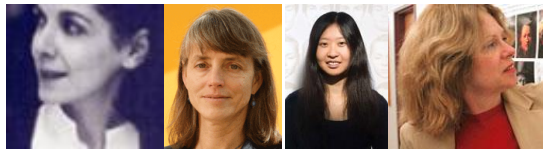
Retinotopy: cells physically near one another respond to parts of the visual field that are also near each other

Tootell et al (1988a)



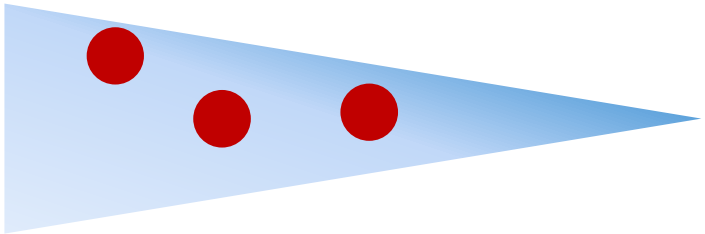
IT cells further from V1 show less and less retinotopy, organizing themselves by feature preference.

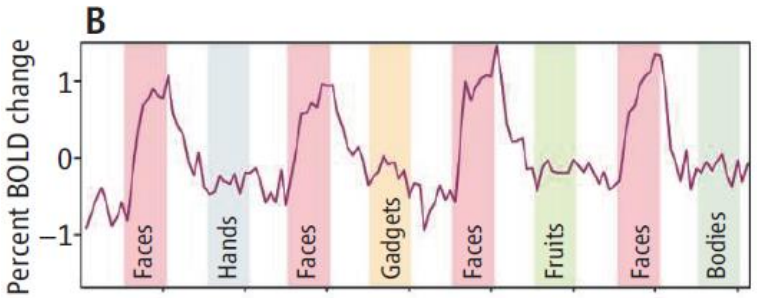
...a stream with interesting cobblestones

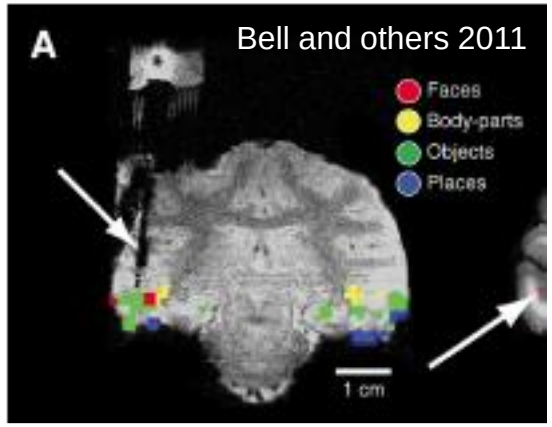


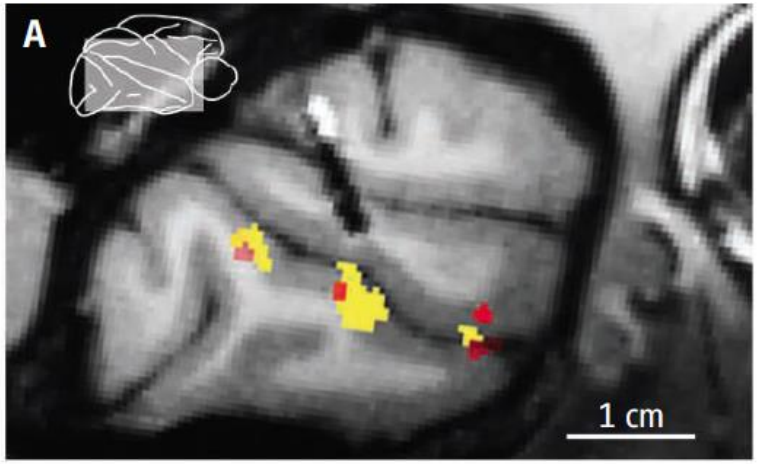
Sergent
Kanwisher
Tsao
Livingstone
Freiwald

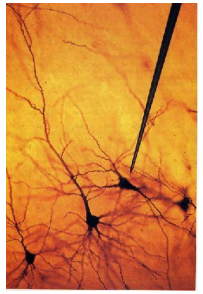
IT contains clusters (“patches”) selective for common ecological categories.

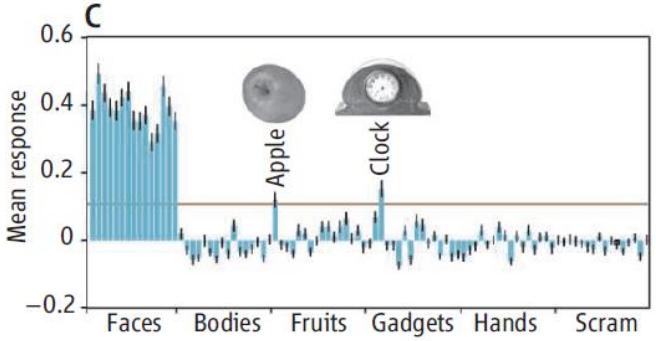












IT cells can band into subnetworks for special tasks

End of anatomy section –

Any questions so far?

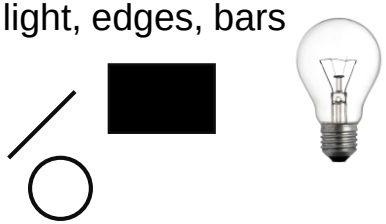
Selectivity

Let us take a closer look at the
preferences of individual cells

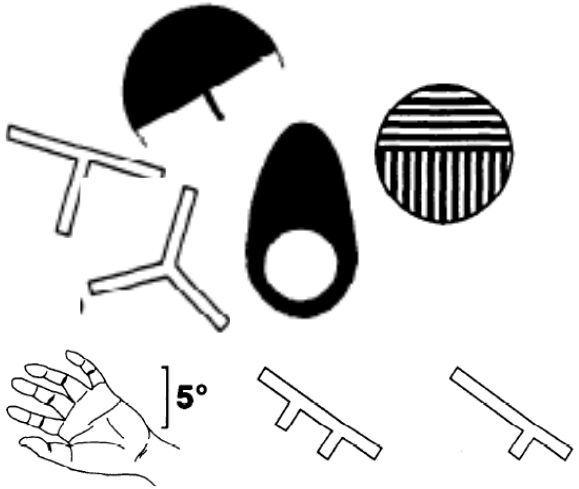
2006: Connor and others

A sample of visual stimuli historically used to stimulate IT cells

1965: Gross: Diffuse light, edges, bars

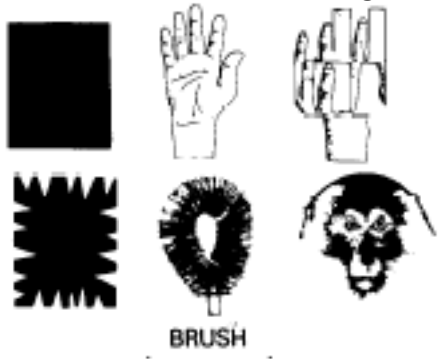


1991: Tanaka, Saito, Fukada and Moriya



2005 - Hung, Kreiman, Poggio and DiCarlo

1984: Desimone, Albright, Gross and Bruce

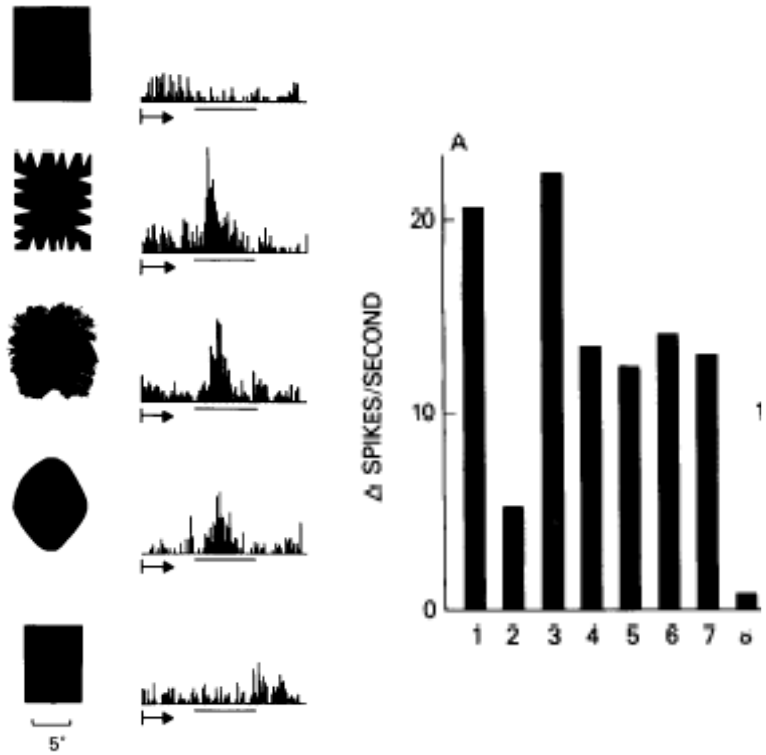


2007: Kiani, Esteky, Mirpour and Tanaka

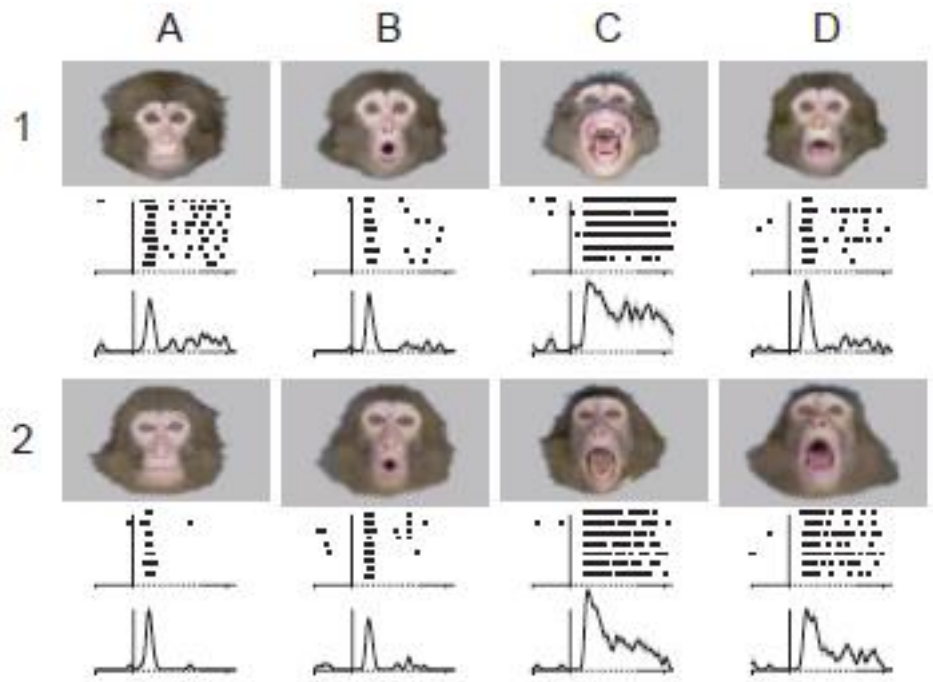
1995: Logothetis, Pauls and Poggio

How do cells express “preferences”?

IT cells emit different number of **action potentials (“spikes”)** in response to different images...



They can be sensitive to small differences in the **same** object.



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

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

Global and fine information coded by single neurons in the temporal visual cortex

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

A historical side note

A tentative approach to complex visual preferences

(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 cutout of a monkey hand, and the more the stimulus looked like a hand, the more strongly the unit responded to it.

Gross et al started with simple stimuli, and eventually moved onto complex stimuli (fingers, burning Q-tips, brushes) to elicit *attention*

Jerry Konorski (1967) proposes "gnostic" units – cells that represented "unitary perceptions." Suggests that they live in IT.

"When we wrote the first draft...we did not have the nerve to include the 'hand' cell until Teuber urged us to do so."

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

Charles G. Gross

How Inferior Temporal Cortex Became a Visual Area

Cells with similar preferences cluster together at different scales

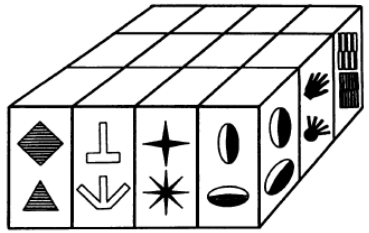
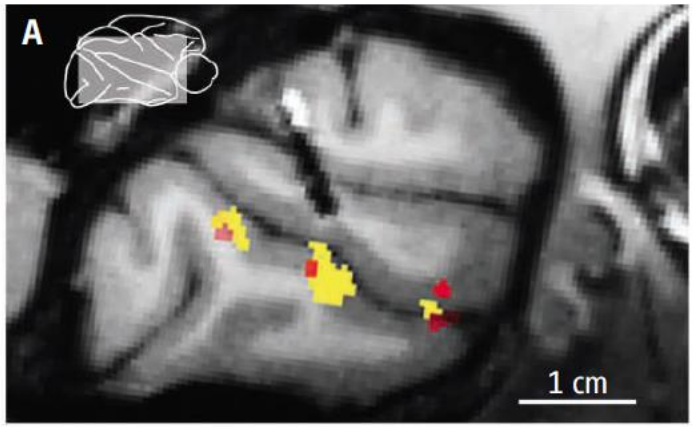


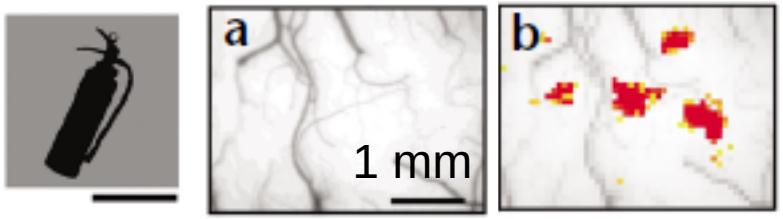
Fig. 3. Schematic diagram of the columnar organization in TE. Cells with similar but slightly different selectivity cluster in elongated vertical columns, perpendicular to the cortical surface.

Clusters can range from several mm...



(visible in fMRI)

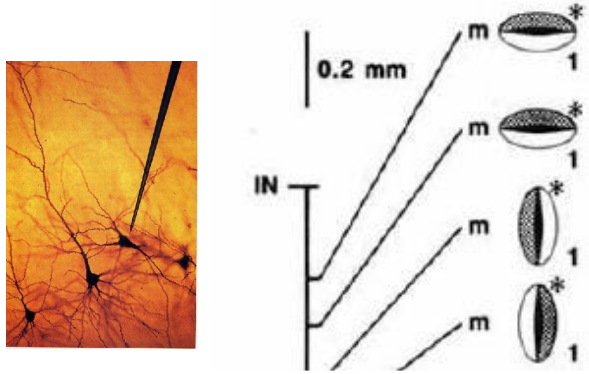
...to scales around 1 mm...



Tsunoda et al 2001

(visible with intrinsic imaging techniques)

...to scales best measured in micrometers.



Fujita et al 1992

(evident with electrophysiology)

Developing preferences for a given object is one problem that IT cells need to solve. There is one trivial solution: develop fixed templates. What is the problem with this?

Imagine you are a new human with a developing IT cortex

Some cells could imprint their RFs to this view of mom's face

Next time mom comes back, context may be a little different

The previously imprinted RFs would not provide a compelling match.

http://thephotobrigade.com

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!)

What type of common variations should IT be ready to handle?

Position

Size

Viewpoint

Illumination

Occlusion

Texture

What else?

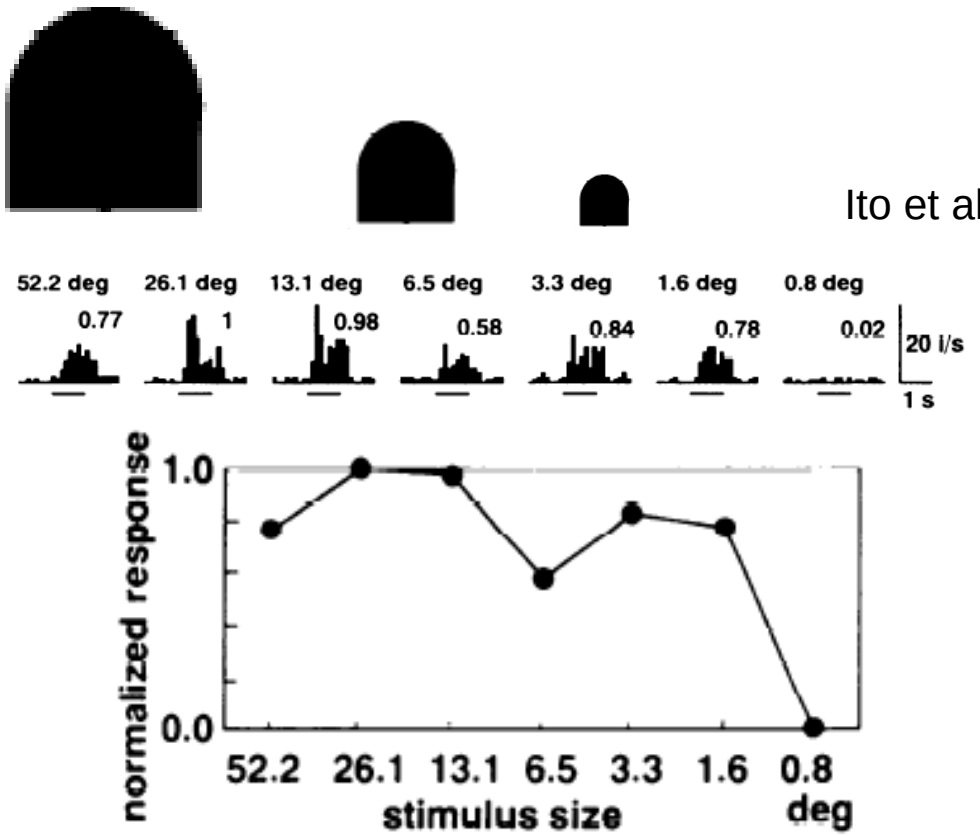
IT neurons can respond to their preferred shapes despite these changes.

This is called “invariance” or “tolerance.”

Let’s review some of the evidence.

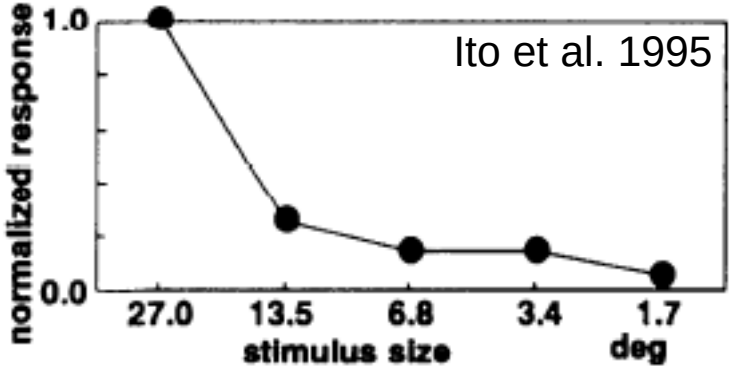
Size invariance

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



Ito et al. 1995

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

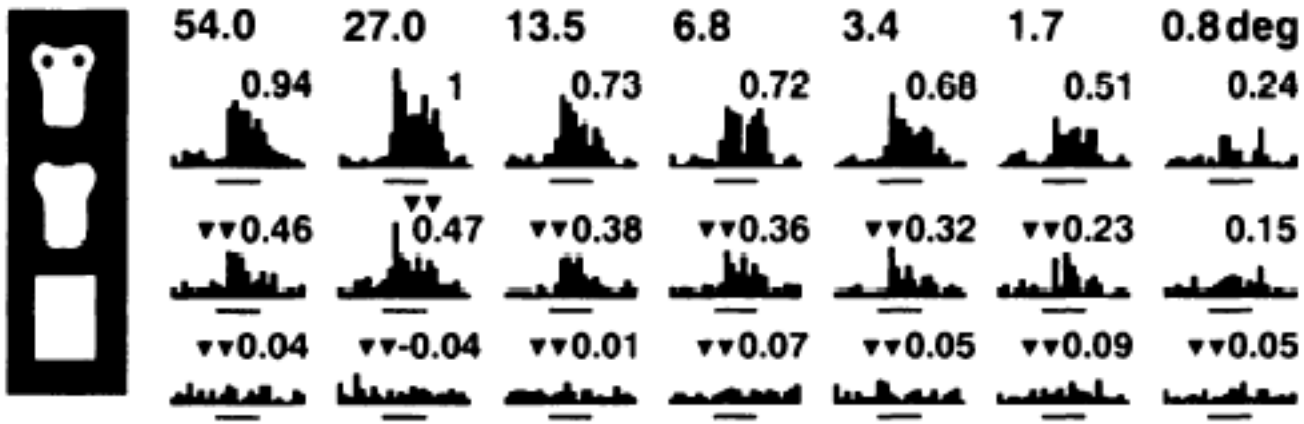


Ito et al. 1995

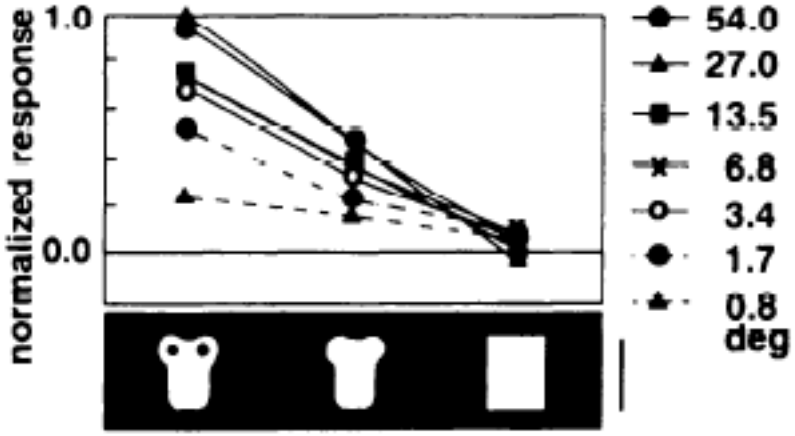
Most of the time, they vary their responses.

Size invariance

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



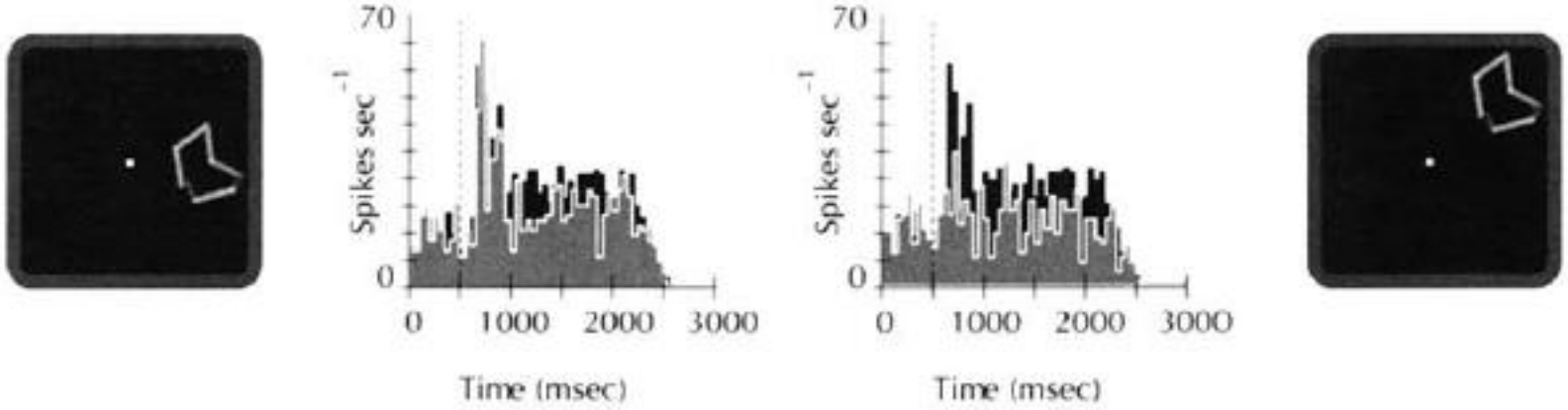
Ito et al. 1995



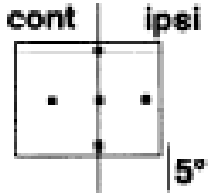
This neuron shows the same relative preference *despite* size changes.

Position invariance

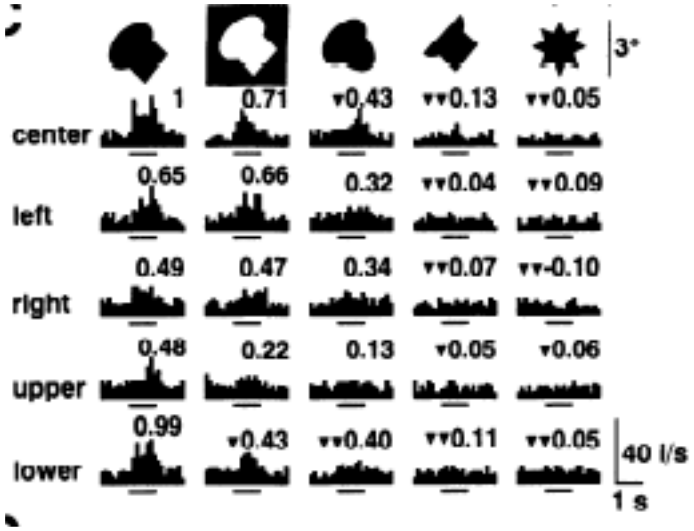
Logothetis et al, 1995

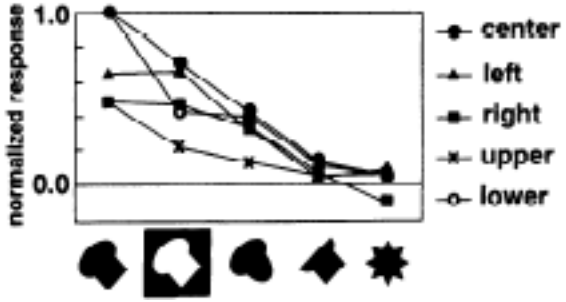


This neuron shows the same firing rate activity AND relative preference despite position changes.



Ito et al. 1995





This neuron shows the same relative preference despite position changes.

Texture invariance

Position

Size

Viewpoint

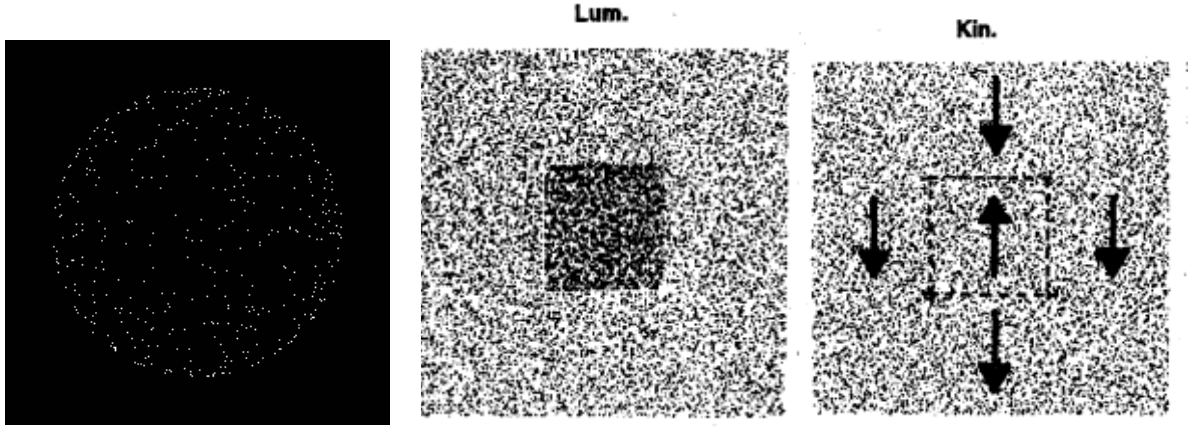
Illumination

Occlusion

Texture

What else?

Visual shapes can be described by simple luminance changes, or by second-order features (motion, textures)



Sary, Vogels and Orban 1993

Texture invariance

Position

Size

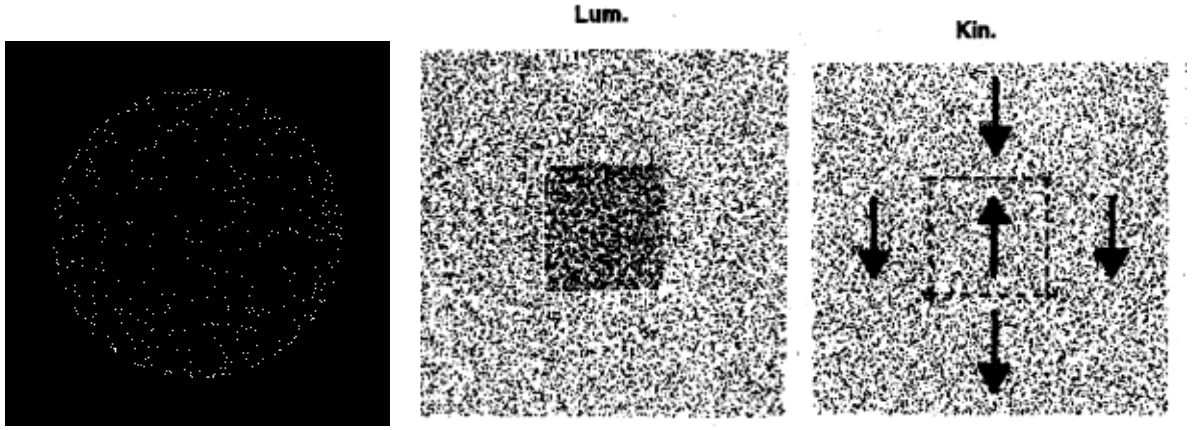
Viewpoint

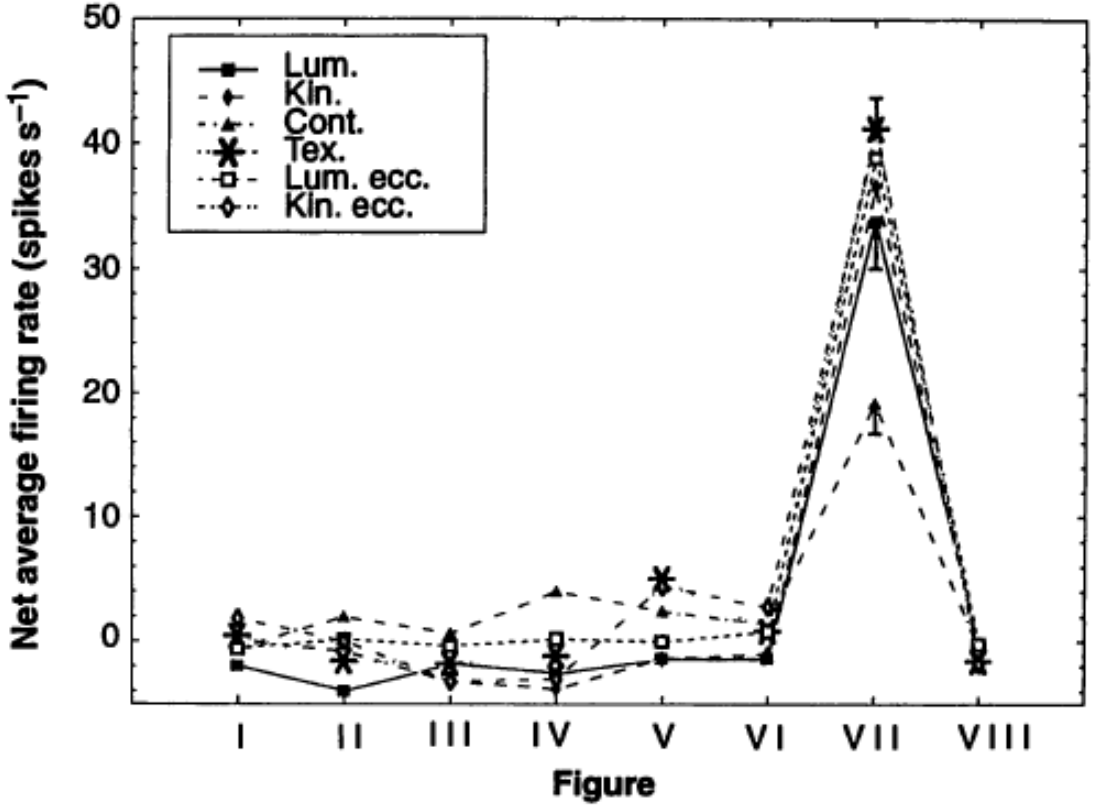
Illumination

Occlusion

Texture

What else?





Sary, Vogels and Orban 1993

Position

Size

Viewpoint

Illumination

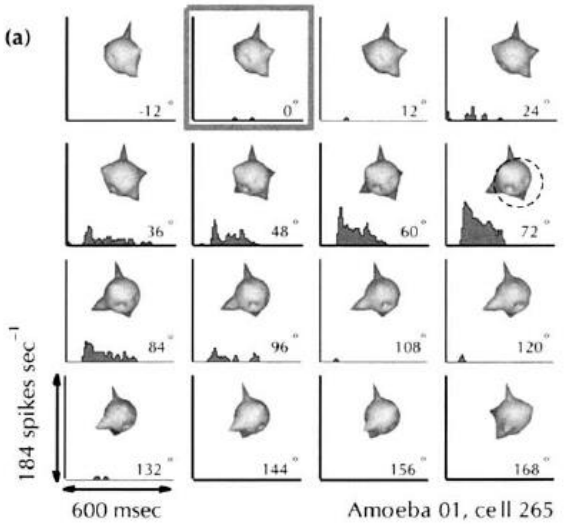
Occlusion

Texture

What else?

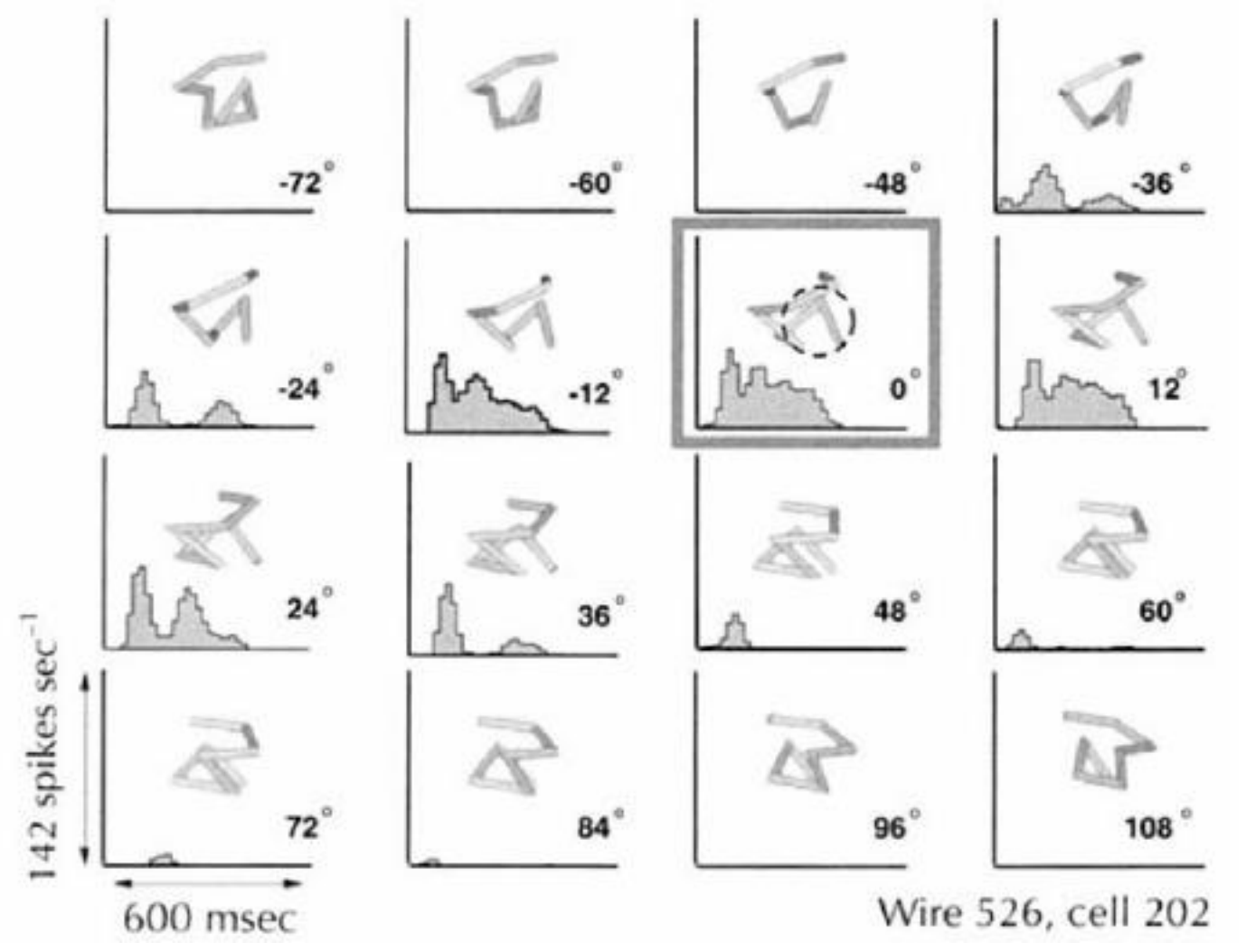
Examples of images used to test viewpoint invariance

Desimone and others, 1984

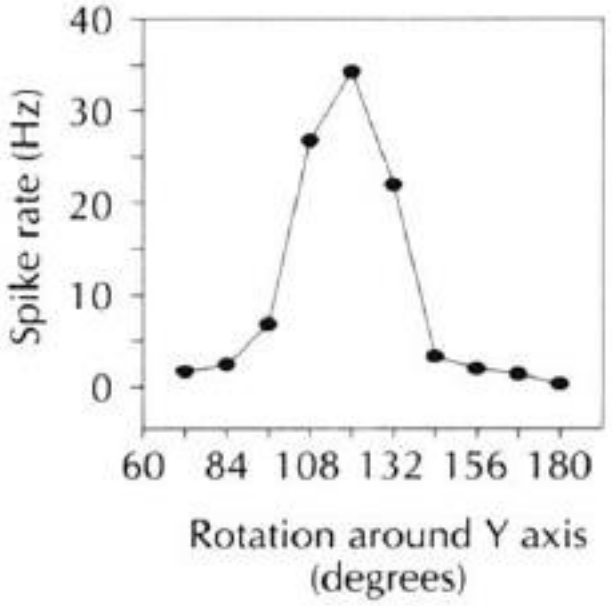


Logothetis and others, 1995

Viewpoint invariance



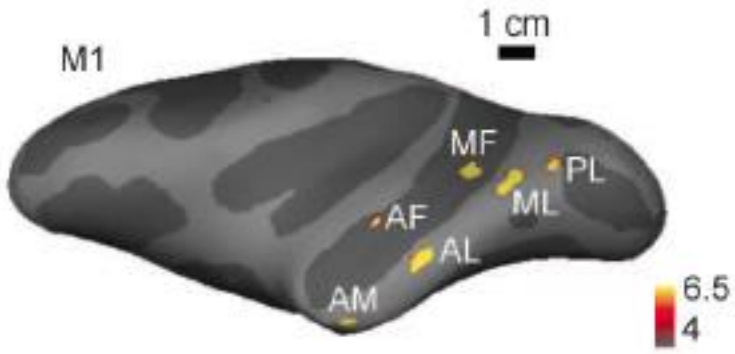
IT neurons view tuning curves have widths of ~ 30° rotation

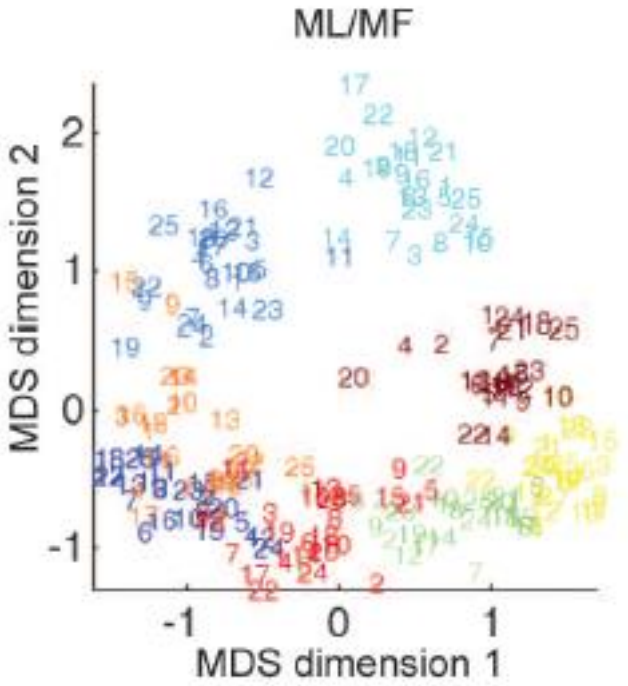


Logothetis and others, 1995

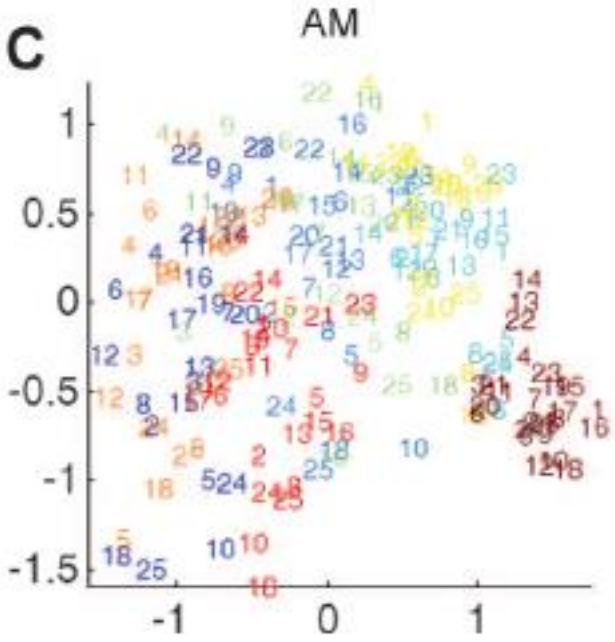
Viewpoint invariance

The face network develops viewpoint invariance along its patches.





Patch ML clusters the faces of different individuals by viewpoint.



Patch AM clusters the faces of different individuals by identity.

Freiwald and Tsao 2010

Lecture parts:

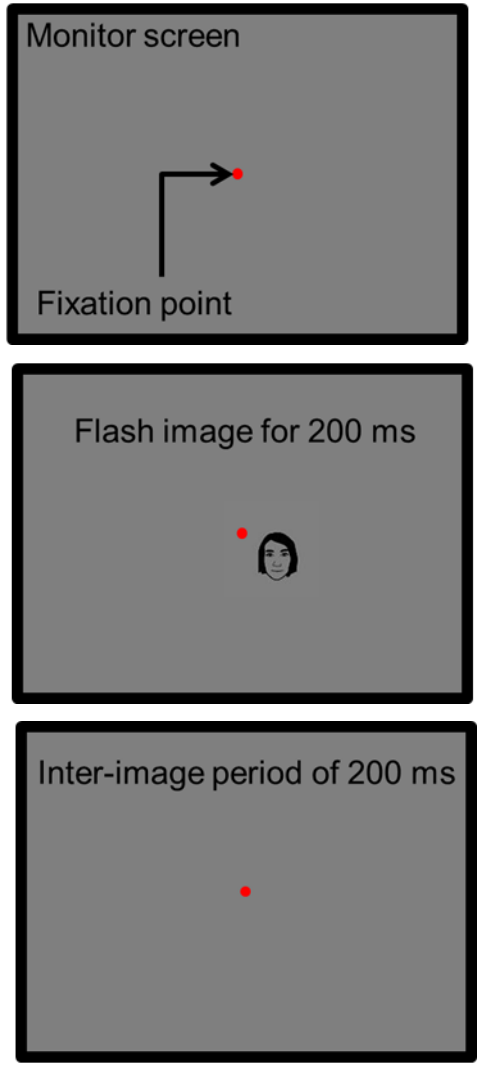
The anatomy of IT

What do IT cells encode? (“selectivity”)

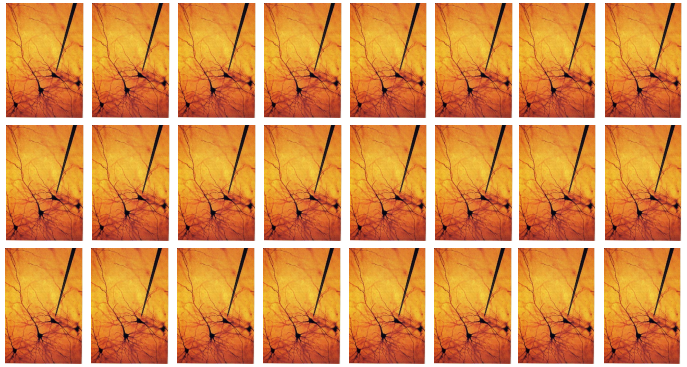
How good are they when contextual noise is introduced? (“tolerance/invariance”)

How we use machine learning techniques to decode information in IT responses

Decoding information from IT populations

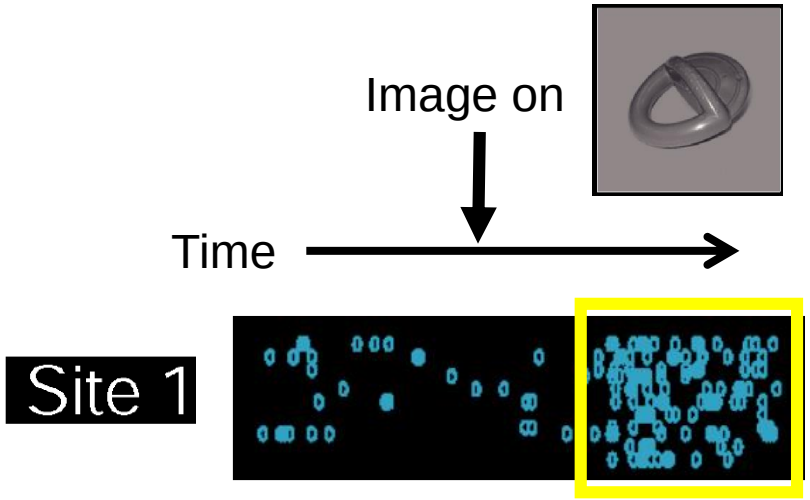


Virtually all studies above were conducted using single-electrode experiments



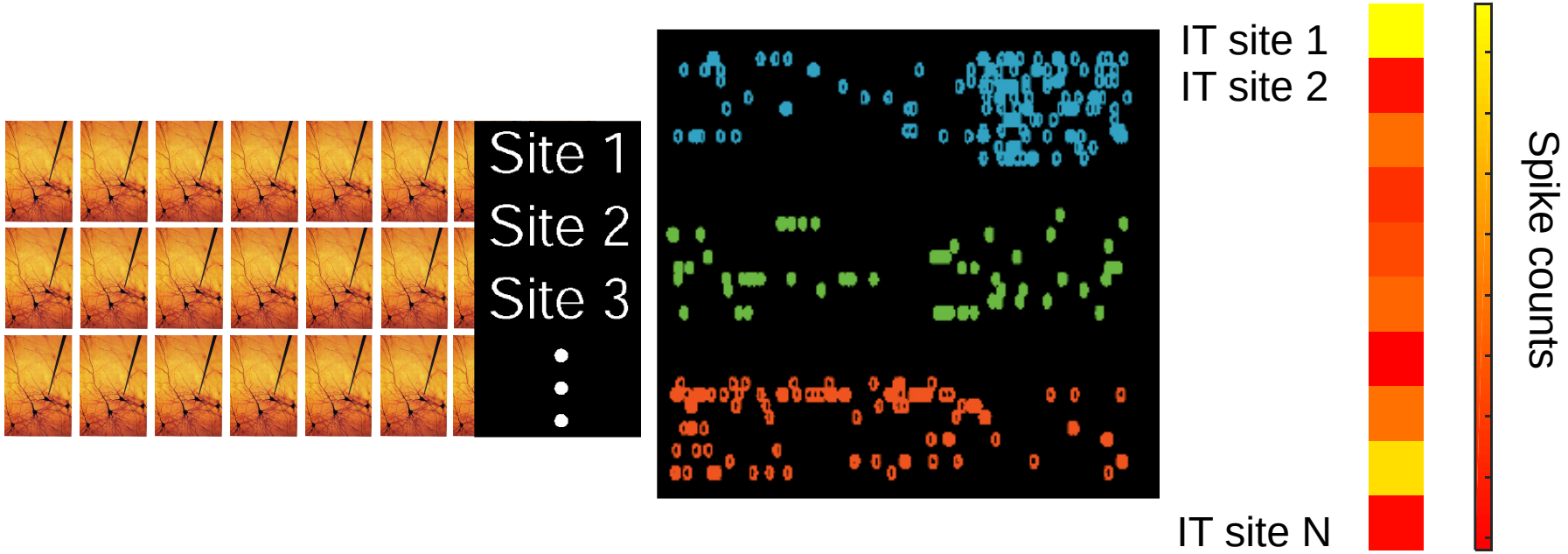
What do we do when we have many, many electrodes?

Firing rates: from scalars to vectors



For each trial: average / time = spikes per s

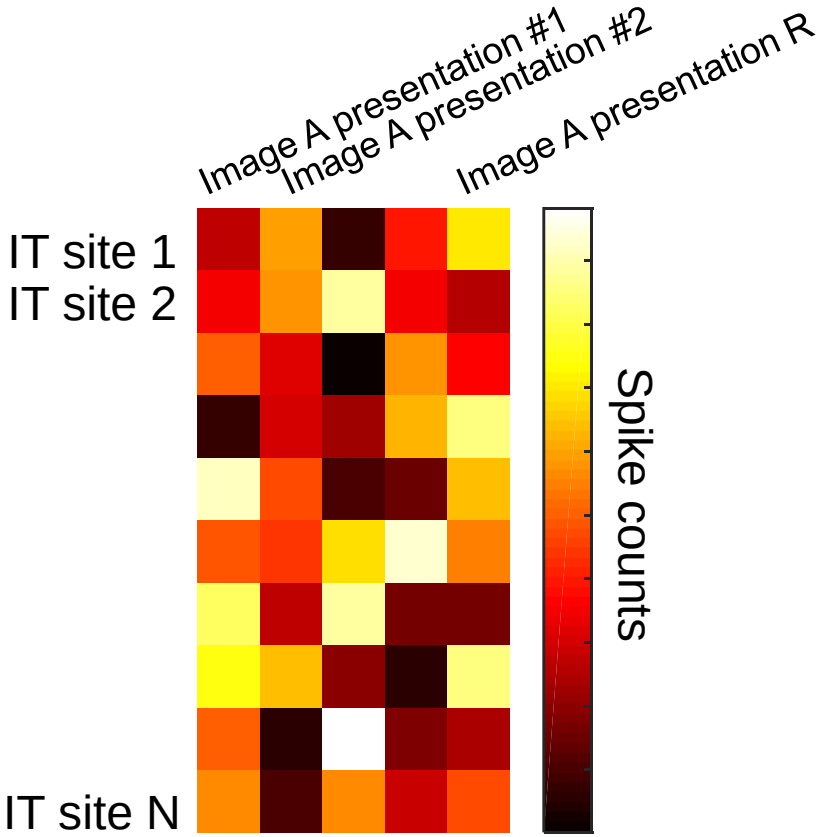
Final datum: *one spike rate per trial*



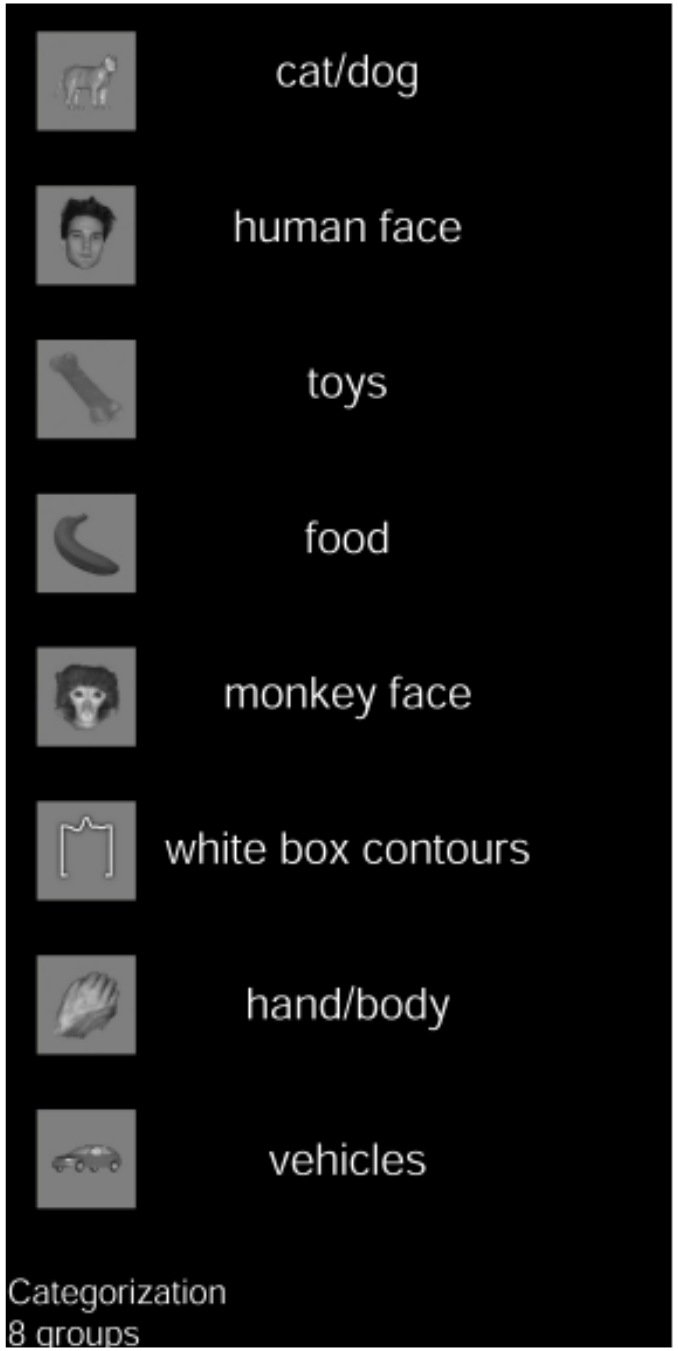
Final datum: *one spike rate vector per trial.*

Chou Hung, Gabriel Kreiman, Jim DiCarlo, Tomaso Poggio

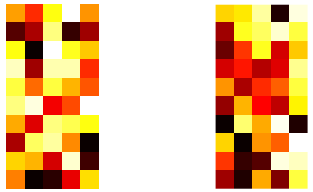
There are as many vectors as there are image presentations.



There are as many matrices as there are categories / individual images.



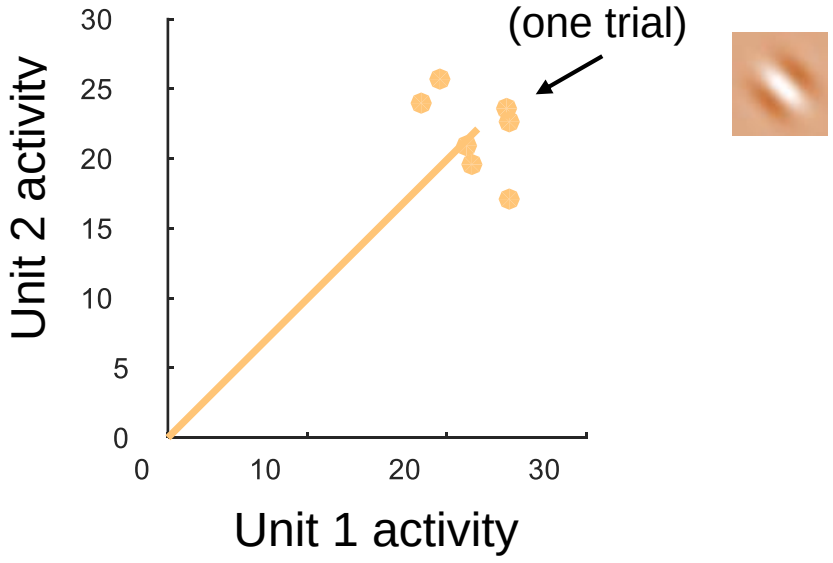
How did we decode information across all response matrices?



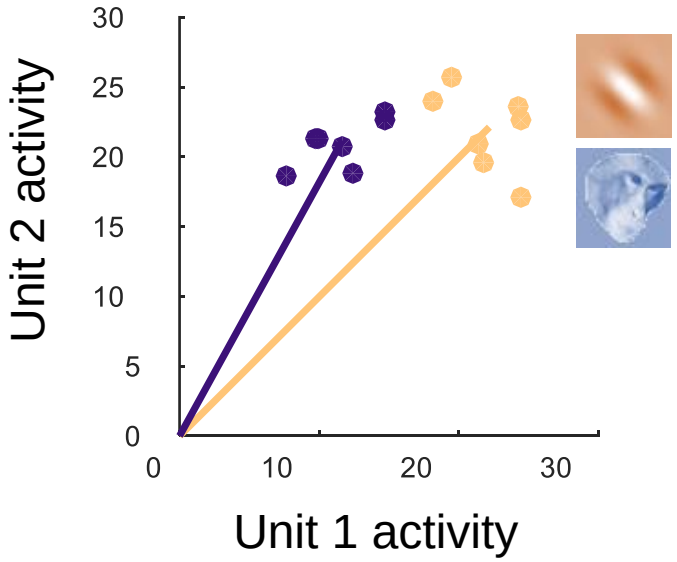
Think of each vector as a *point* in a coordinate space

(Let's simplify and imagine that the number of elements in the vector is 2)

Response cloud for image 1

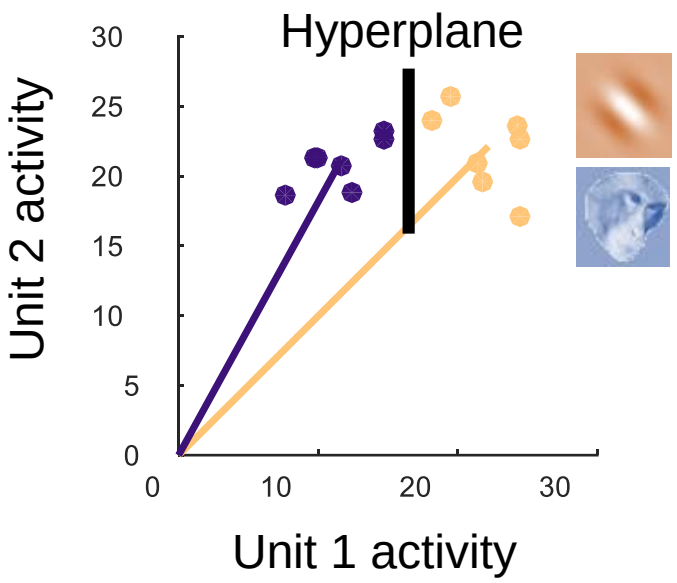


Response clouds for images 1 and 2



Different coordinate positions suggest differential encoding.

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:

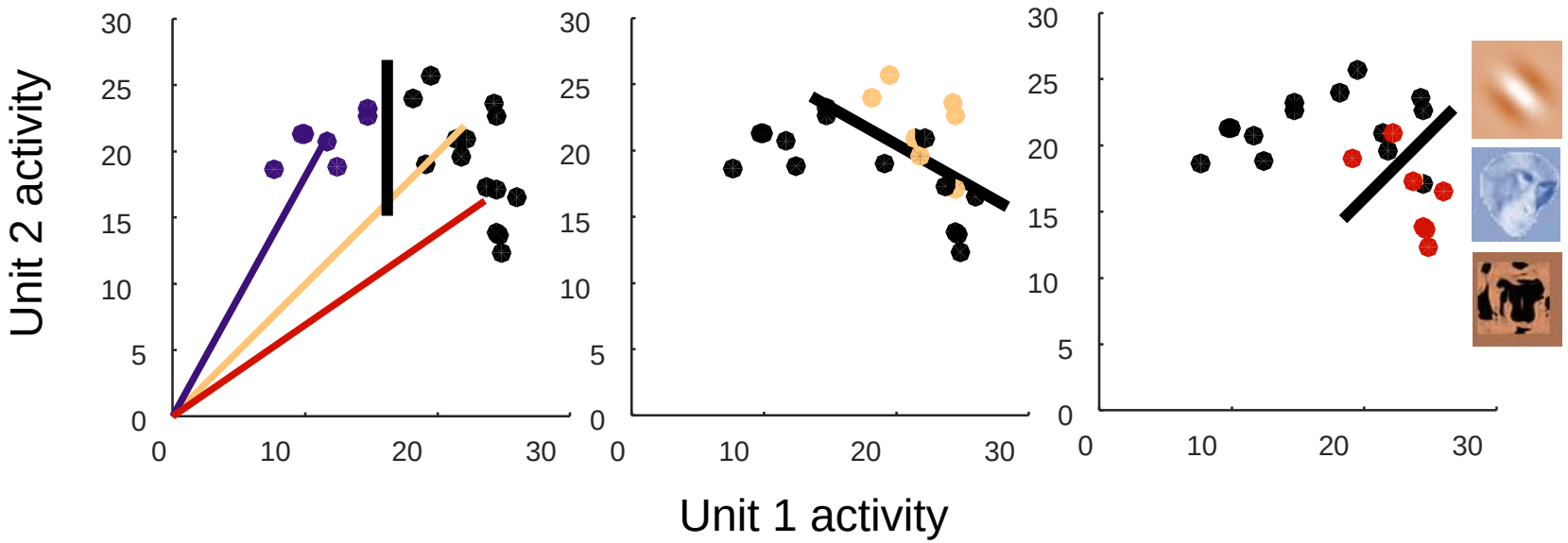
Support vector machines

-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.

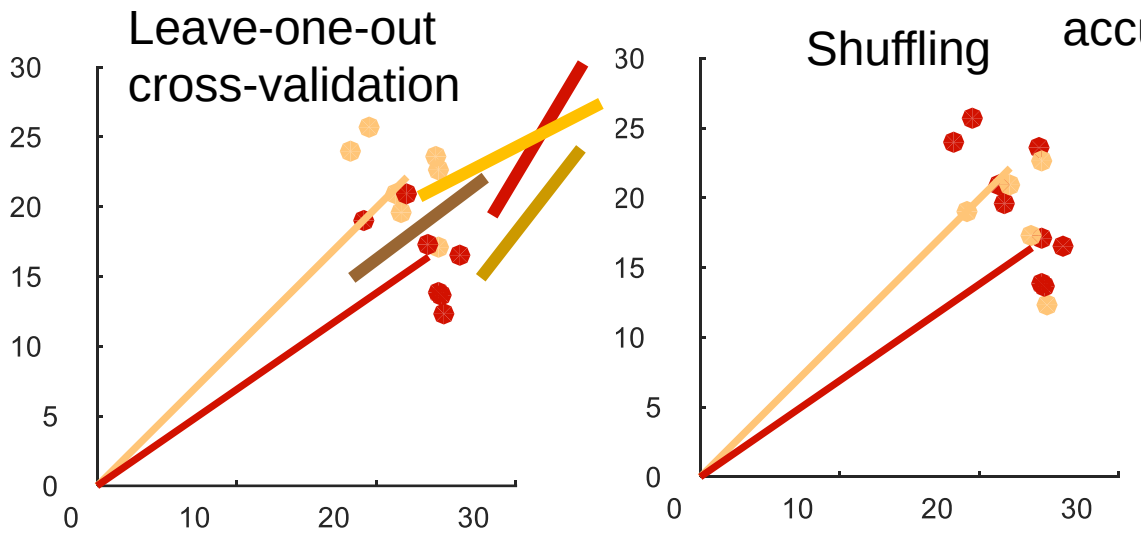
Label one category as positive, everything else as negative



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

How do we define the statistical reliability of classification accuracy?

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



Agenda

A brief recap: what you have seen so far in the course.

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

Lecture parts:

The anatomy of IT
What do IT cells encode? ("selectivity")
How good are they when contextual noise is introduced? ("invariance")
How do we use machine learning techniques to decode information in IT responses?

Paper discussion

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]

What is the scientific premise of the paper (i.e. background)?

What questions do the authors aim to answer?

FIGURE 1

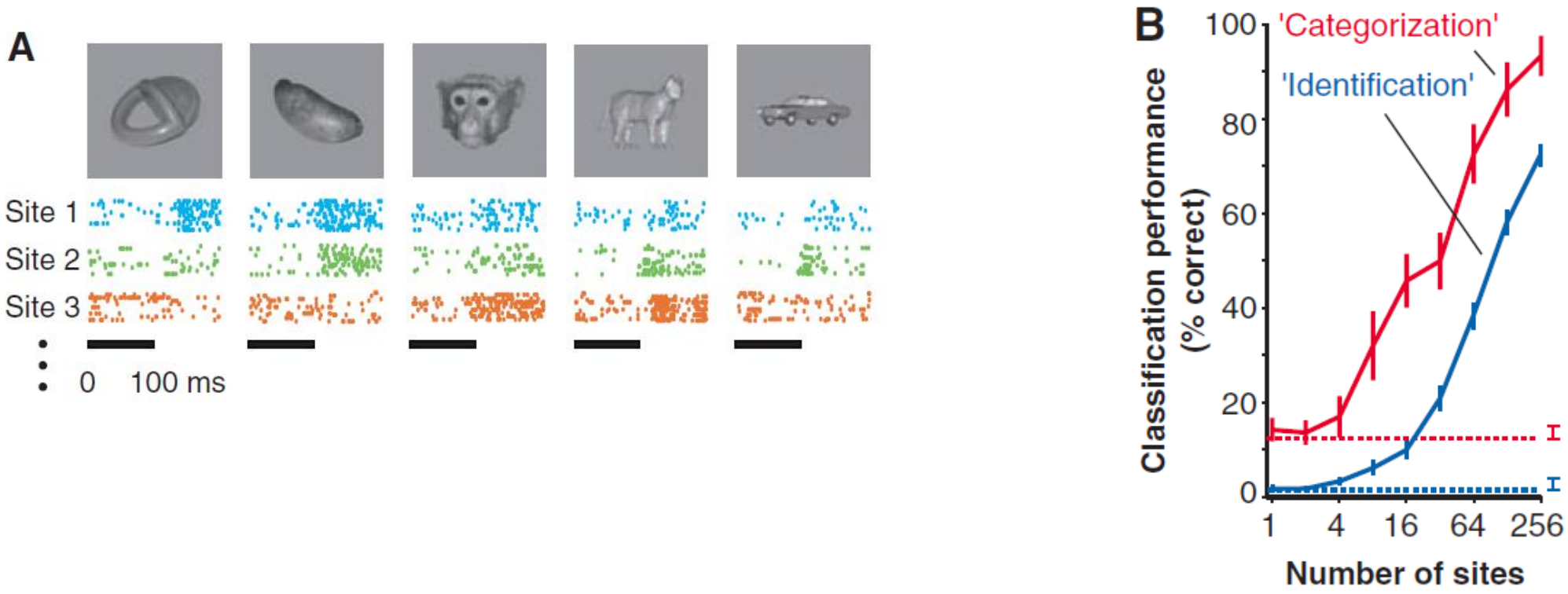
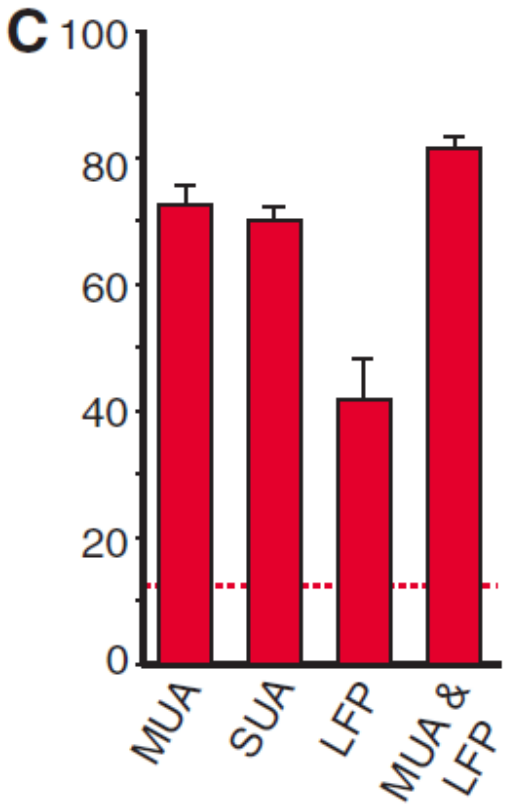

A
Site 1
Site 2
Site 3
0
100 ms
B
'Categorization'
'Identification'
Classification performance
(% correct)
100
80
60
40
20
0
1
4
16
64
256
Number of sites

FIGURE 1



Toys
Foodstuffs
Human faces
Monkey faces
Hand/body parts
Vehicles
Boxes
Cats and dogs

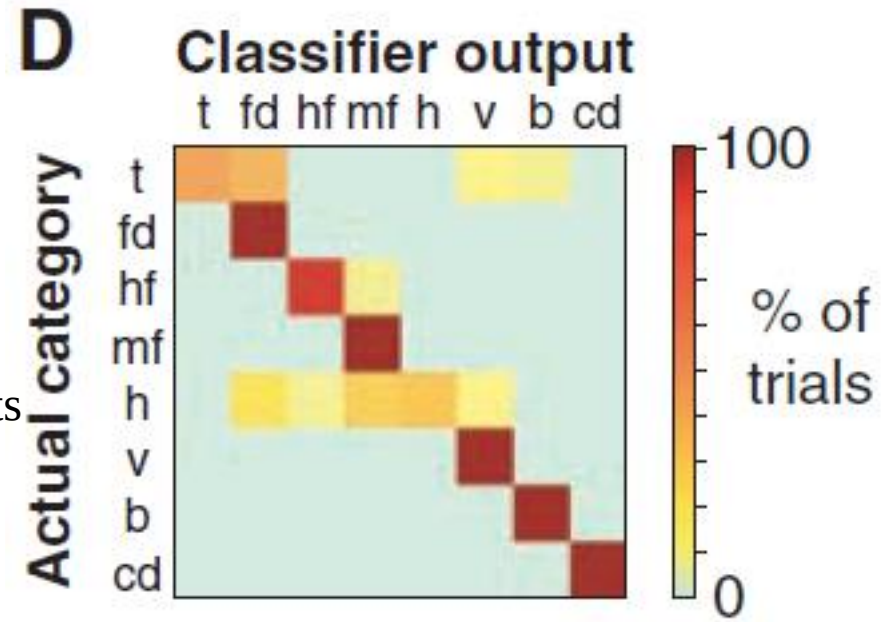


FIGURE 2

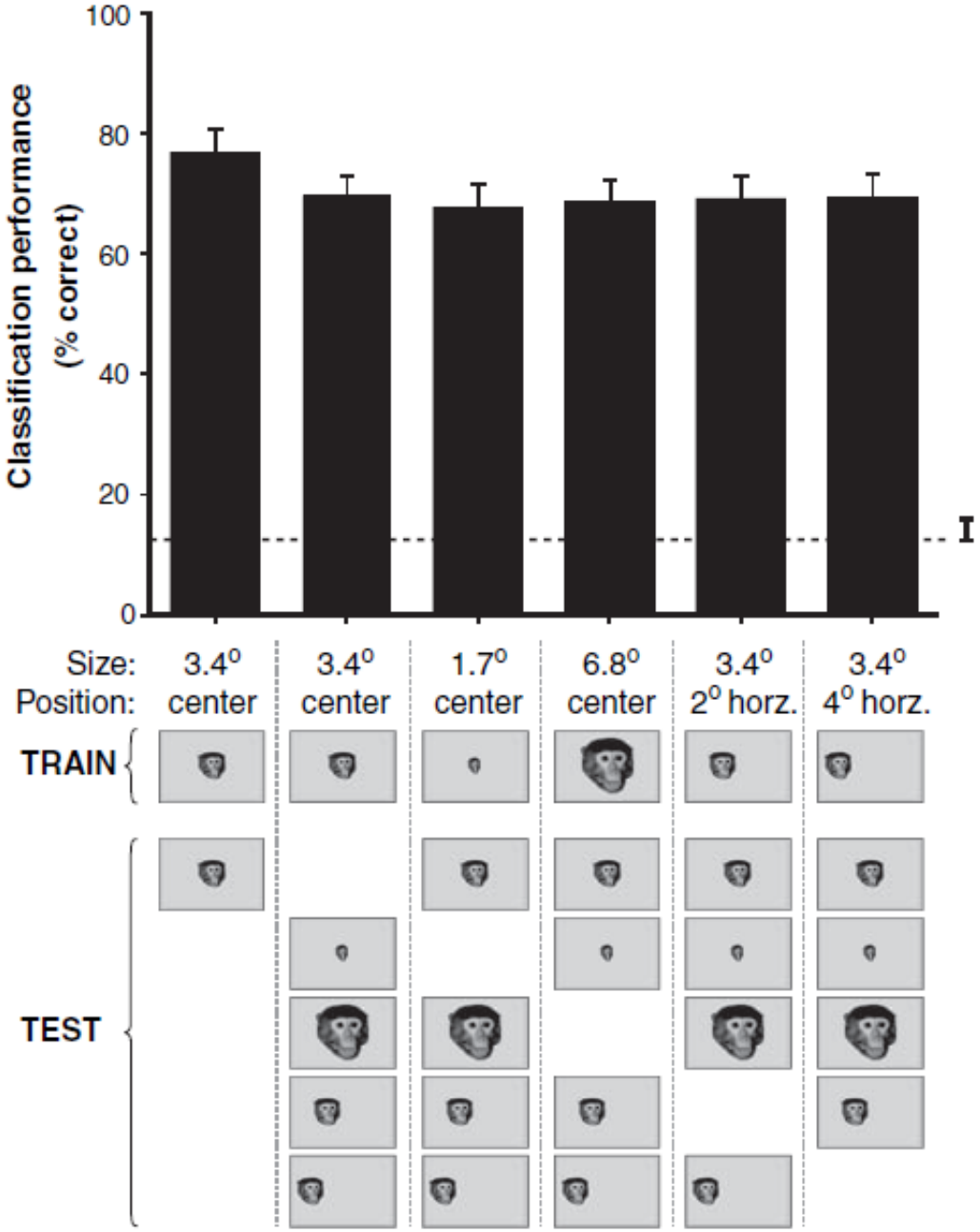
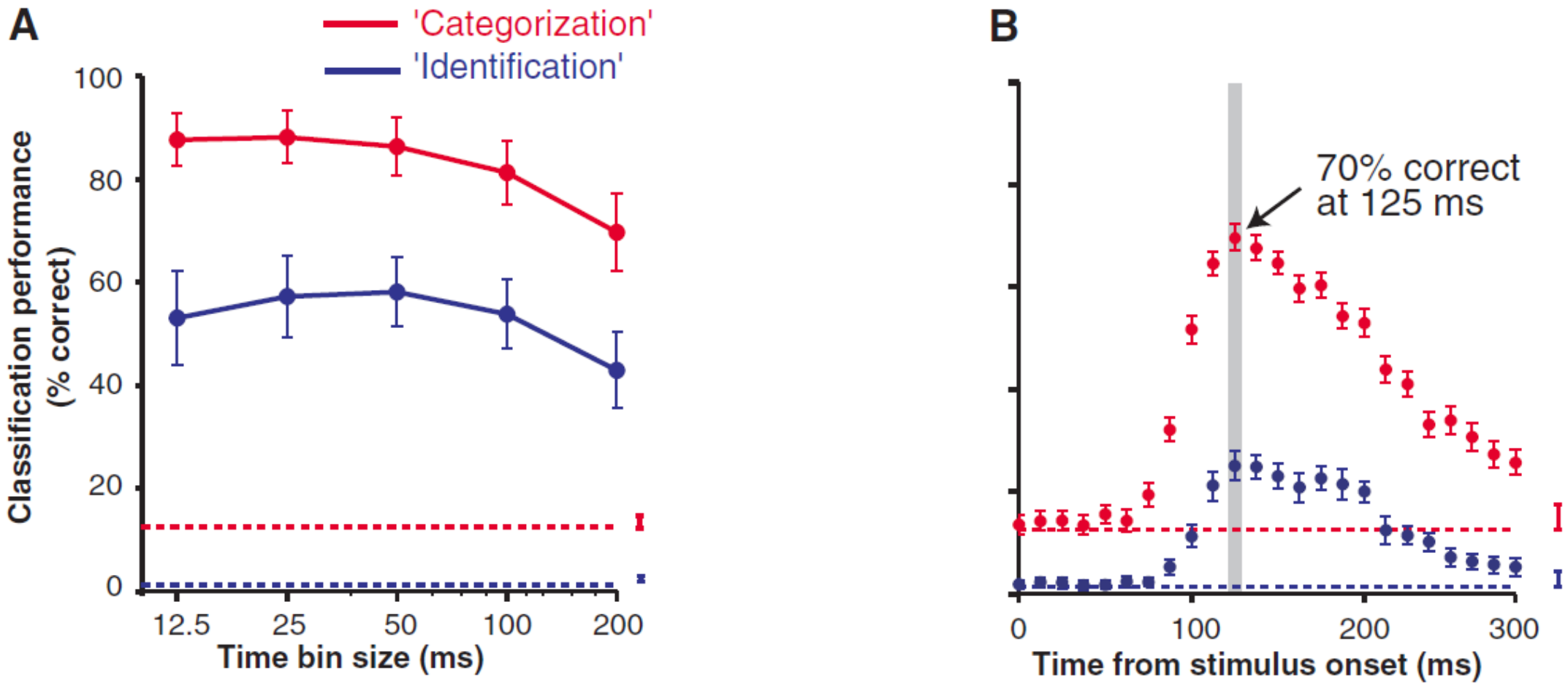

Classification performance
(% correct)
100
80
60
40
20
0
Size:
Position:
3.4°
center
3.4°
center
1.7°
center
6.8°
center
3.4°
2° horz.
3.4°
4° horz.
TRAIN
TEST

FIGURE 3


A
'Categorization'
'Identification'
Classification performance (% correct)
100
80
60
40
20
0
12.5
25
50
100
200
Time bin size (ms)
B
70% correct at 125 ms
0
100
200
300
Time from stimulus onset (ms)

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

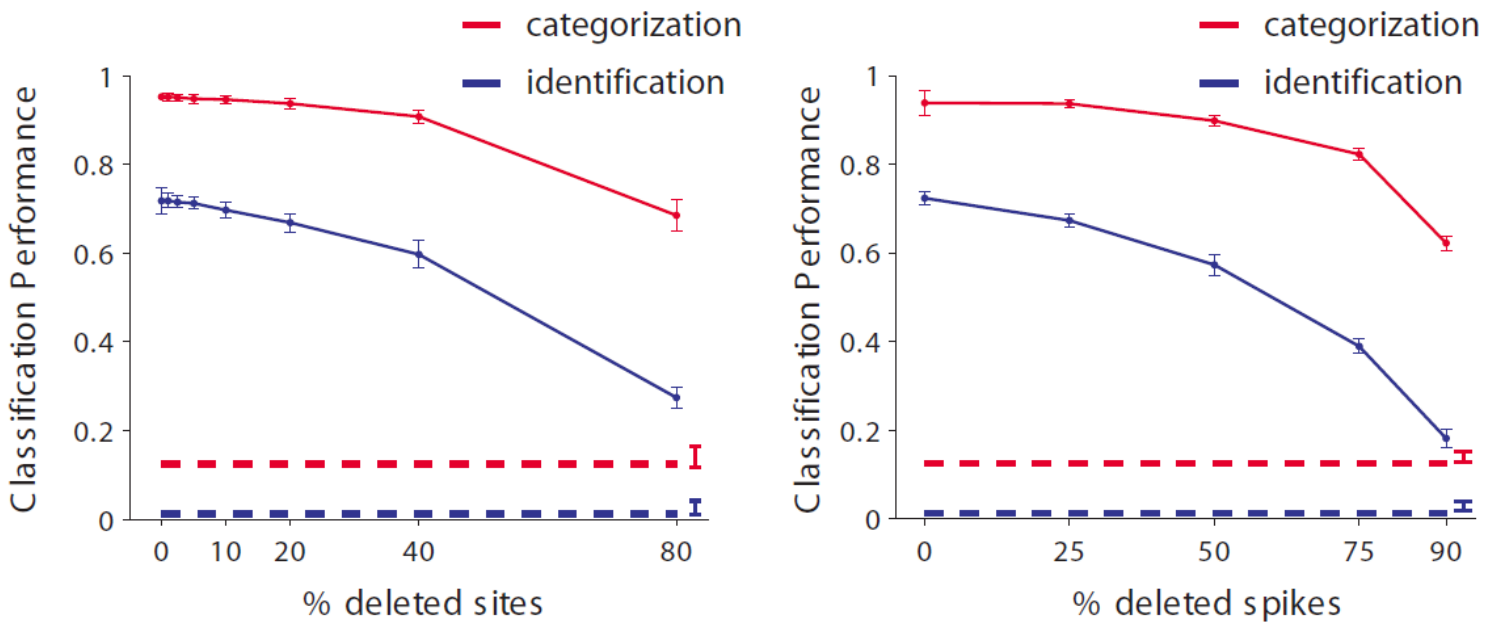
2009 - Liu H, Agam Y, Madsen J, Kreiman G. Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62:281-290

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

Fig. S1: The neural code is robust to neuronal drop-out and spike deletion



range of performances obtained using the 100 ms before stimulus onset

Fig. S2: Specific wiring significantly improves classifier performance

randomly selected sites (solid lines) or pre-selected sites (dotted lines).

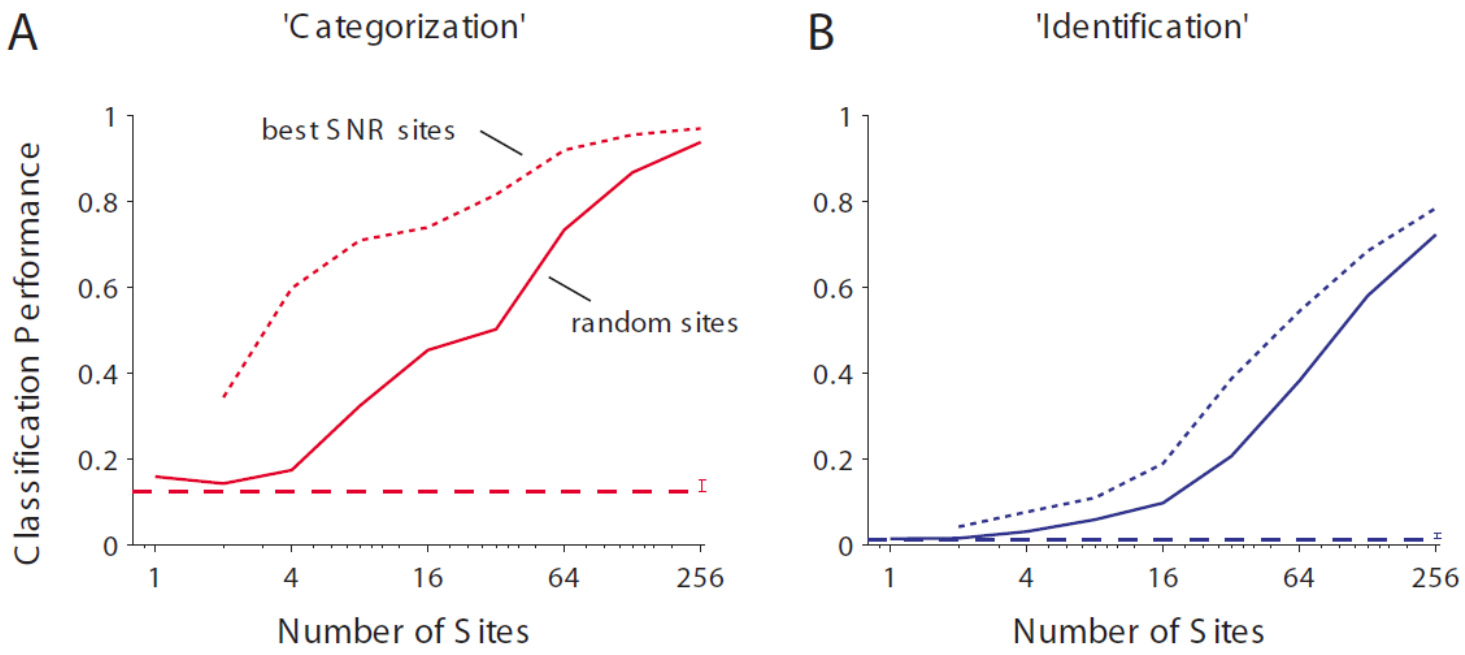


Fig. S3: Extrapolation to novel pictures within the same categories

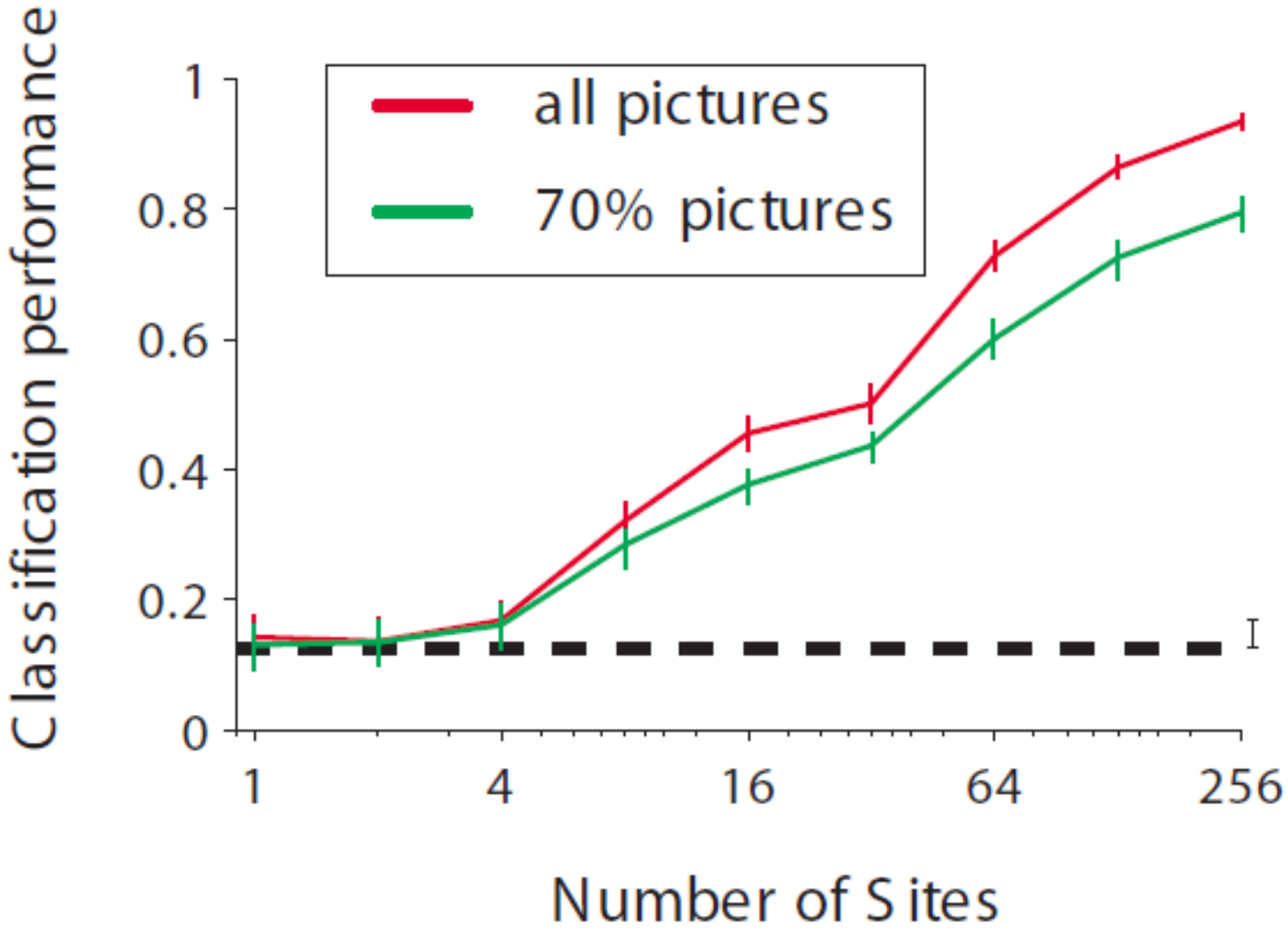
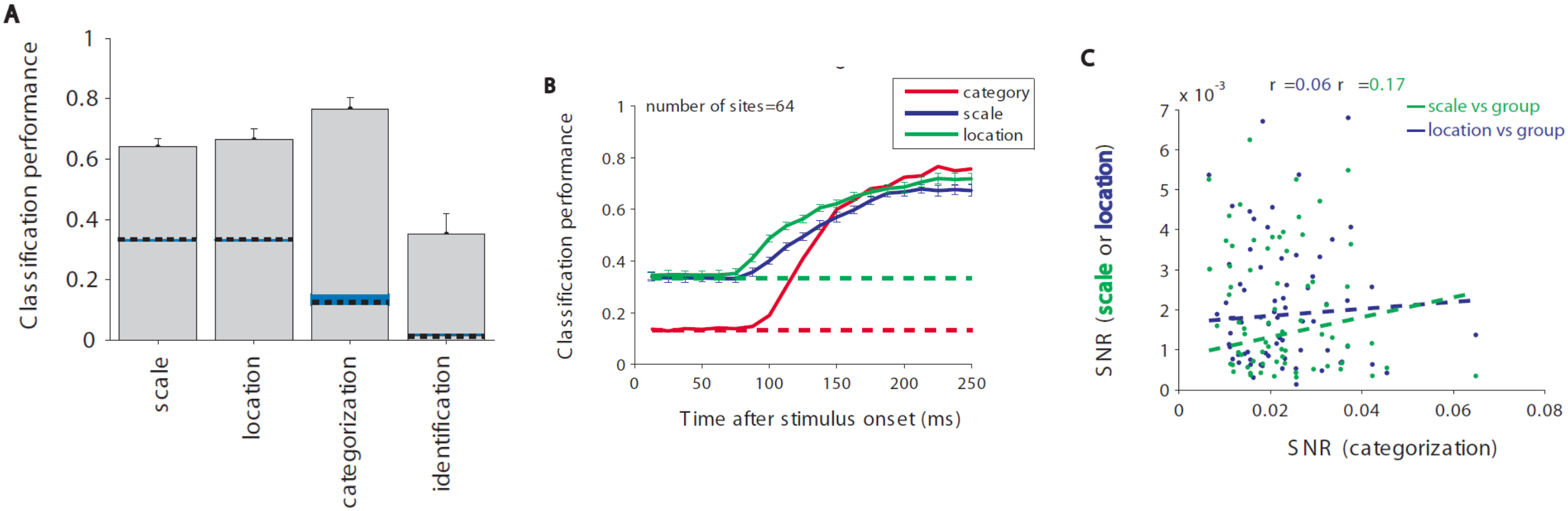


Fig. S4: Scale and position ('where') can be read out independently of object identity ('what')



chance was 1/3 for read-out of scale and position.

Fig. S5: Read-out of stimulus onset

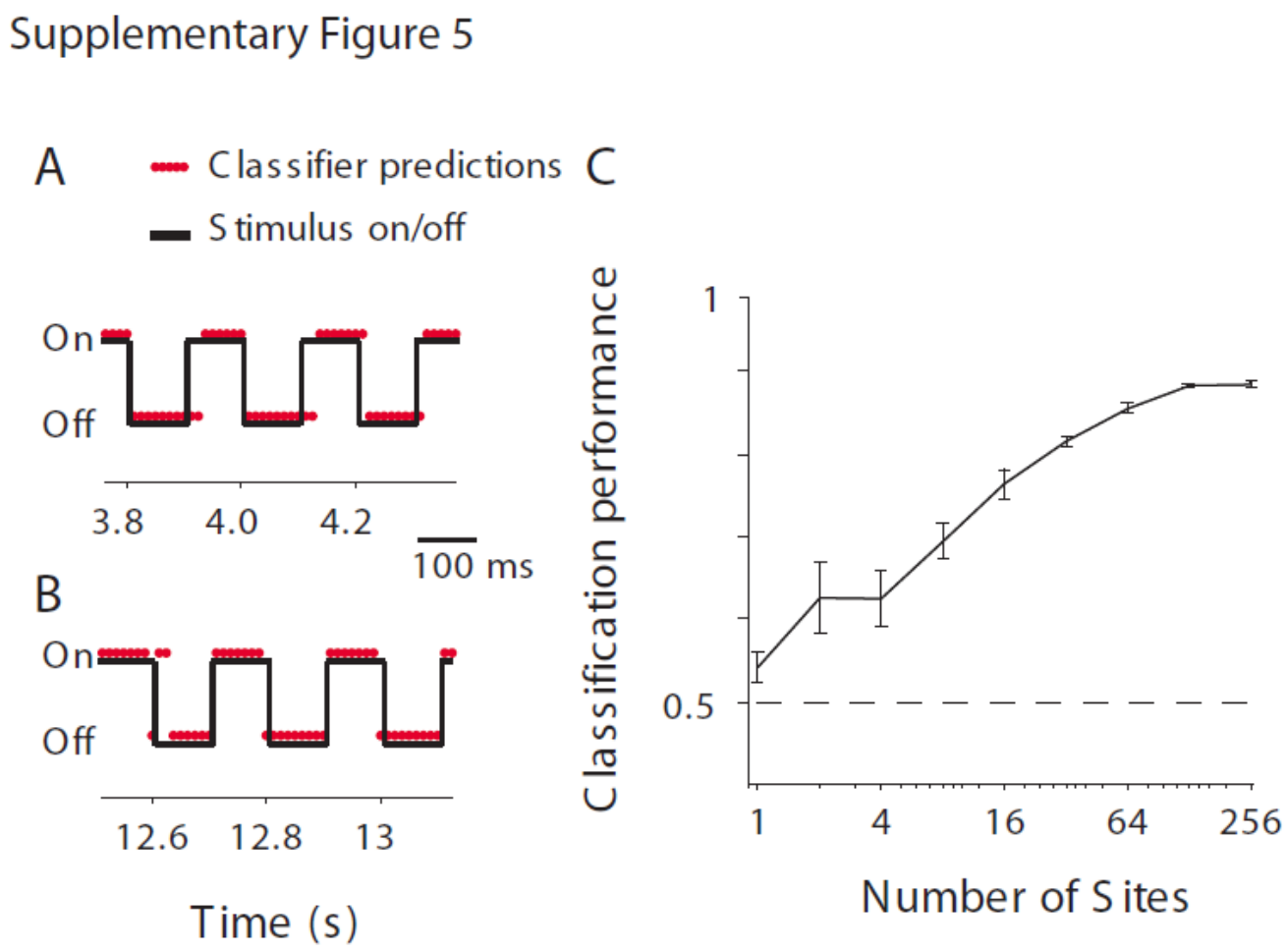
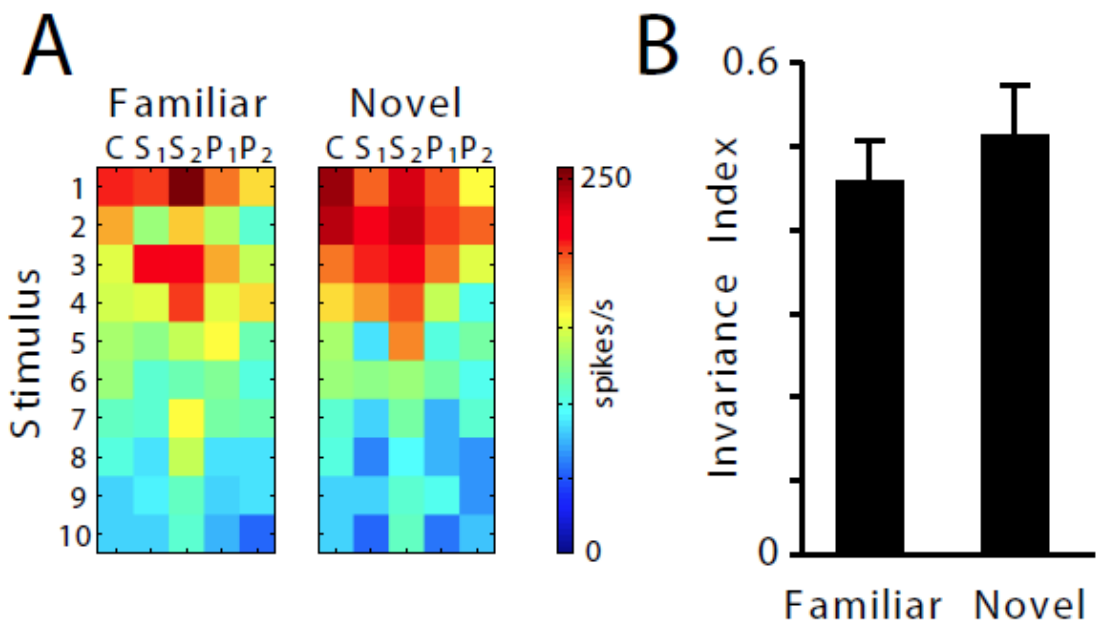


Fig. S6: Invariance to novel objects



(C = center of gaze, 3,4° size; S1=center of gaze, 1.7° size; S2 = center of gaze, 6.8° size; P1 = 2° shift, 3.4° size; P2 = 4° shift, 3.4° size.

Fig. S7: Comparison of responses during active versus passive tasks