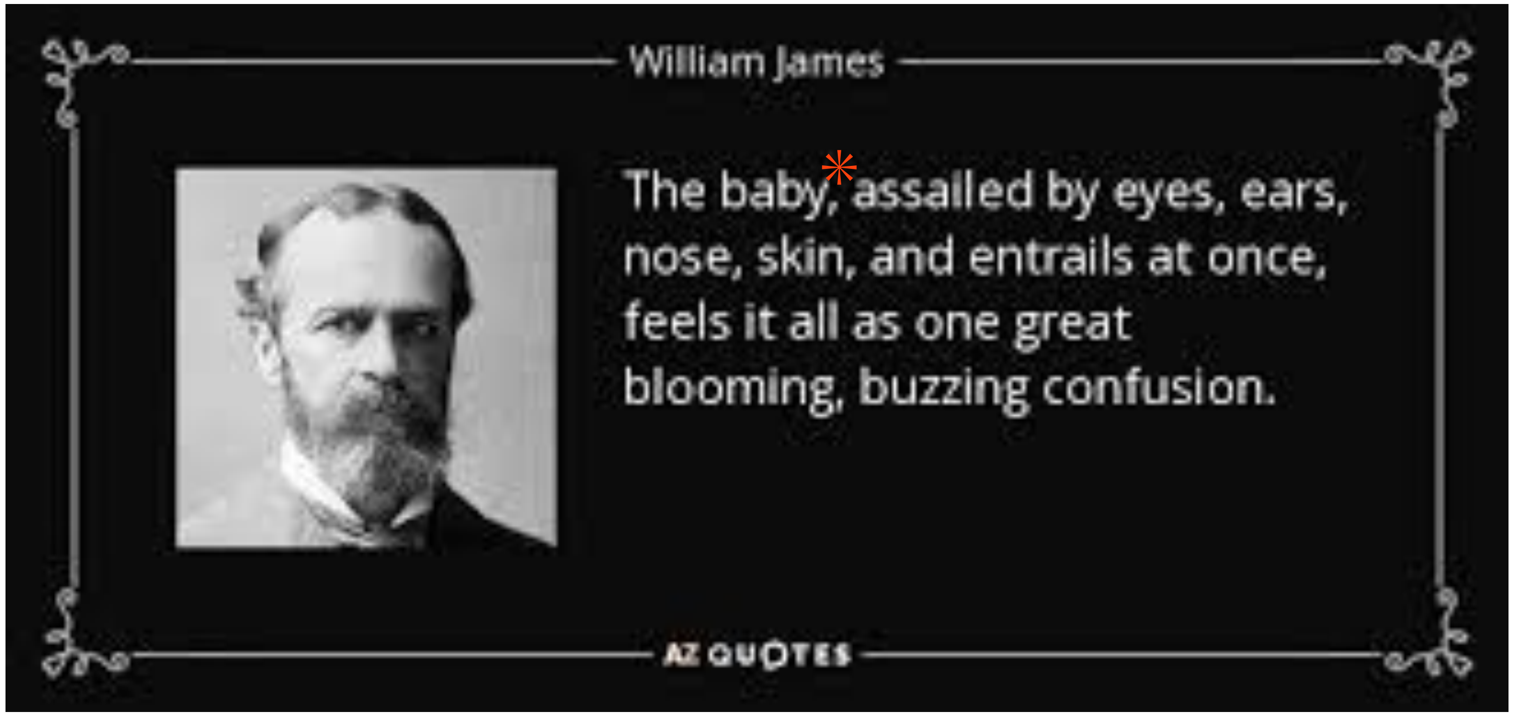
CS375 / Psych 249:
Large-Scale Neural Network Models for Neuroscience

Lecture 2: The biological inspiration of CNNs
2025.01.07

Daniel Yamins
Departments of Computer Science and of Psychology
Stanford Neuroscience and Artificial Intelligence Laboratory
Wu Tsai Neurosciences Institute
Stanford University

Problem: Entity Extraction

Understanding complex, noisy data streams is a critical part of cognition.



Without sophisticated parsing and entity extraction, the world would be "as one great blooming, buzzing confusion" (for babies or otherwise).

*actually not clearly true for babies . . .

Problem: Entity Extraction

Why is the problem hard computationally?

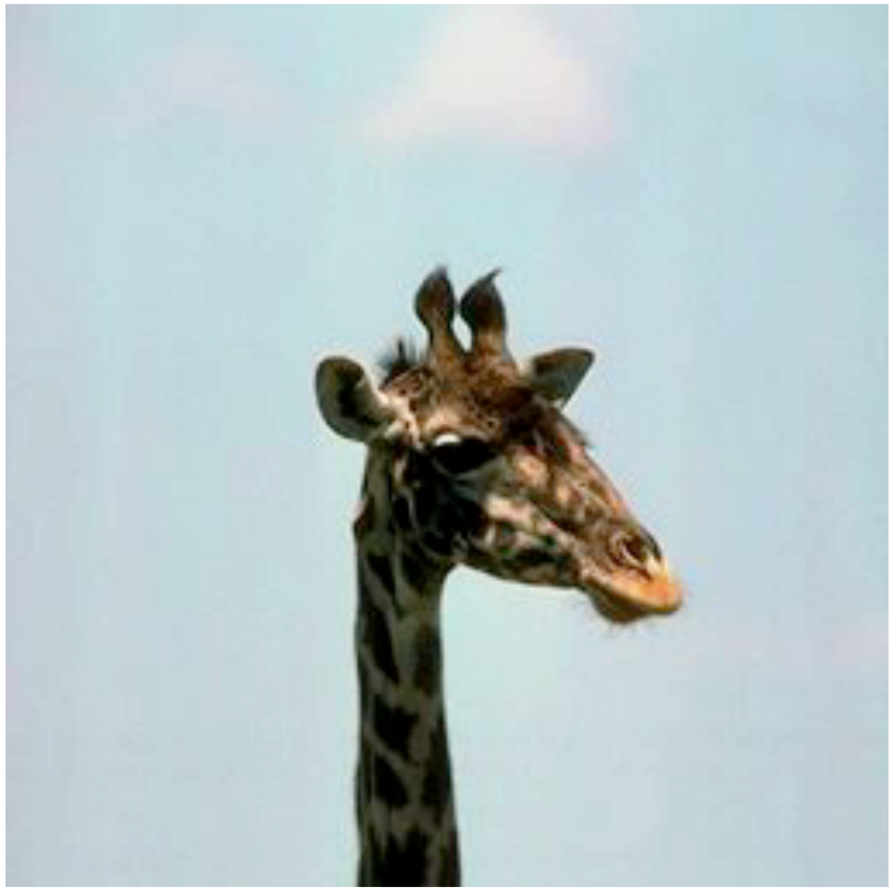
1. Nonlinear misalignment between physical and behavioral dimensions

Problem: Entity Extraction

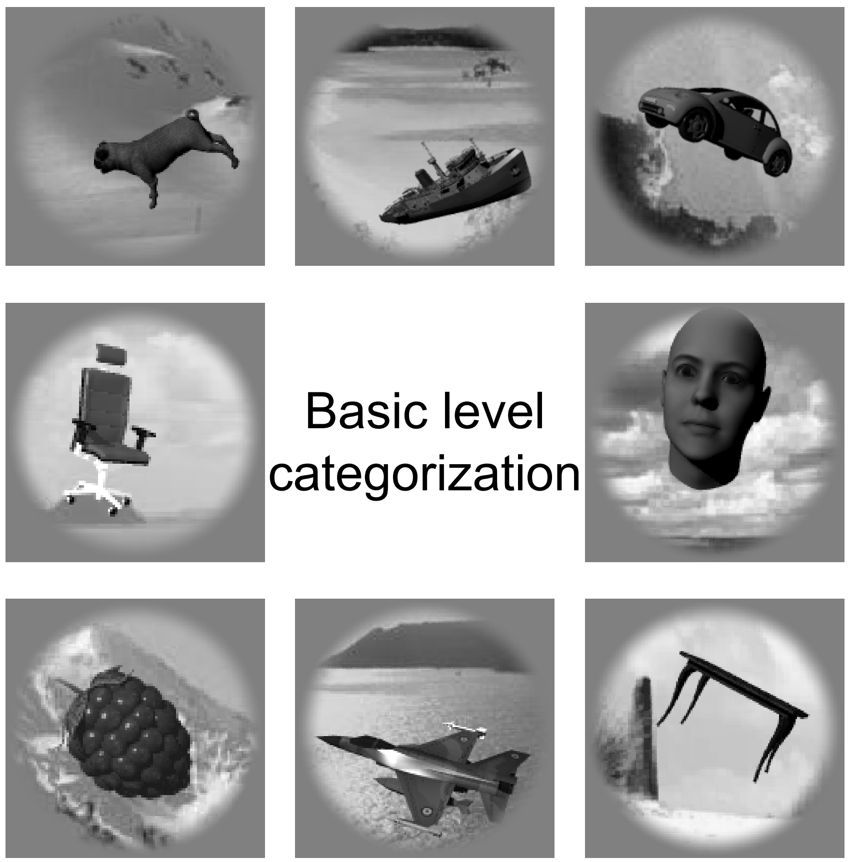
Why is the problem hard computationally?

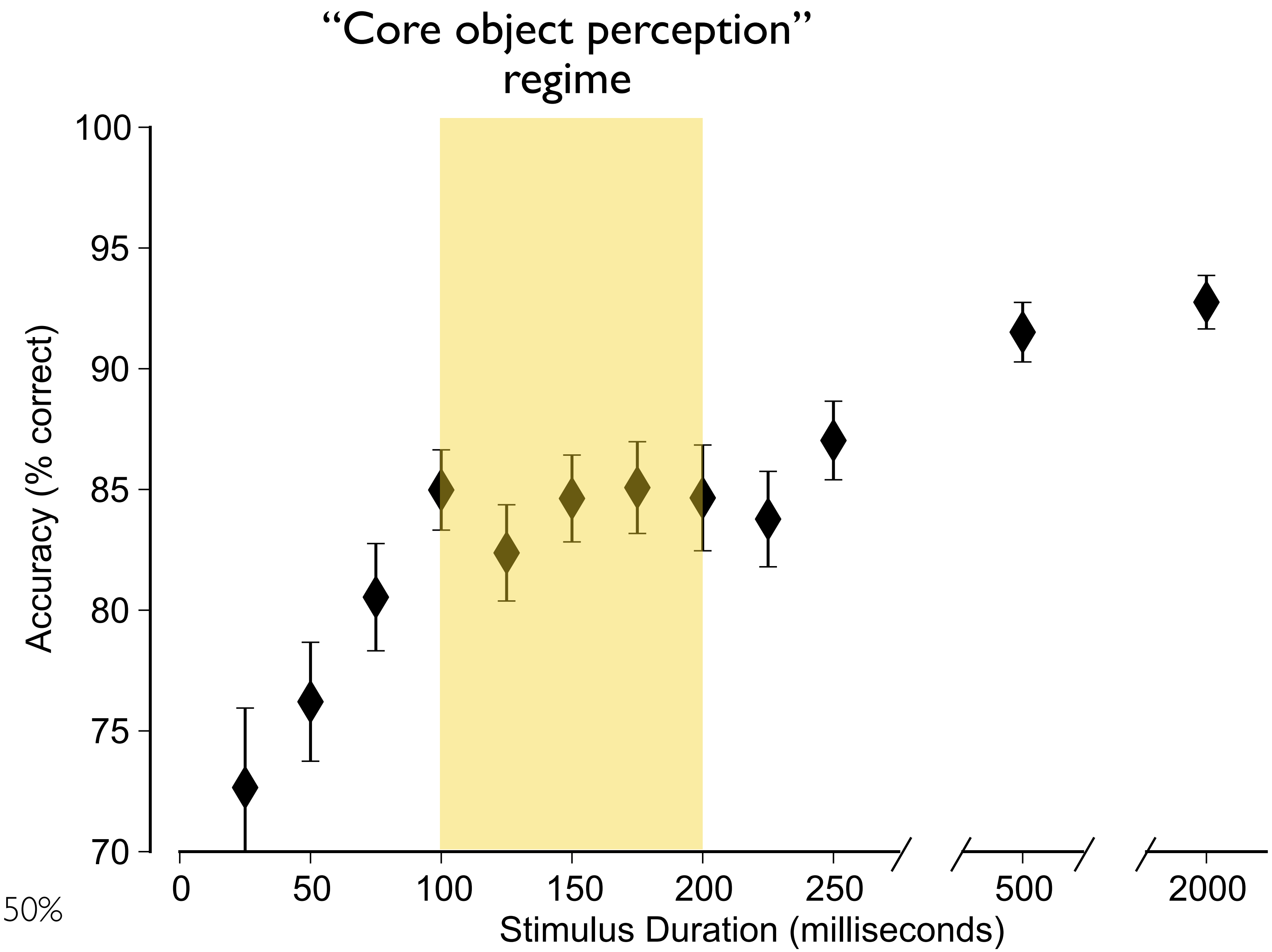
1. Nonlinear misalignment between physical and behavioral dimensions

2. Needs to be done ***fast***, and thus, presumably, massively in parallel



Problem: Entity Extraction

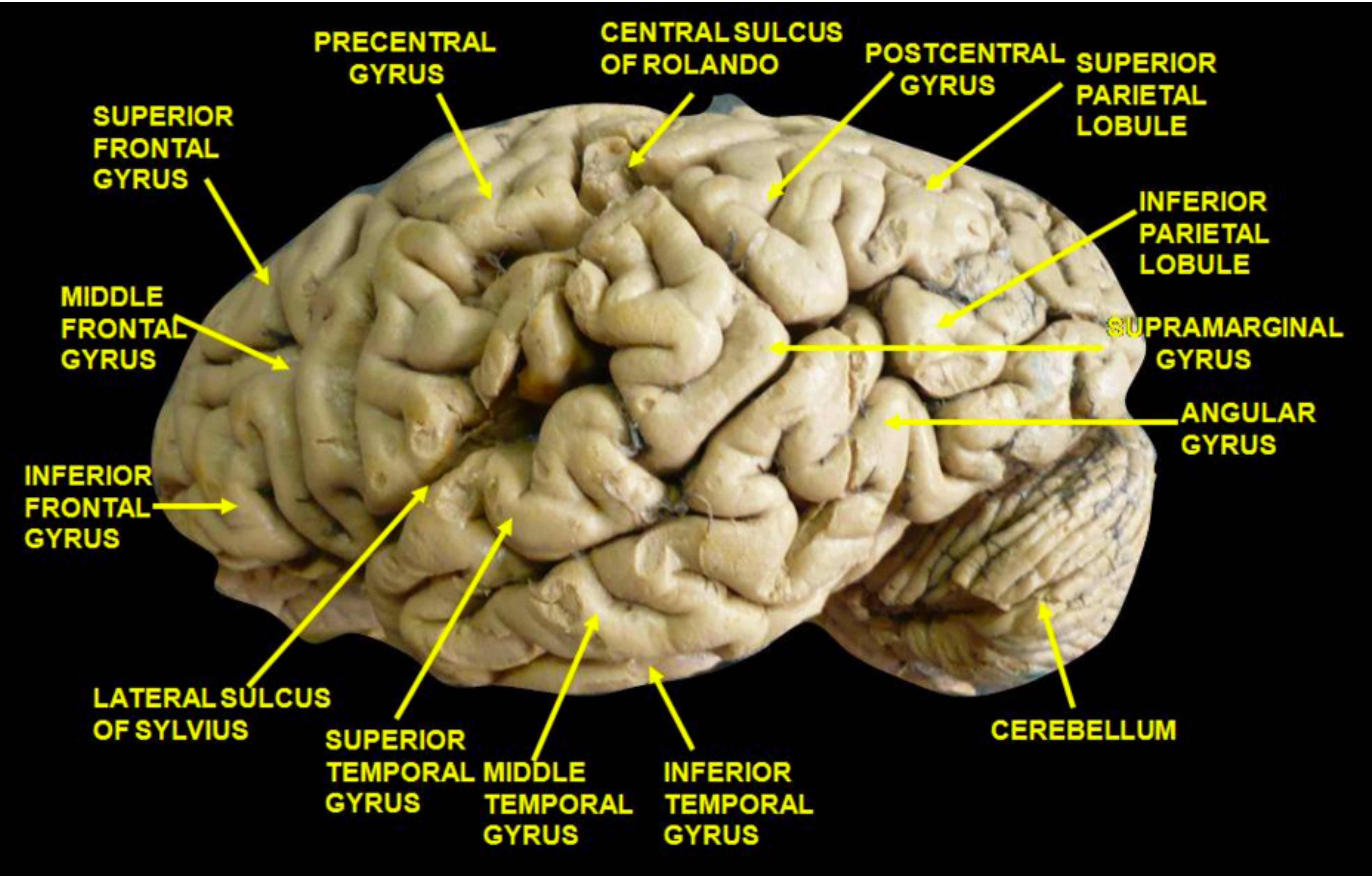




Chance is 50%

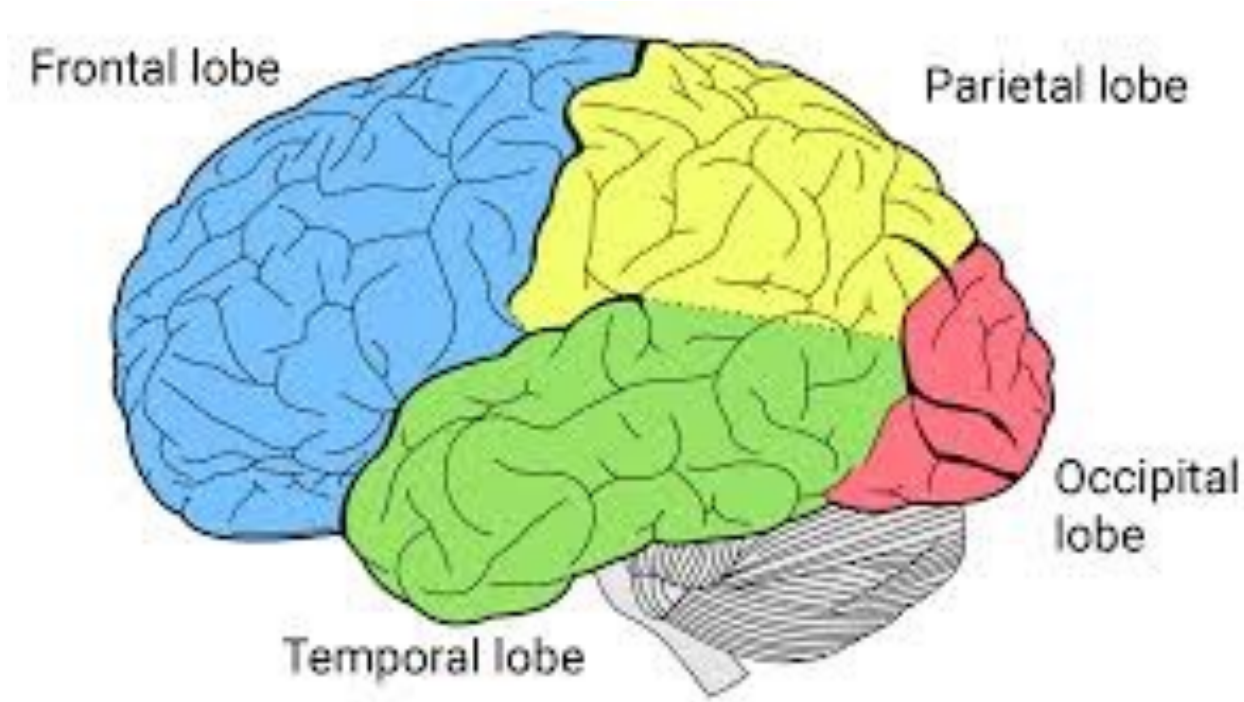
All the data I will show you today

Typical primate fixation duration during natural viewing


PRECENTRAL GYRUS
CENTRAL SULCUS OF ROLANDO
POSTCENTRAL GYRUS
SUPERIOR PARIETAL LOBULE
SUPERIOR FRONTAL GYRUS
INFERIOR PARIETAL LOBULE
MIDDLE FRONTAL GYRUS
SUPRAMARGINAL GYRUS
ANGULAR GYRUS
INFERIOR FRONTAL GYRUS
LATERAL SULCUS OF SYLVIUS
SUPERIOR TEMPORAL GYRUS
MIDDLE TEMPORAL GYRUS
INFERIOR TEMPORAL GYRUS
CEREBELLUM

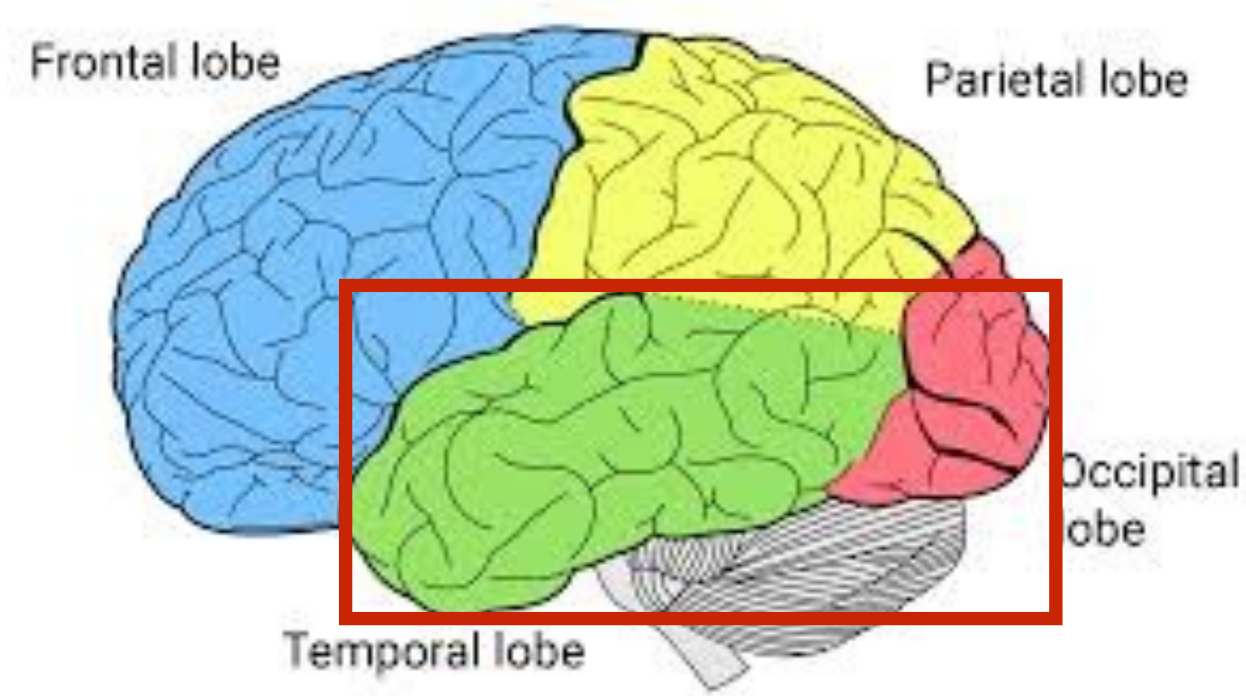
Many Different Computational Goals

- Sensory processing
  - visual, auditory, somatosensory recognition (occipital, temporal)
  - navigation (hippocampus?)
- motor command production & execution (motor cortex)
- memory, decision making and planning (hippocampus, prefrontal cortex)
- language
- emotions, theory-of-mind



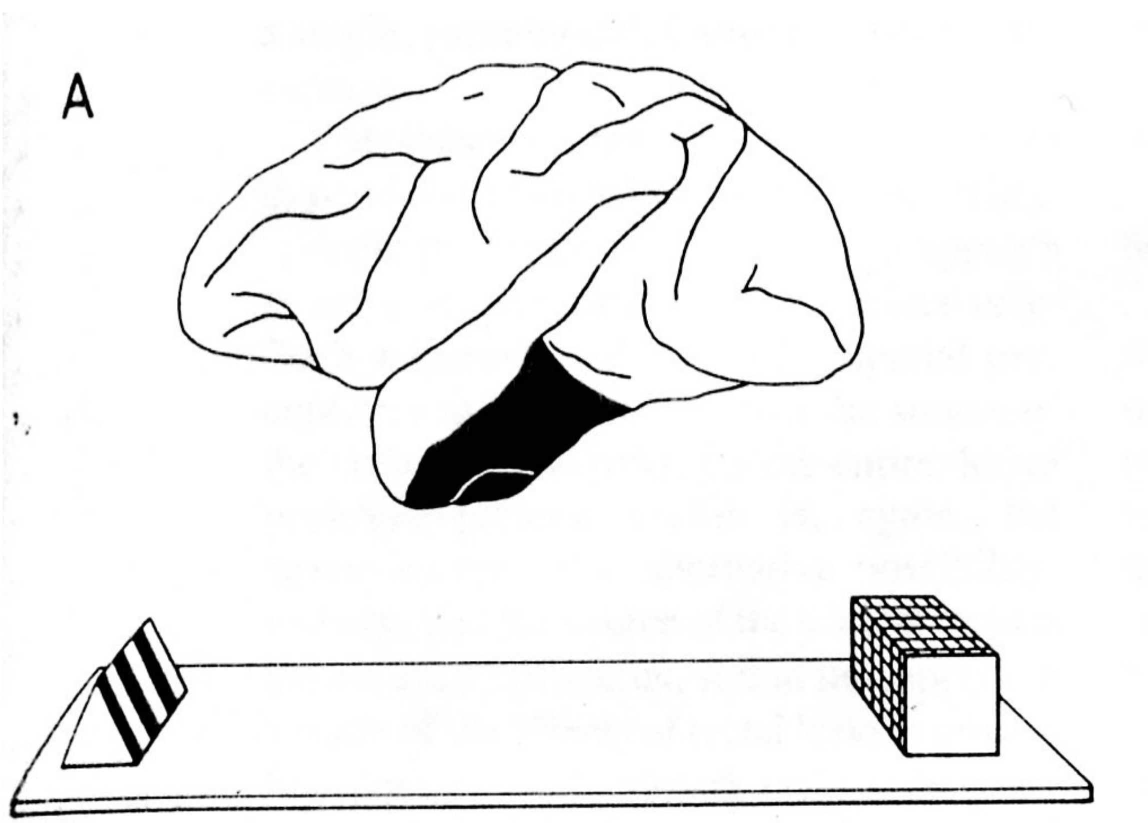
Many Different Computational Goals

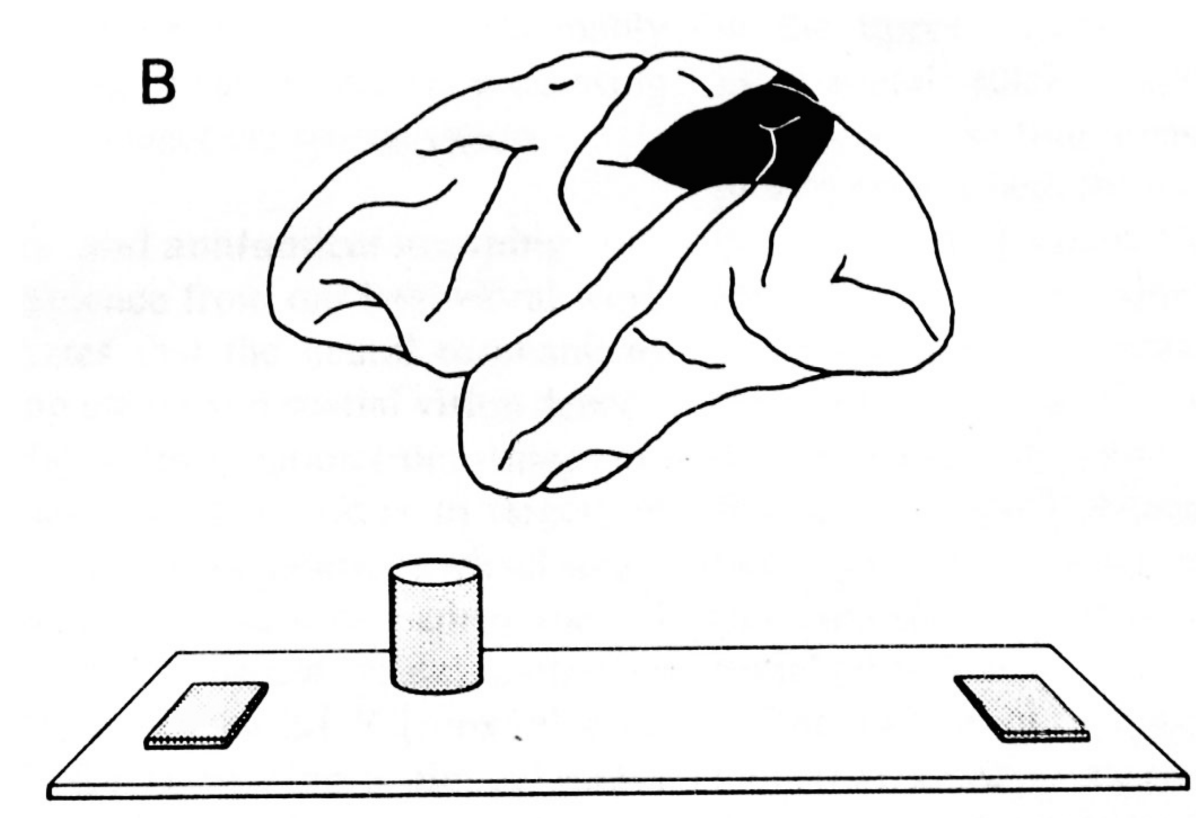
- Sensory processing
  - visual, auditory, somatosensory recognition (occipital, temporal)
  - navigation (hippocampus?)
- motor command production & execution (motor cortex)
- memory, decision making and planning (hippocampus, prefrontal cortex)
- language
- emotions, theory-of-mind



Background: Ventral visual stream

Mishkin & Ungerleider, 1982





"what" / ventral / occipitotemporal

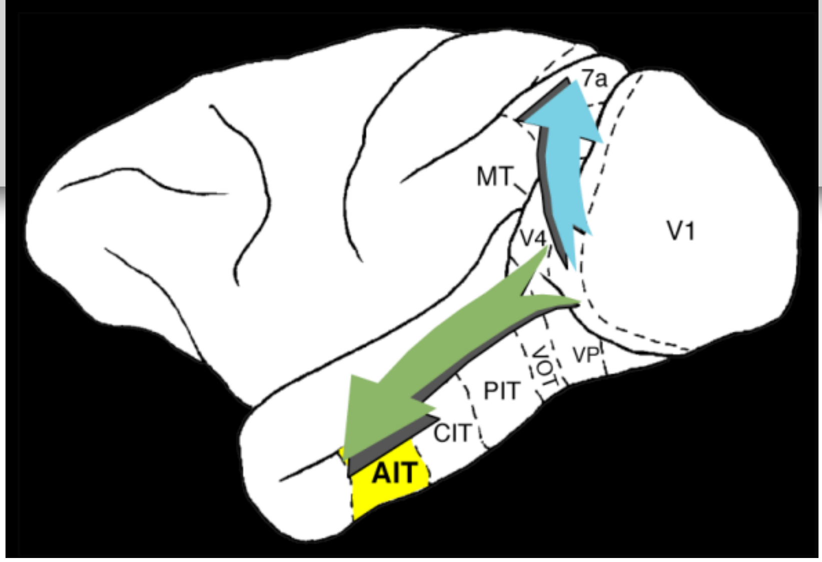
Lesions in IT cortex produce deficits in shape discrimination tasks (Gross et al, 1973, Mishkin 1982)

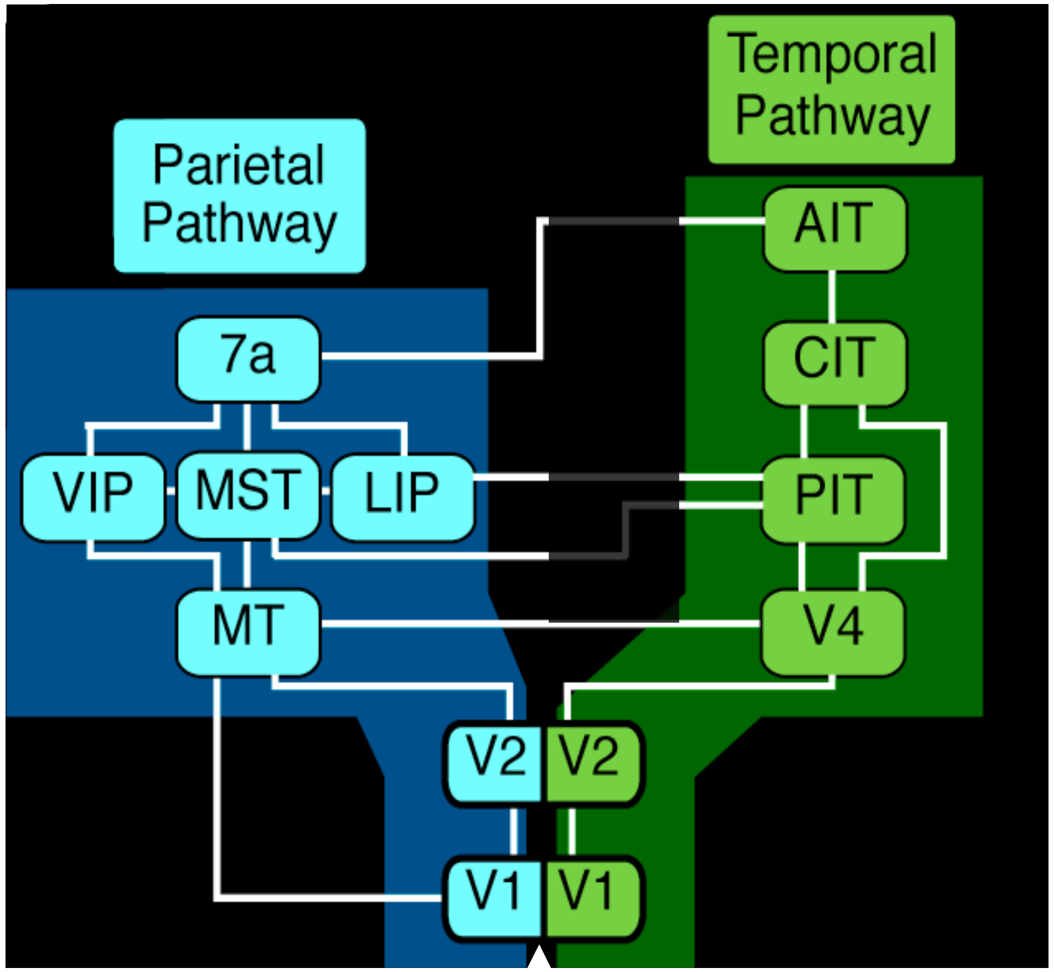
"where" / dorsal / parietal

Lesions in parietal cortex produce deficits in landmark task (Pohl et al. 1973)

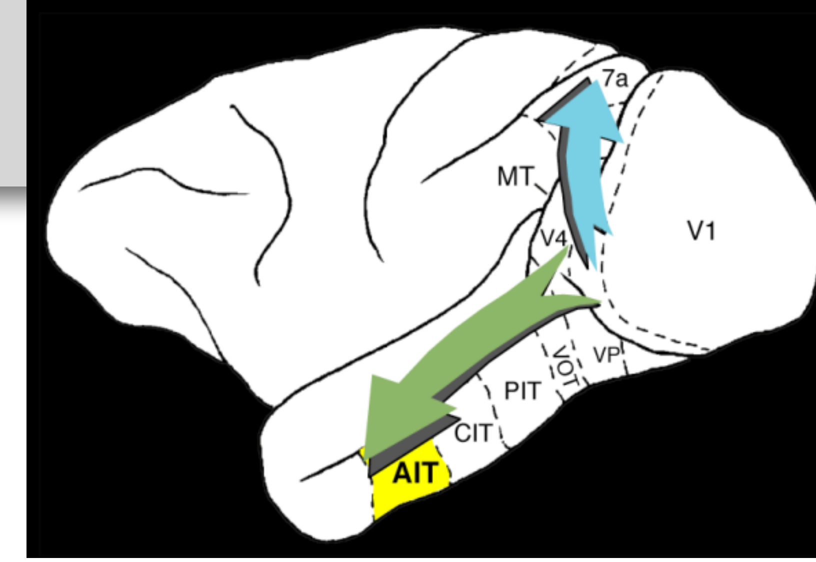
Source of Slide: Jody Culham

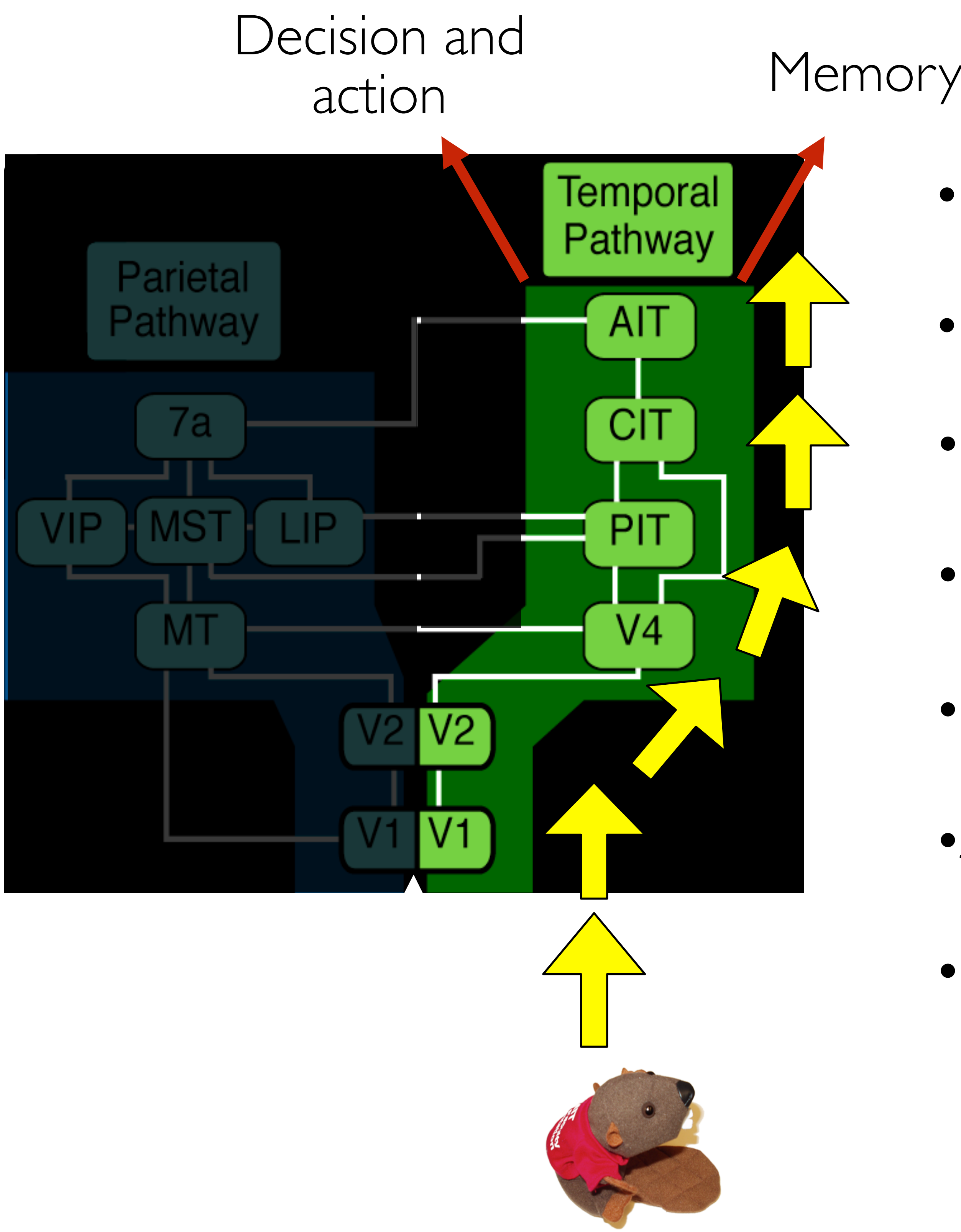
Background: Ventral visual stream





Background: Ventral visual stream

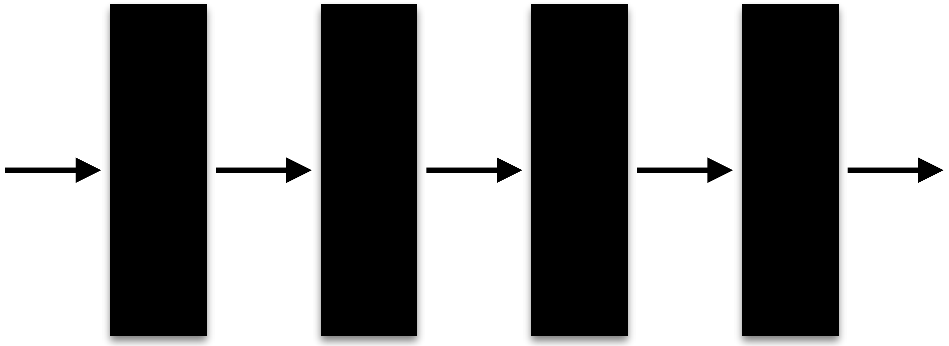




- *<u>Tolerance</u> to identity-preserving transforms*
- *Ability to support visual recognition*
- *Correlation with perceptual report*
- *Sensitivity to behavioral state (e.g. attention)*
- *Visually-evoked latency*
- *Selectivity to visual "feature" conjunctions*
- *Effects of experience (plasticity)*

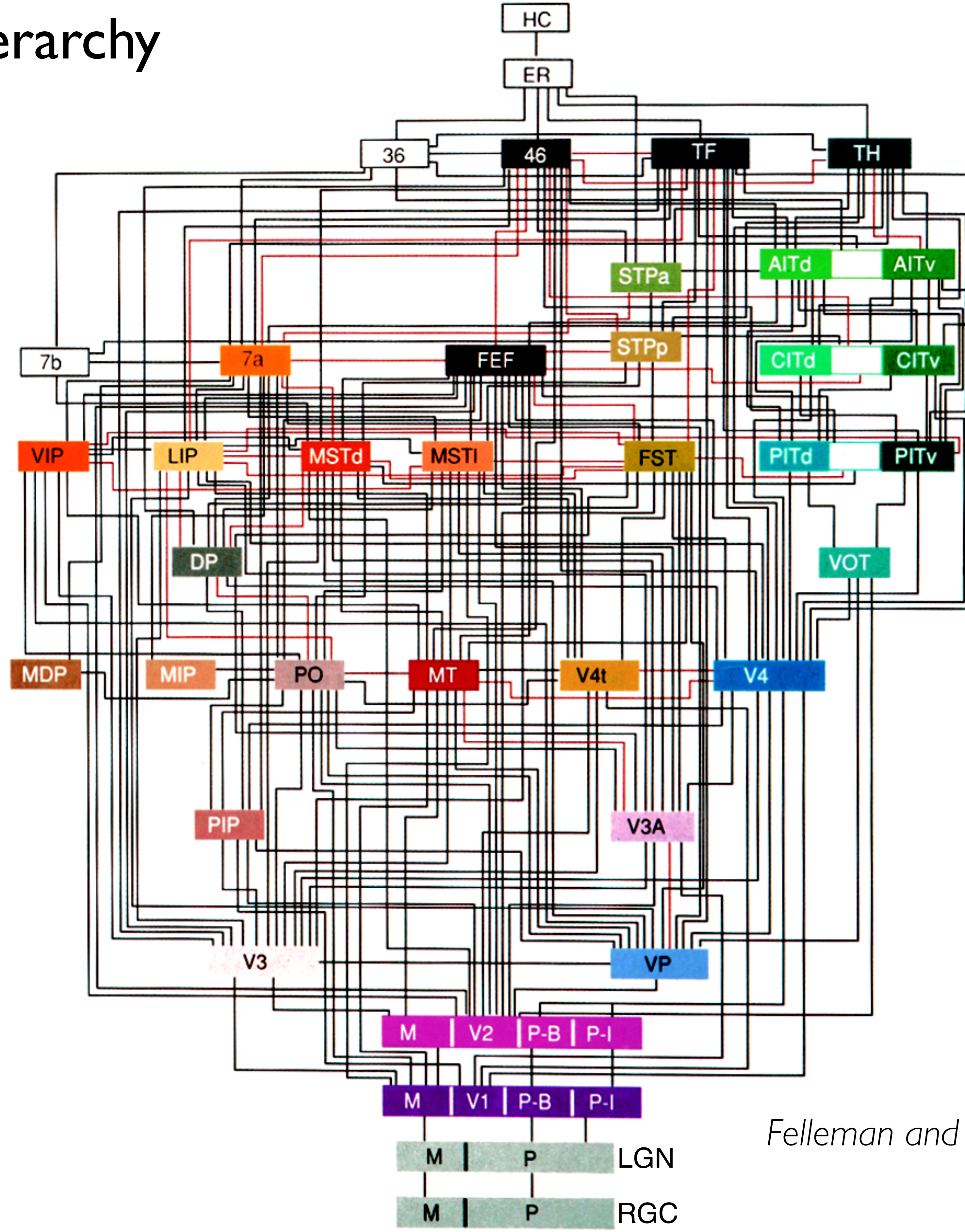
Sensory cascade

sensory cascade in
visual (mostly-) cortex



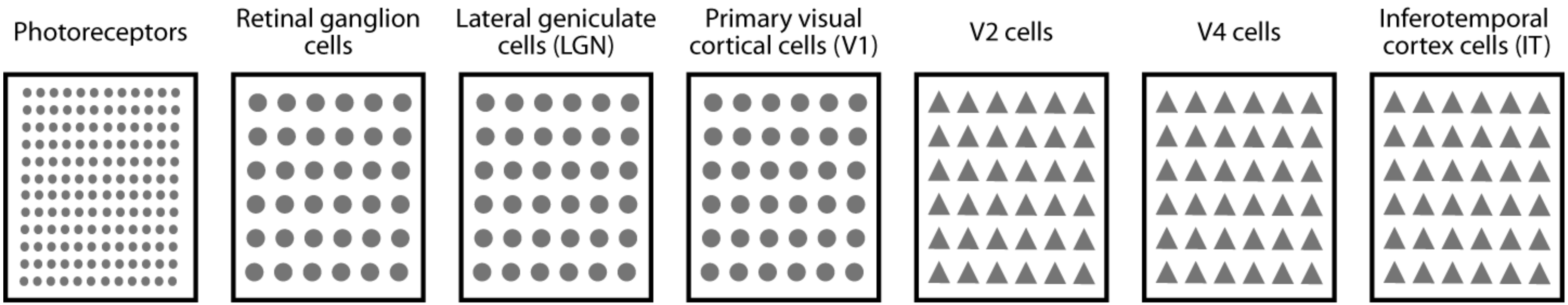
Madame Curie!

Visual area hierarchy

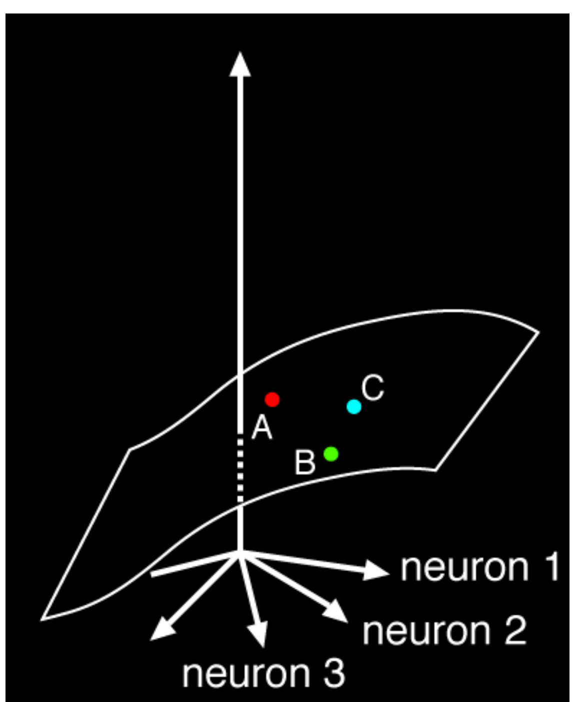


Felleman and Van Essen, 1991

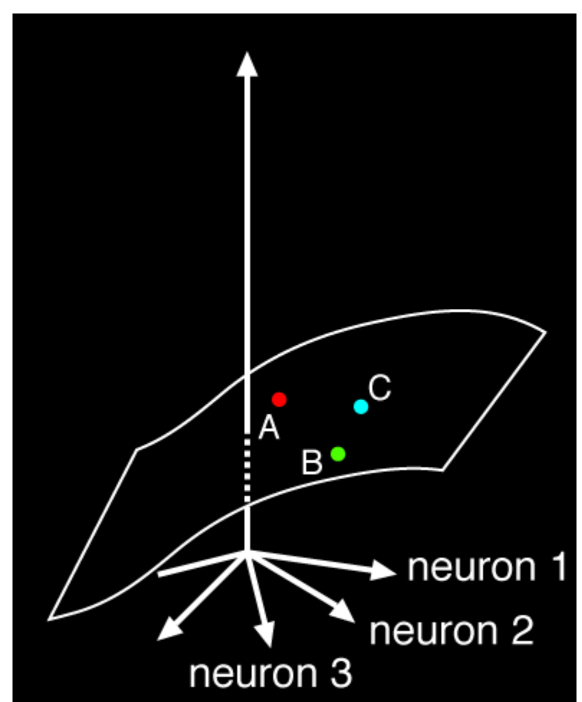
How does the brain represent the visual world?



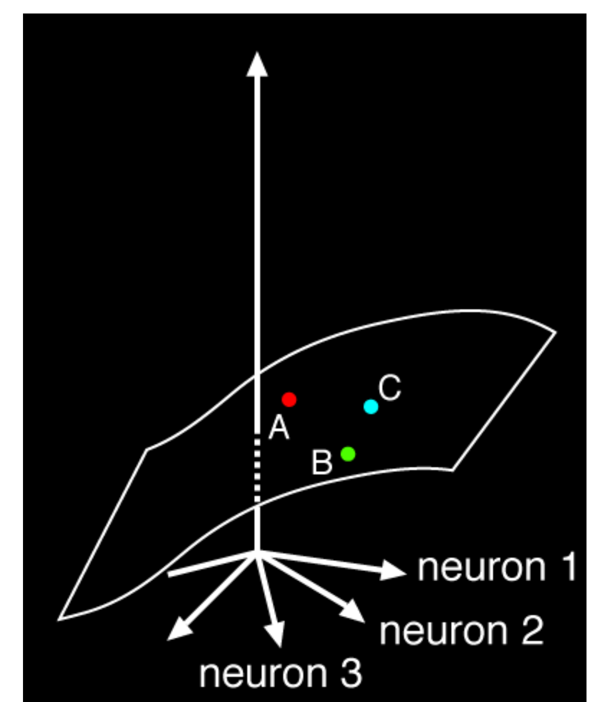
photoreceptor representation



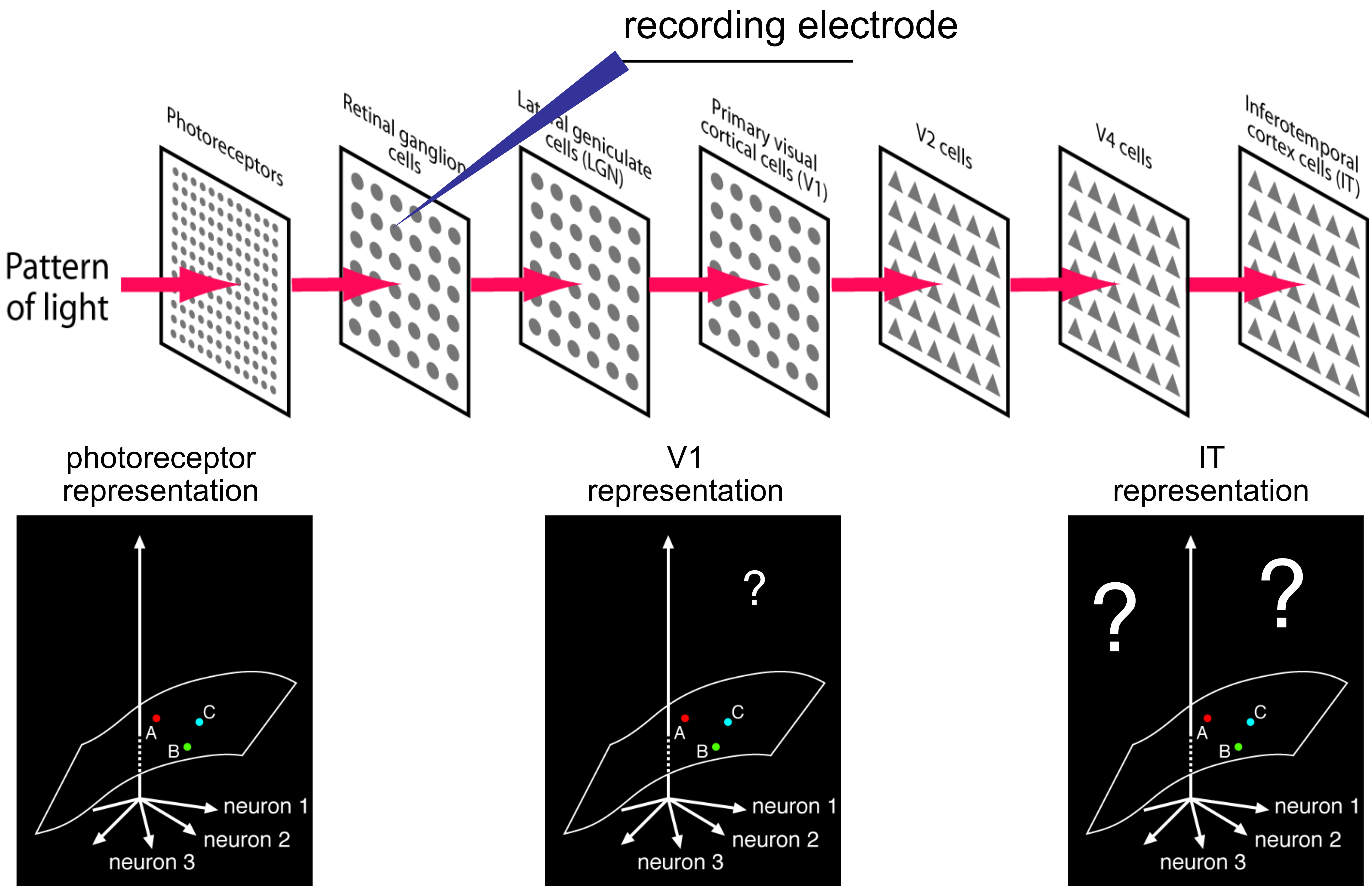
V1 representation



IT representation



How does the brain re-represent the visual world?



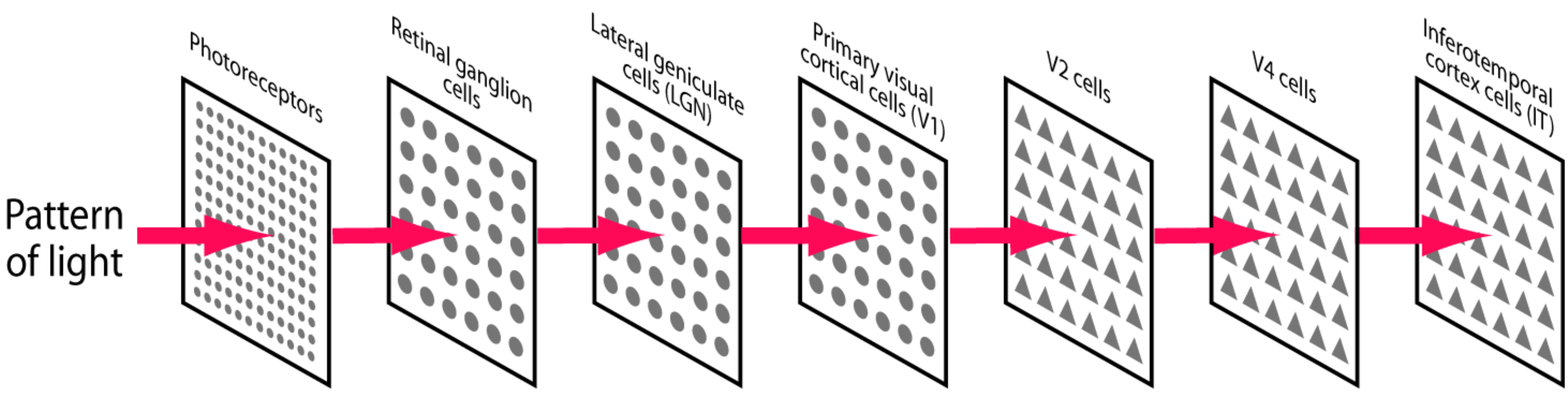
Four important pieces on information

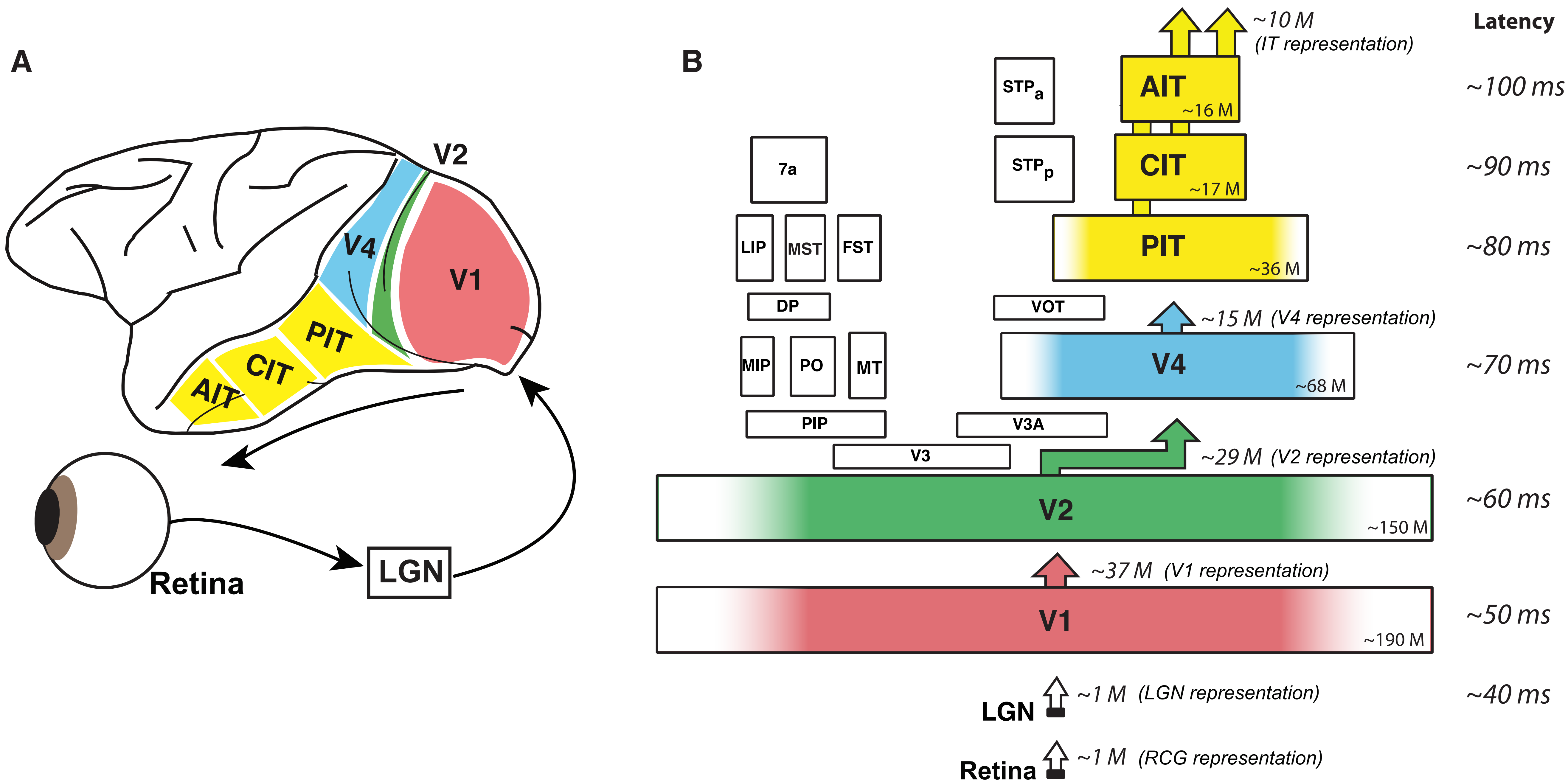
1) Neuronal selectivity generally increases as we move up the cortical hierarchy

2) Receptive field (RF) size generally increases as we move up the cortical hierarchy

3) Selectivity pattern is typically apparent at the time first spikes are elicited by a visual stimulus ("feedforward" assumption)

4) There is hierarchy of times at which first spikes are detected.

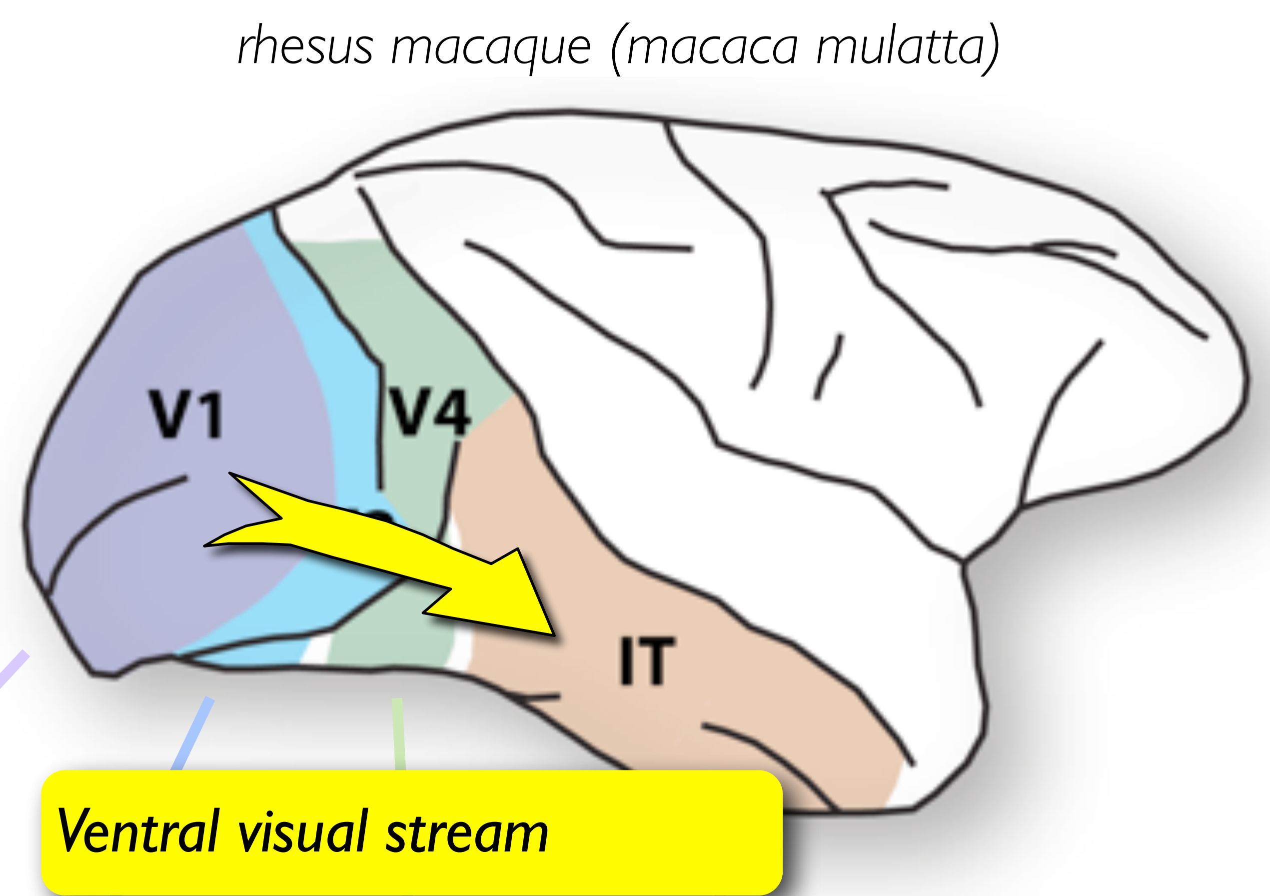


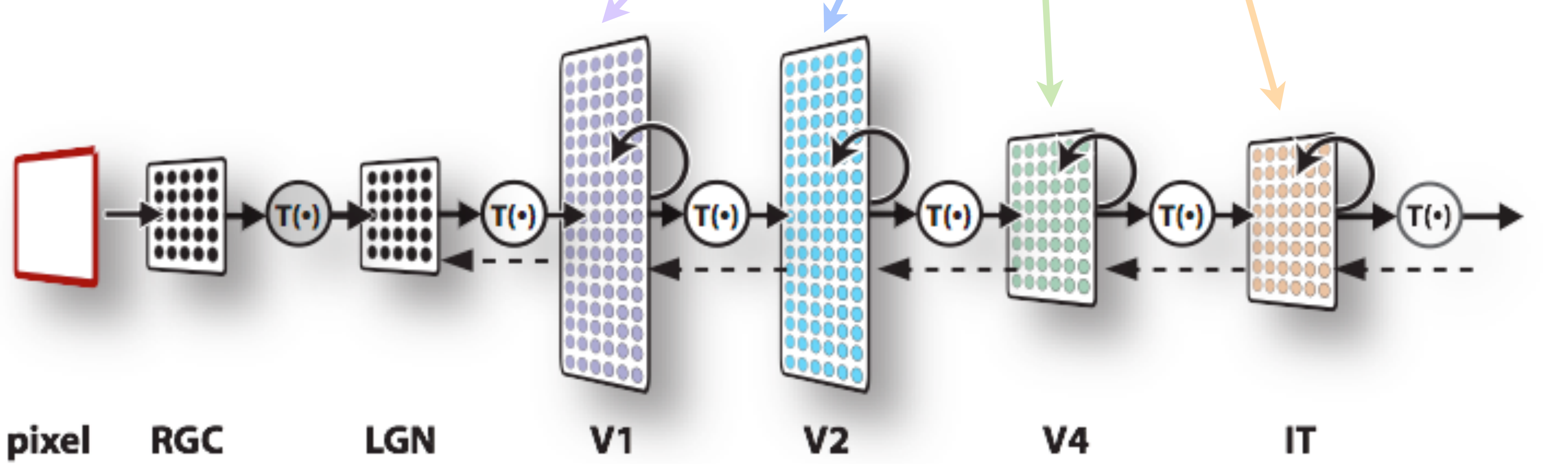


Adapted from DiCarlo et al. 2012

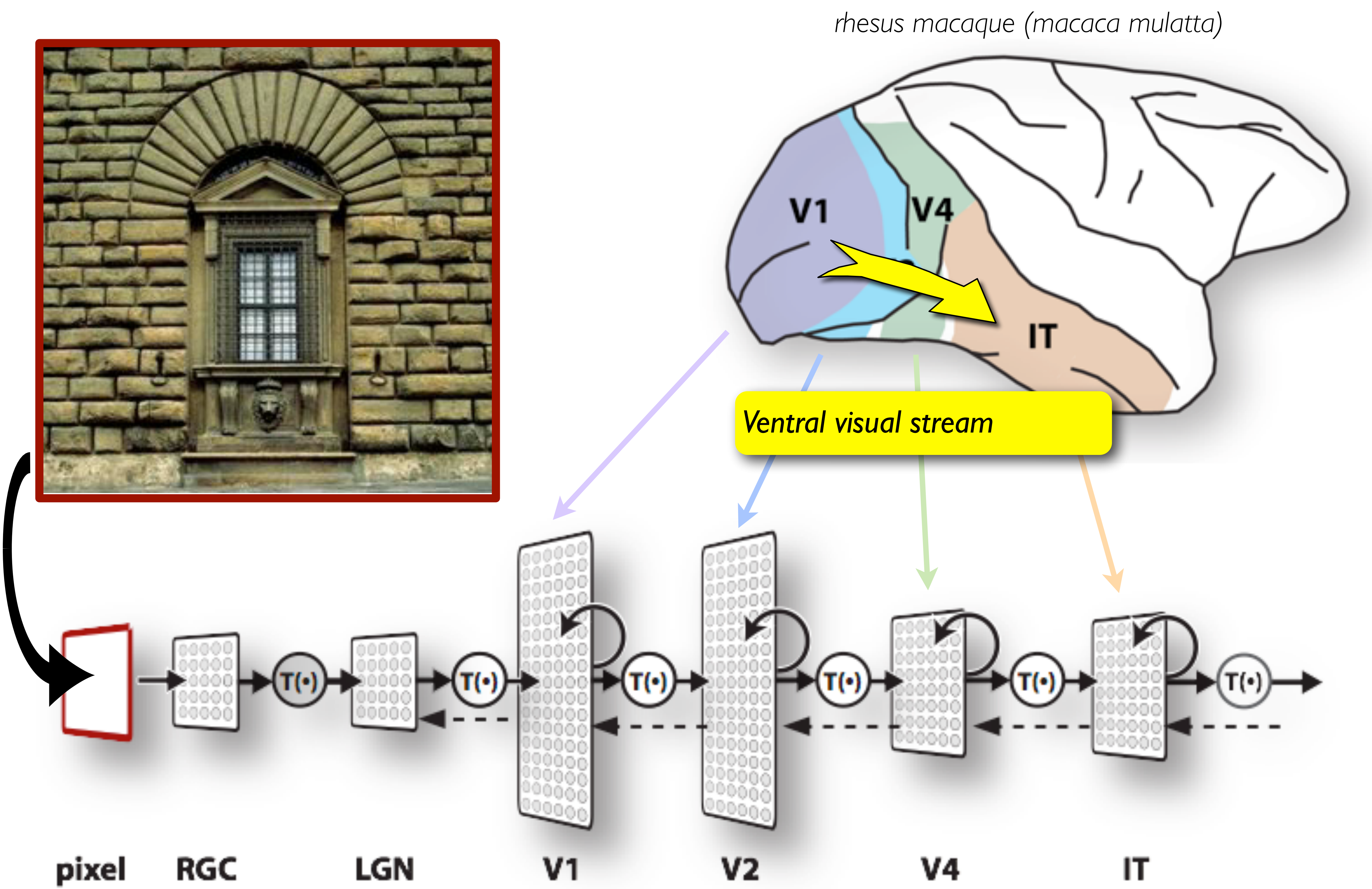
Background: Ventral visual stream

Kaas (2003), Van Essen (2003), Valois and Morgan (1974)
Gross (1973), Mishkin and Ungerleider (1983), Holmes and Gross (1984)
Horel *et al.*,(1987); Freiwald and Tsao (2010), Pitcher, *et al.* (2009)
Yaginuma (1982), Holmes (1984), Weiskrantz (1984), Schiller (1995)
Afraz (2006), Verhoef (2012)
Rust (2010), Freiwald (2010), Lehky (2007)
Majaj (2012)

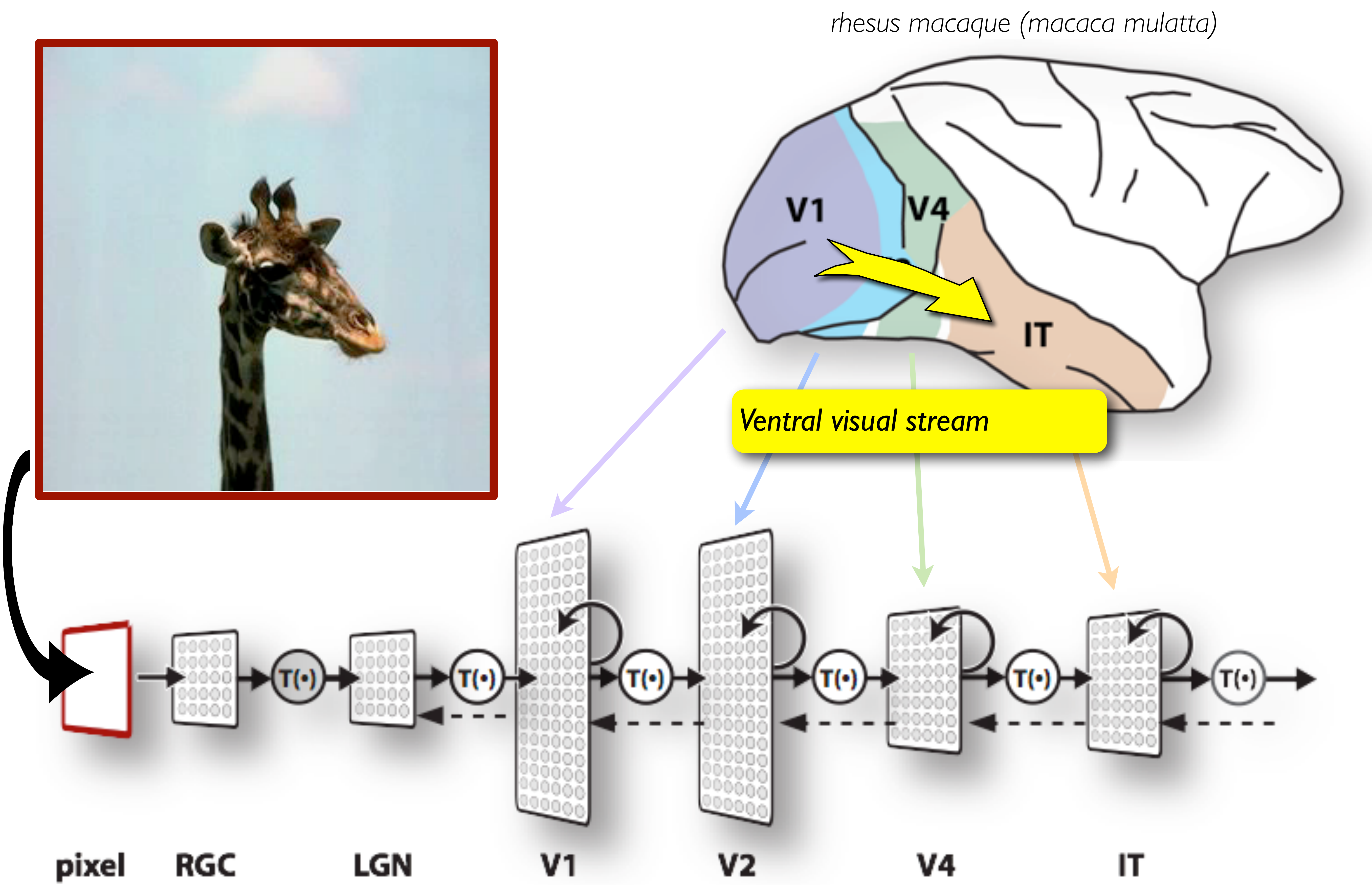




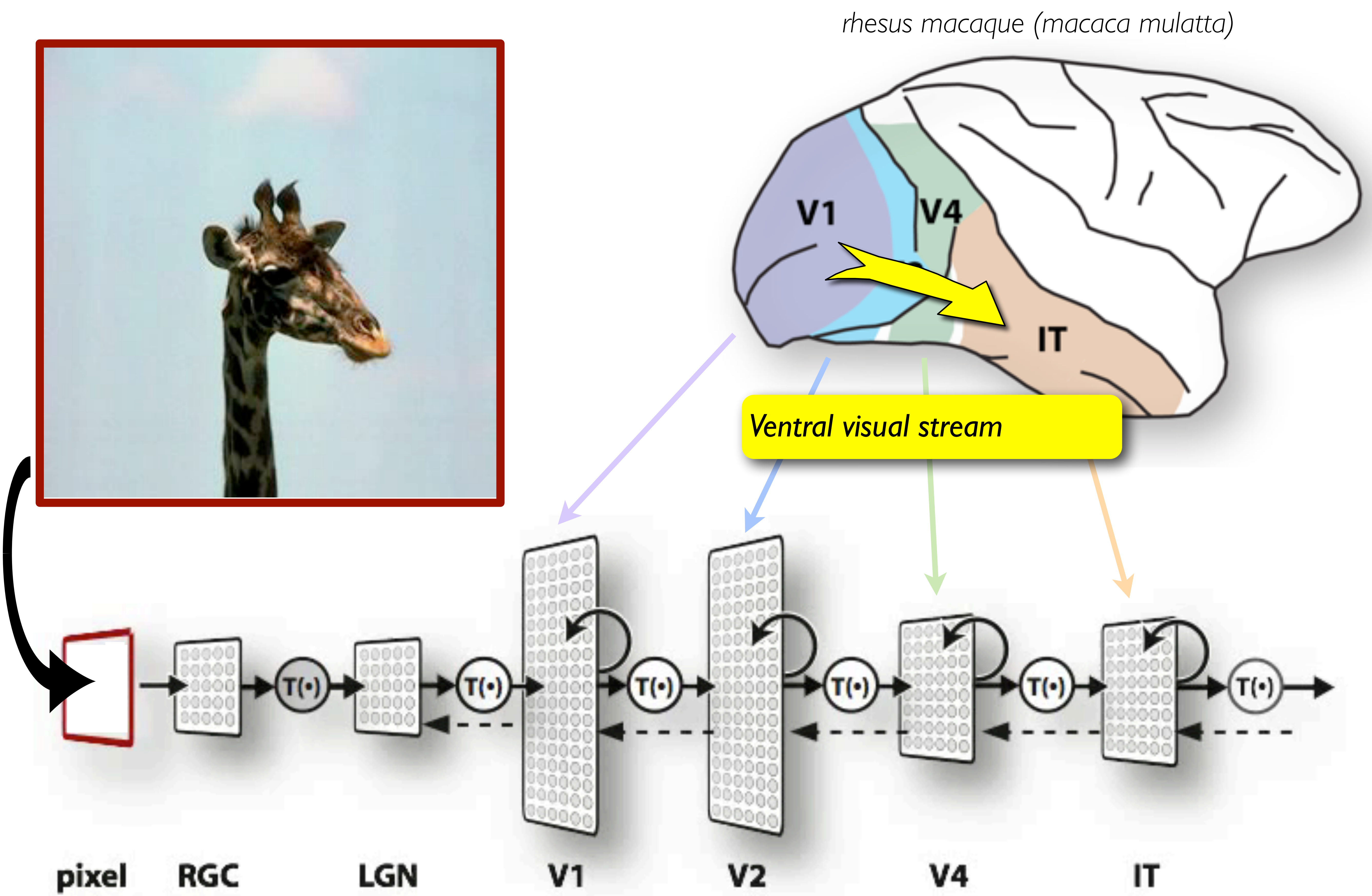
Background: Ventral visual stream



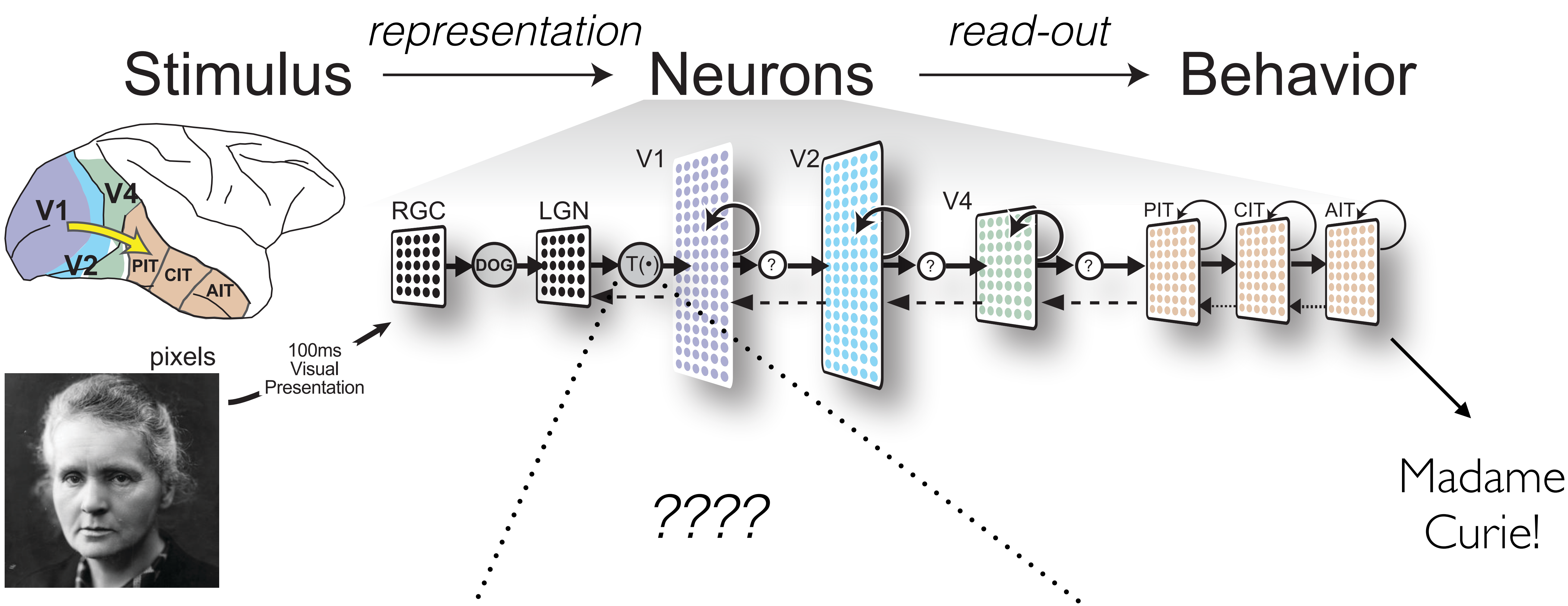
Background: Ventral visual stream



Background: Ventral visual stream

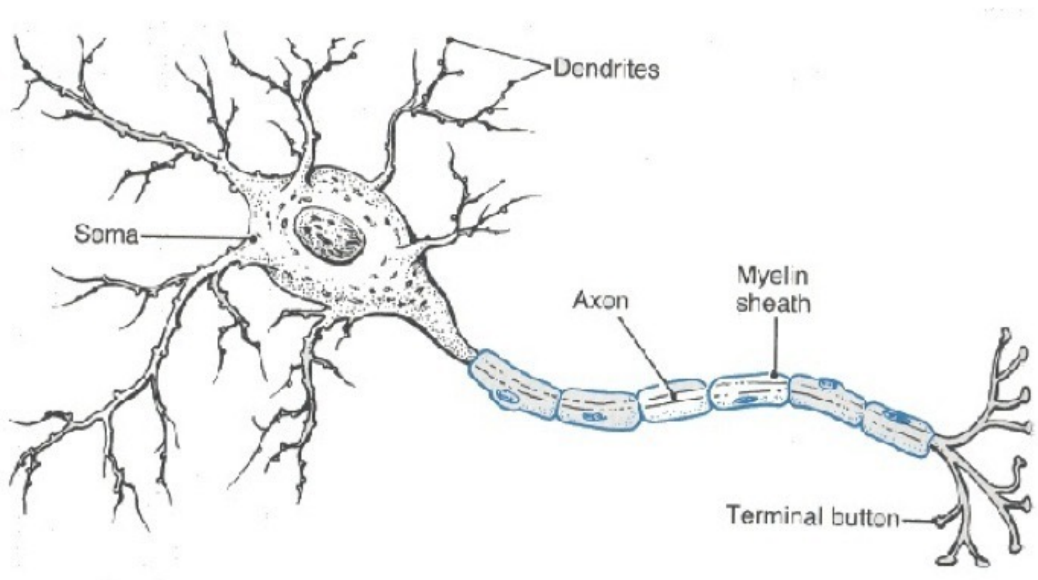


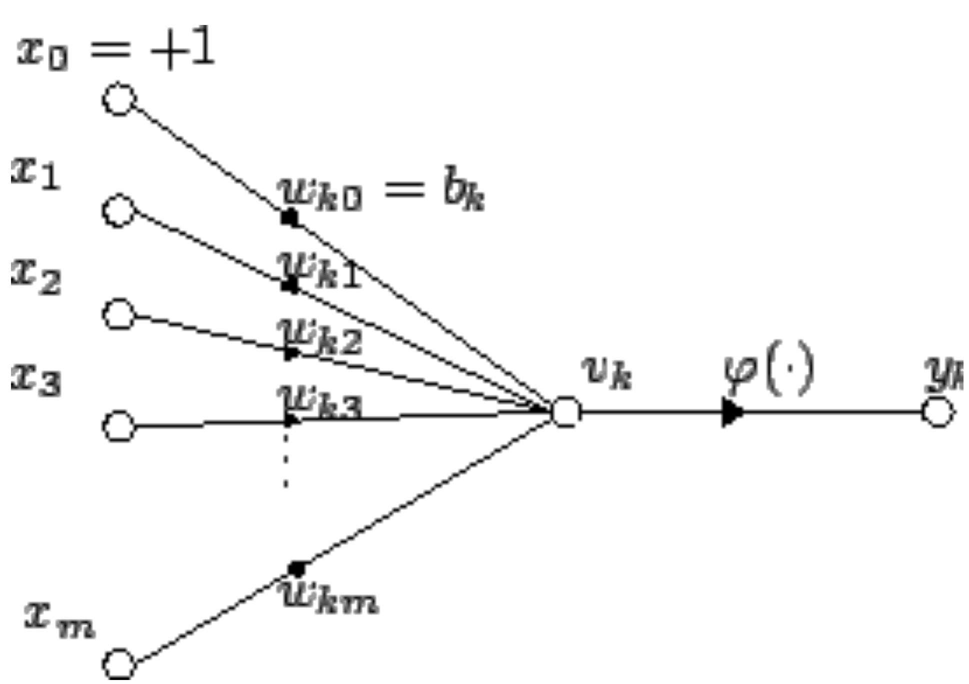
Ventral Stream = Connected series of brain areas



Recall …

McCulloch and Pitts (1943)





$$y_k = \phi\left(\sum_{j=0}^{m} w_{kj} x_j + b_k\right)$$

$$\phi : \mathbb{R} \longmapsto \mathbb{R}$$

some nonlinear activation function

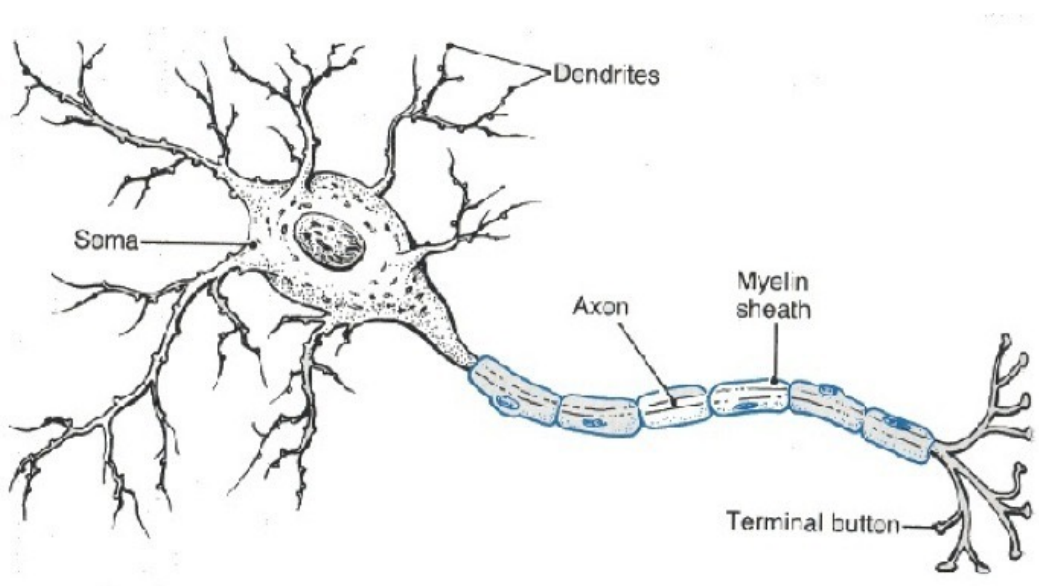
$$w_{kj} \in \mathbb{R}^{m+1}$$

"synaptic strengths"

$$b_j \in \mathbb{R}$$

"biases"

McCulloch and Pitts (1943)

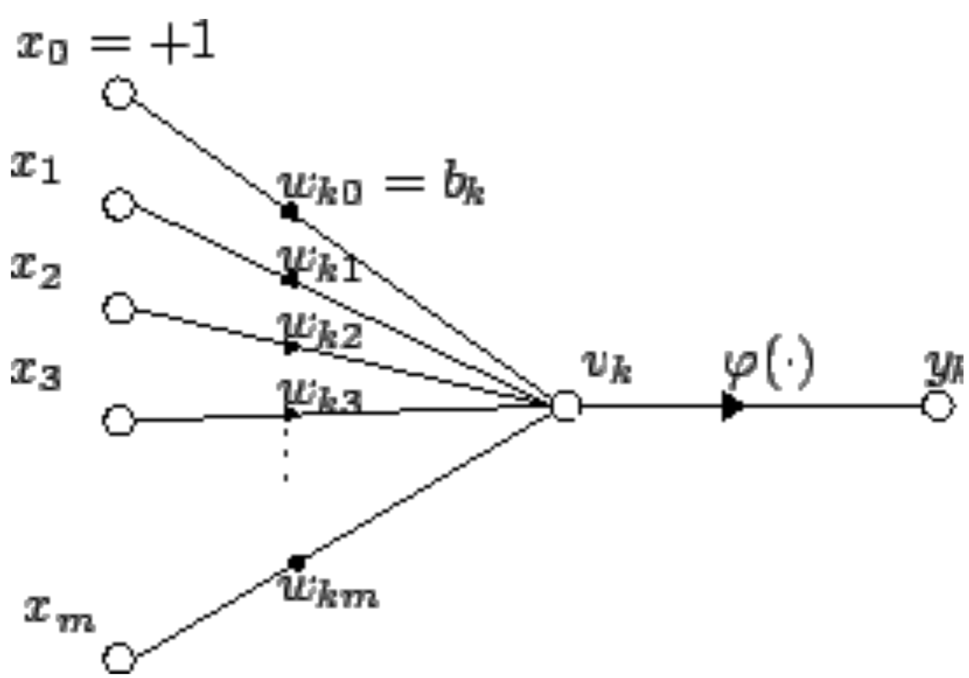


$$y_k = \phi\left(\sum_{j=0}^{m} w_{kj} x_j + b_k\right)$$

???

$\phi : \mathbb{R} \longmapsto \mathbb{R}$

some (nonlinear) activation function



???

$w_{kj} \in \mathbb{R}^{m+1}$

"synaptic strengths"

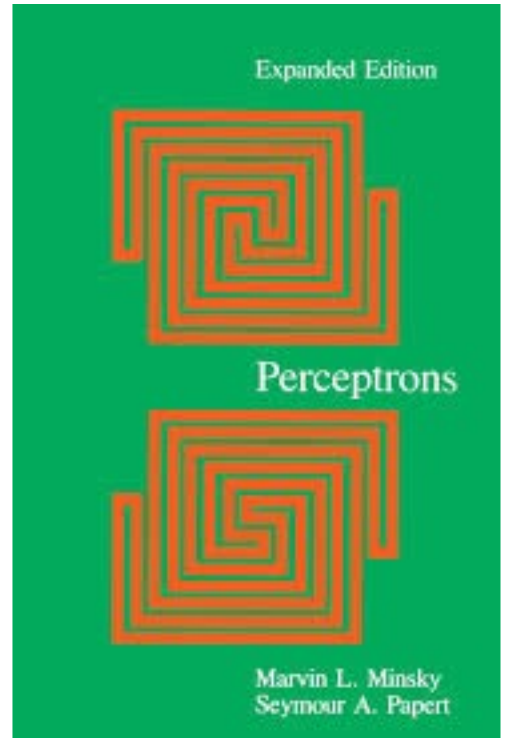
$b_j \in \mathbb{R}$

"biases"

and what's the connectivity?

Cautionary tale

Minsky & Papert (1969)



$$y_k = \phi\left(\sum_{j=0}^{m} w_{kj} x_j + b_k\right)$$

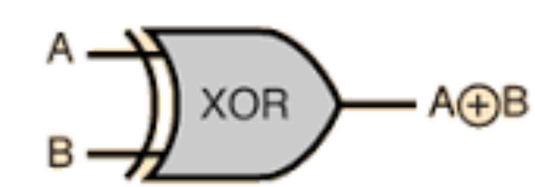
$$\phi : \mathbb{R} \longmapsto \mathbb{R}$$

<u>1. better have more than one layer</u>

<u>2. better be actually nonlinear</u>

$\sum_{i=0}^{N} v_i \phi(w_i^T x + b_i)$ at least *(and which, according to the UAT, is enough)*

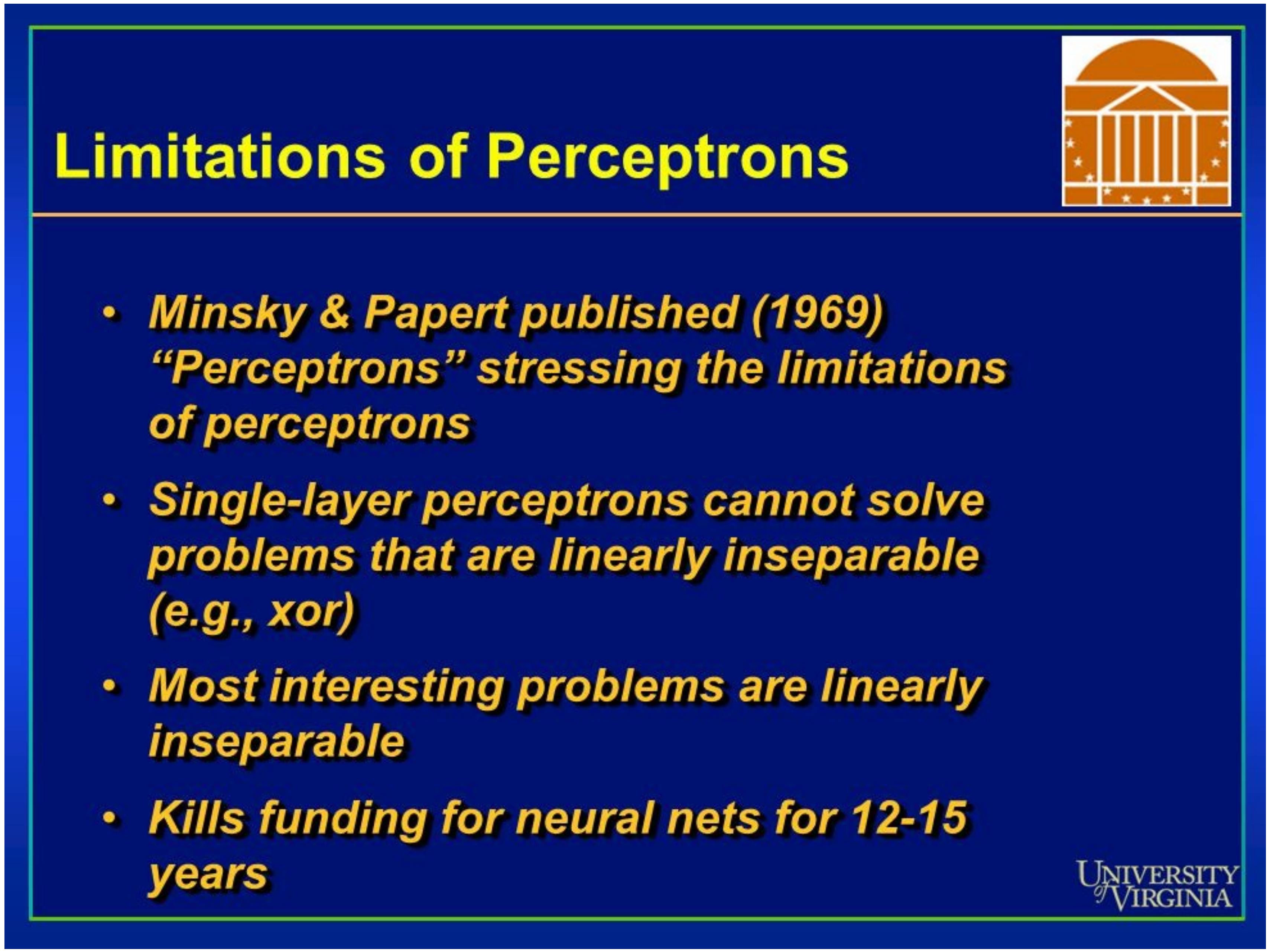
<u>cause otherwise … ain’t no XOR</u>



A	B	Out
0	0	0
0	1	1
1	0	1
1	1	0

<u>and what’s the connectivity?</u>

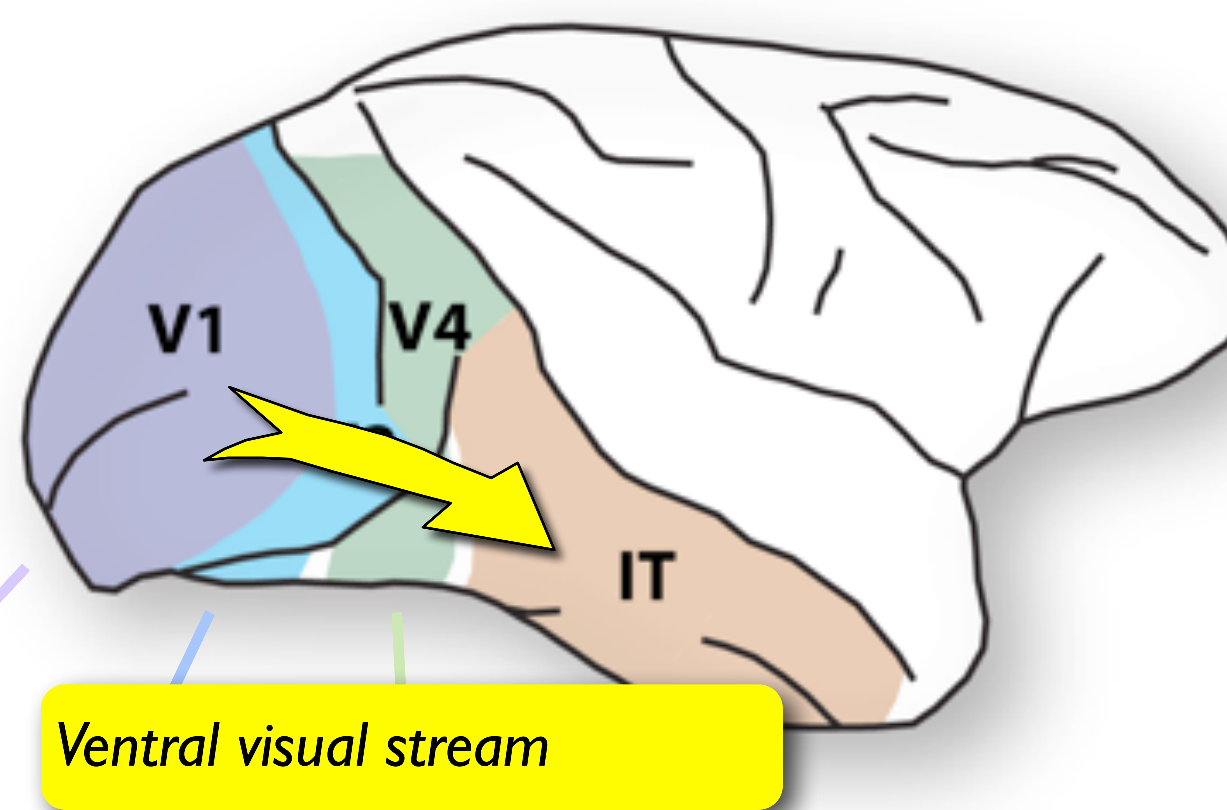
Cautionary tale

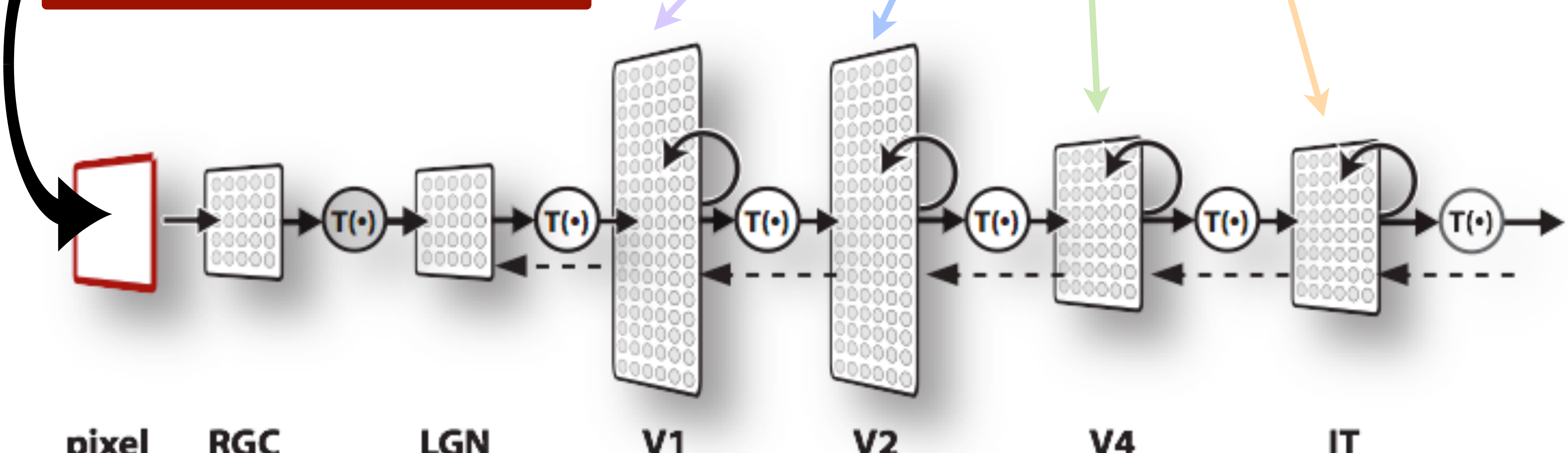


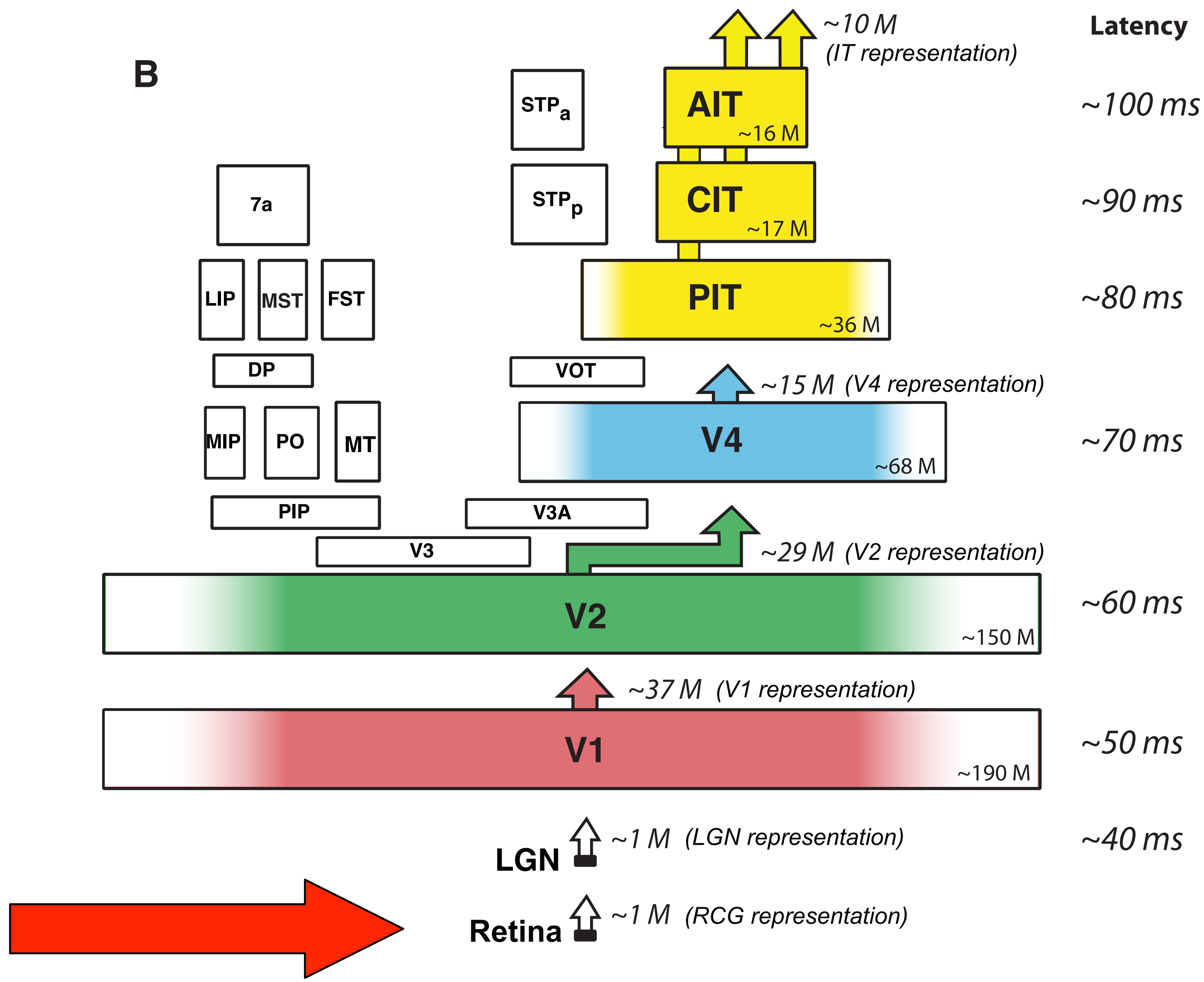
Maybe a bit apocryphal …. but I can definitely say from personal experience that MIT CSAIL felt very "anti-neural networks" as late as 2012

Ventral Stream = Connected series of brain areas

neuroanatomy + neurophysiology tell us:



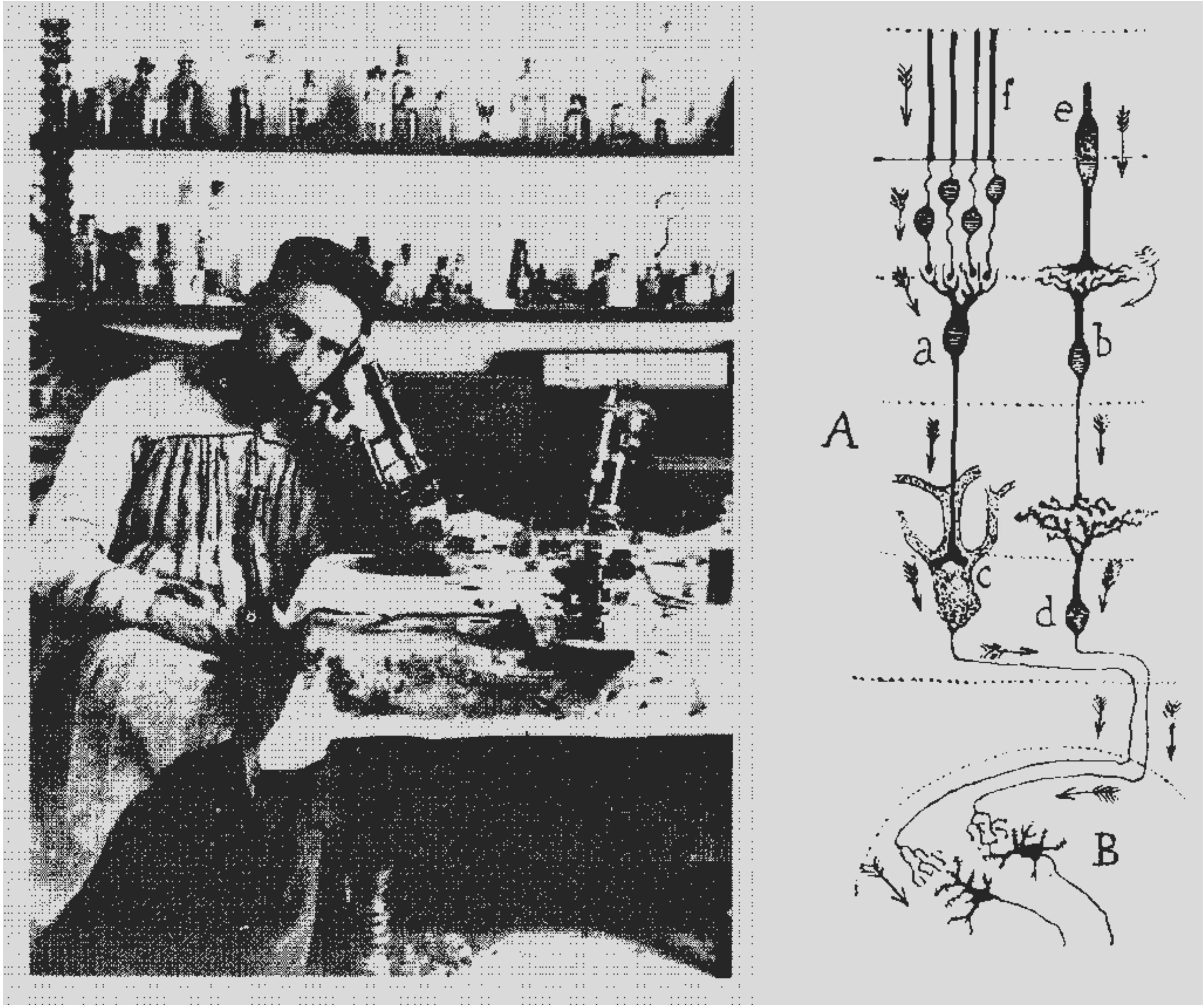




You are here.

Adapted from DiCarlo et al. 2012

Origins in the Retina



Ramon y Cajal *from Rodieck (1973)*

Origins in the Retina

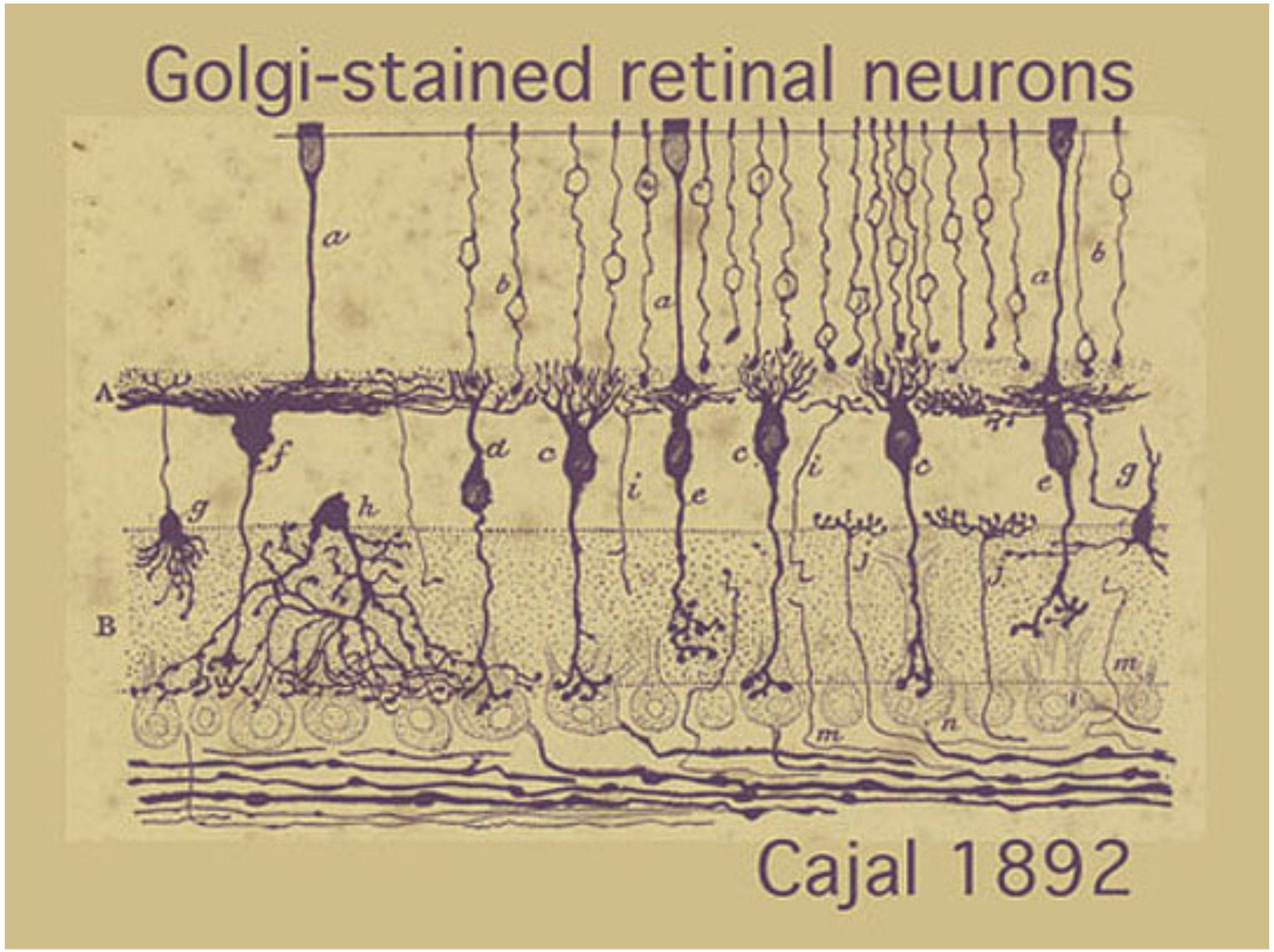


Fig. 2. A drawing done by Cajal to show some of the neurons of the retina in vertical section.

Origins in the Retina

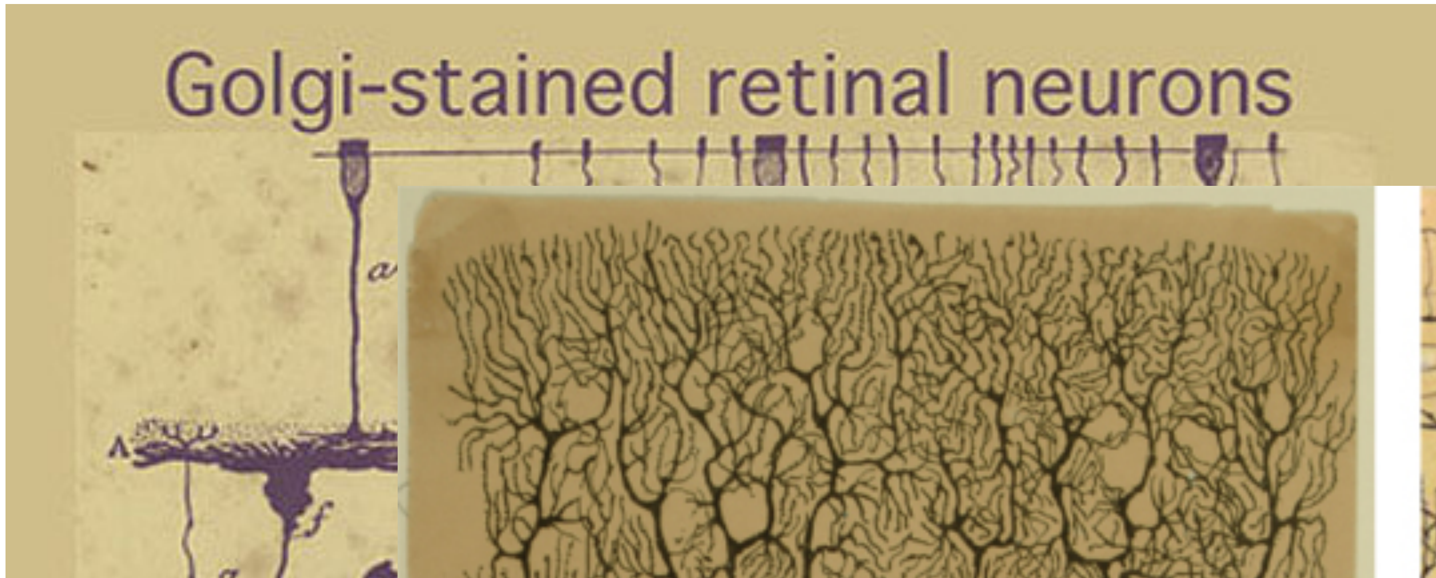
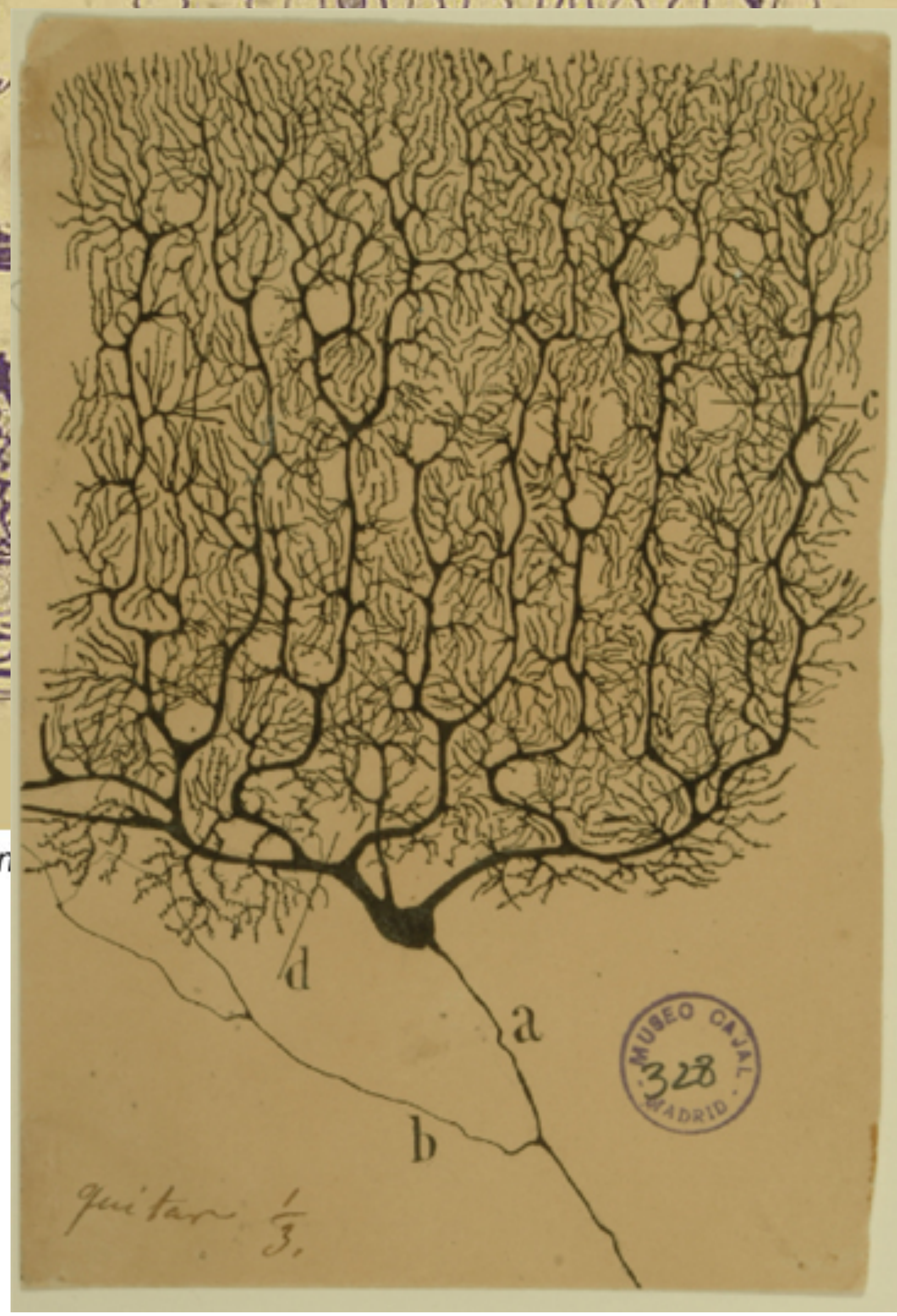
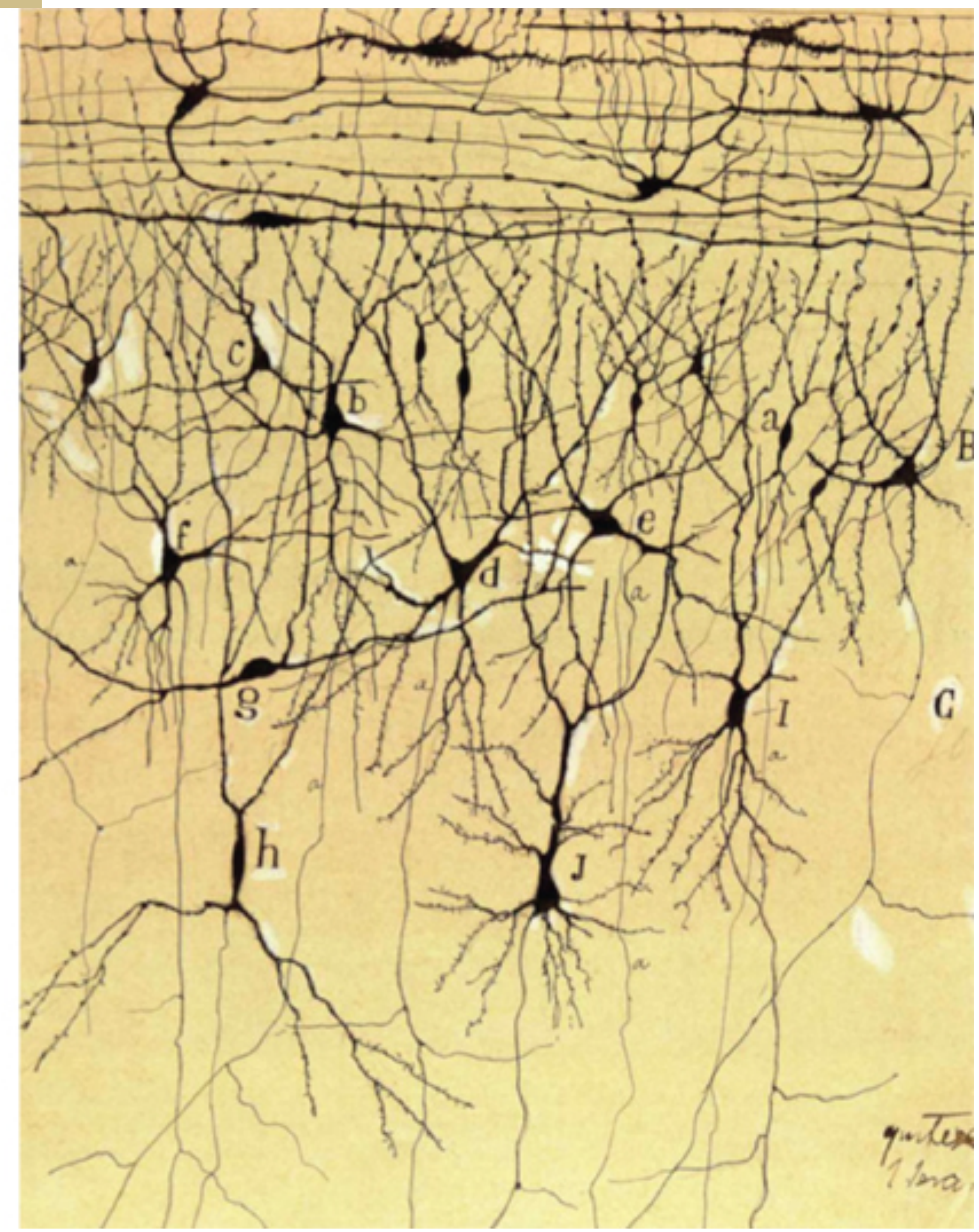
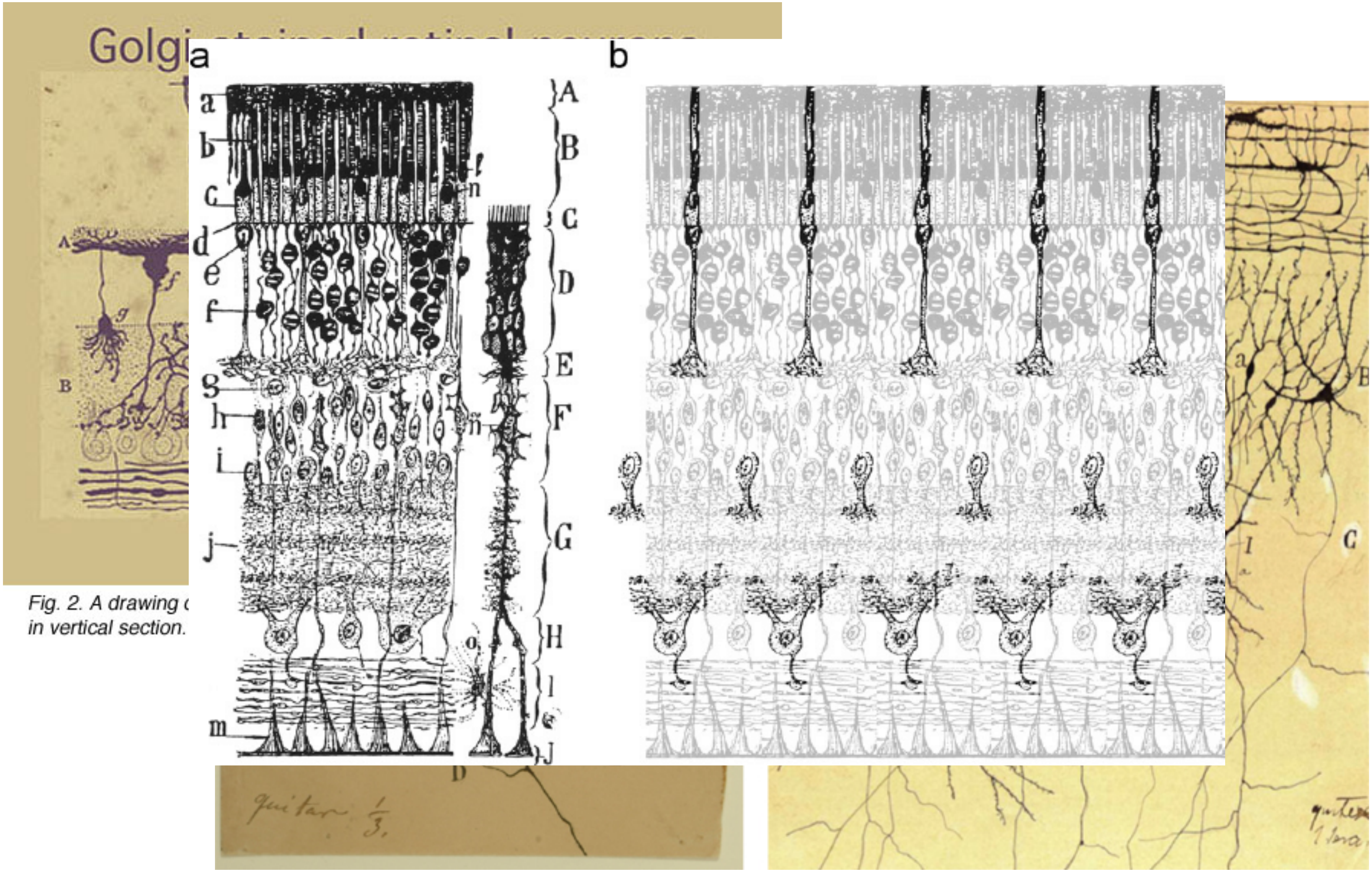


Fig. 2. A drawing don
in vertical section.

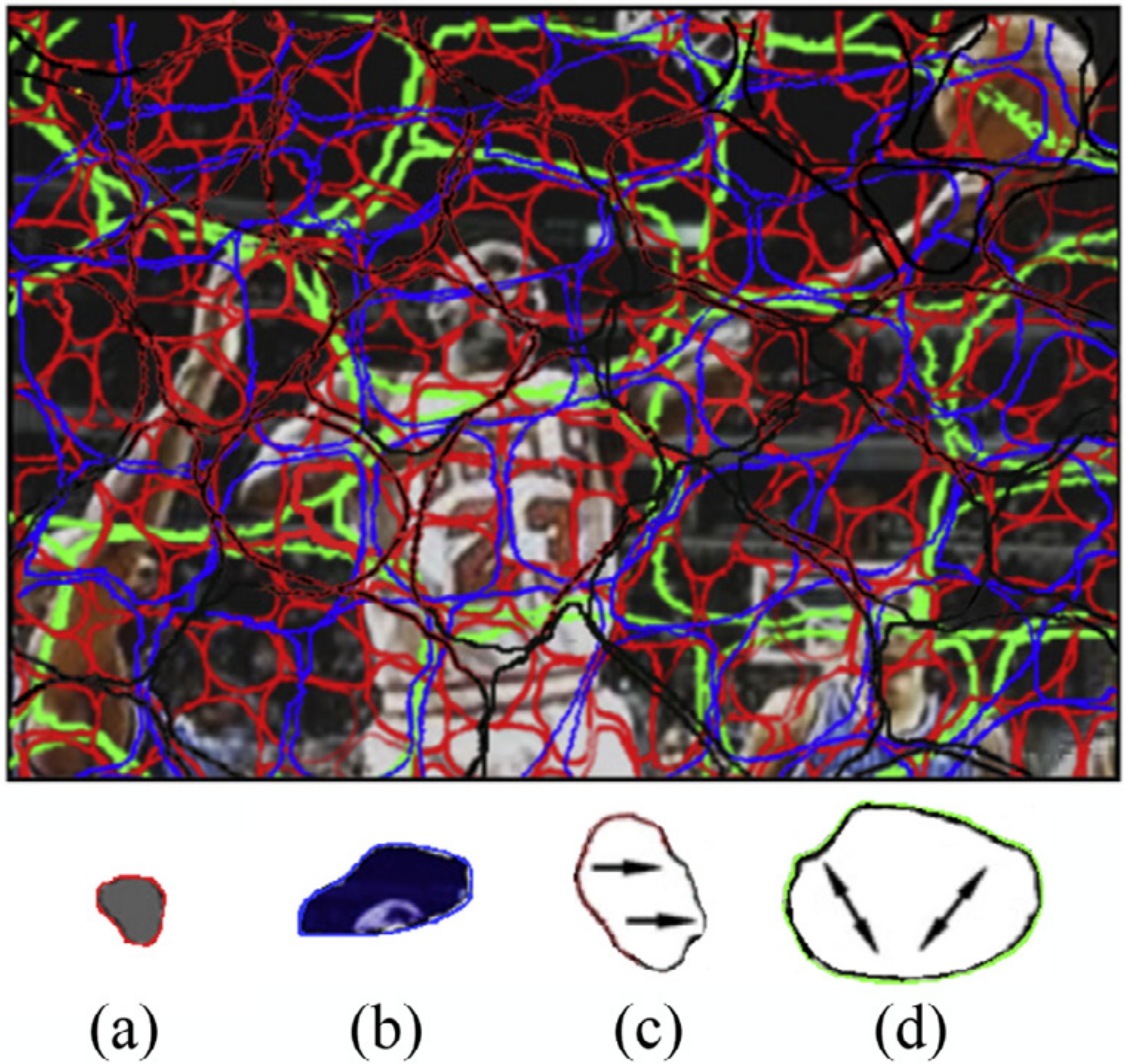




Origins in the Retina

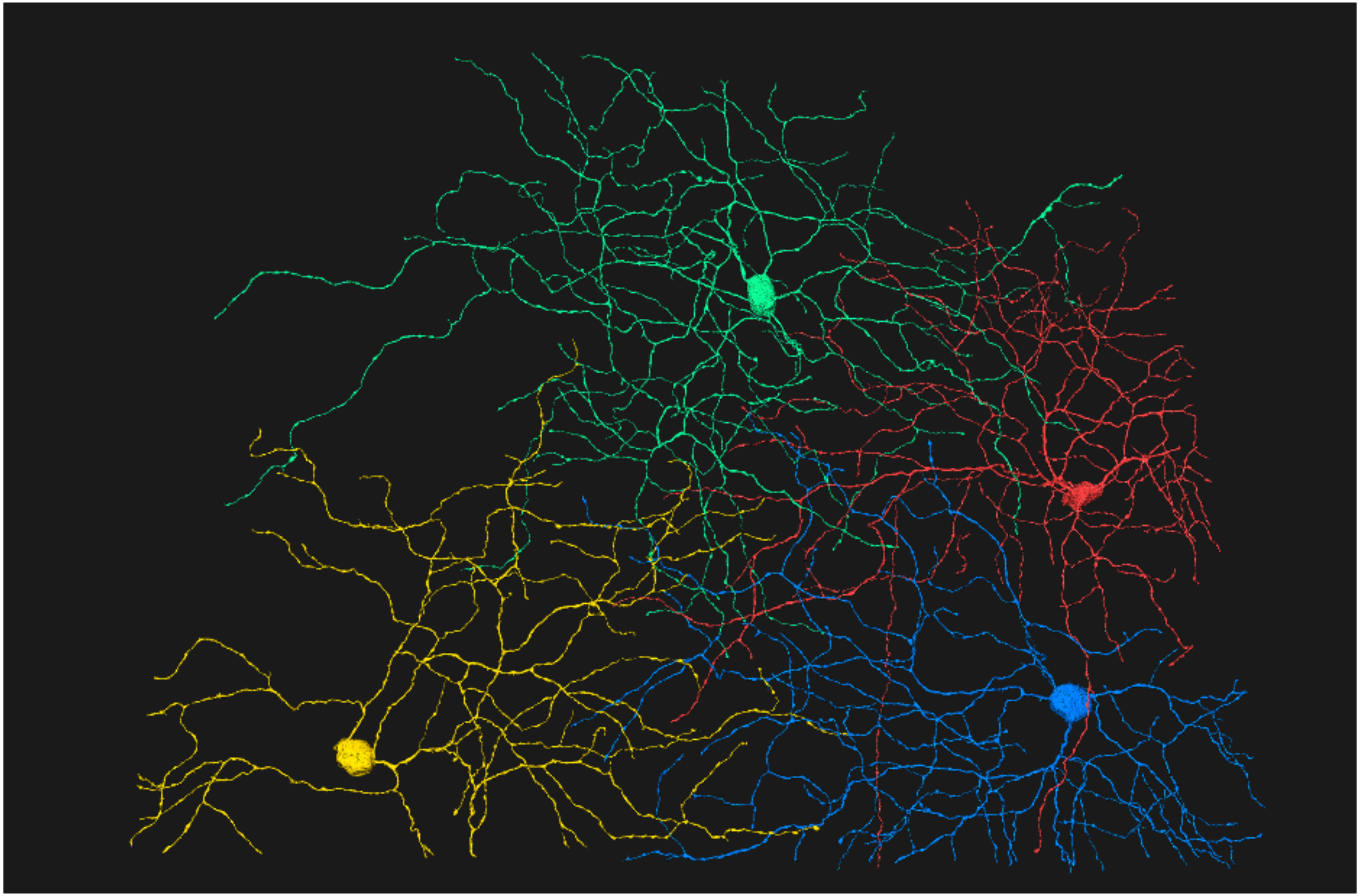


Origins in the Retina



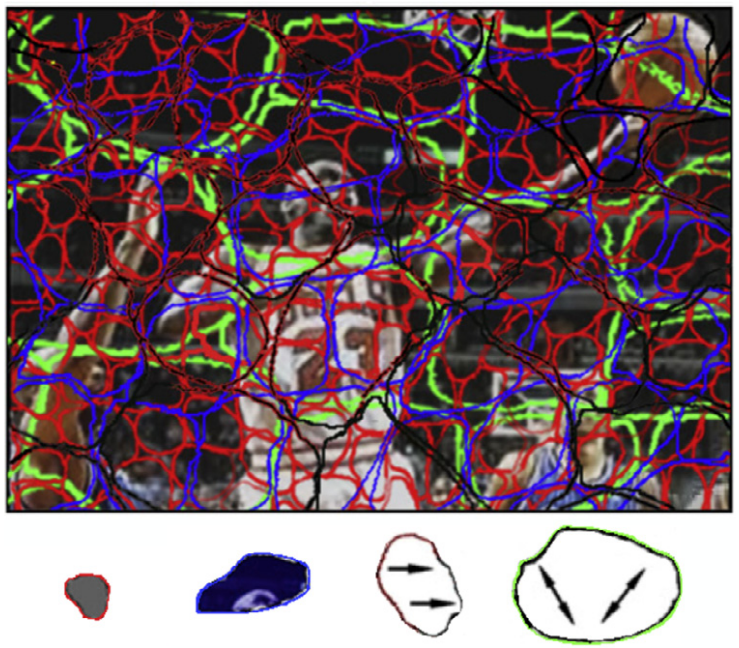
Masland (2012)

Origins in the Retina



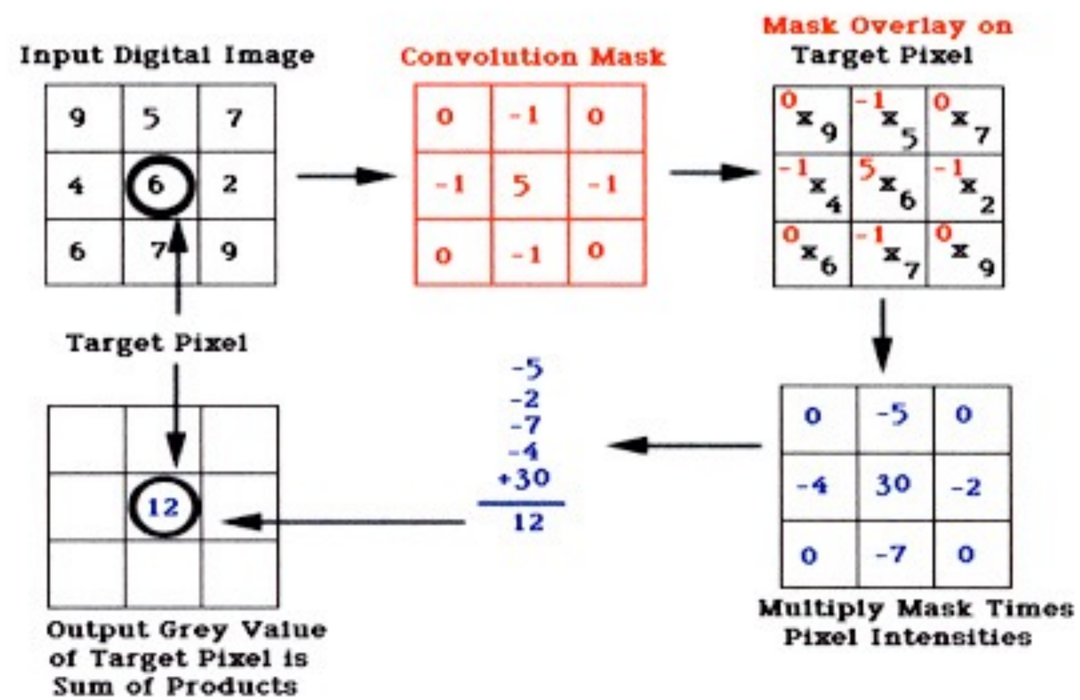
http://museum.eyewire.org

Origins in the Retina



(a) (b) (c) (d)

SPATIAL CONVOLUTION



1. The convolution mask is overlated on the original image so that the center pixel of the mask is matched with a pixel location on the image (Target Pixel- to be convolved).

2. Each pixel value in the original image is multiplied by the corresponding value in the overlying mask..

3. The grey value of the target pixel is replaced by the sum of all the products in the second step.

4. The operation is repeated for each pixel in the original image (the mask scans the entire image) and each pixel is replaced by the weighted average of its 3 x 3 neighbors.

Origins in the Retina

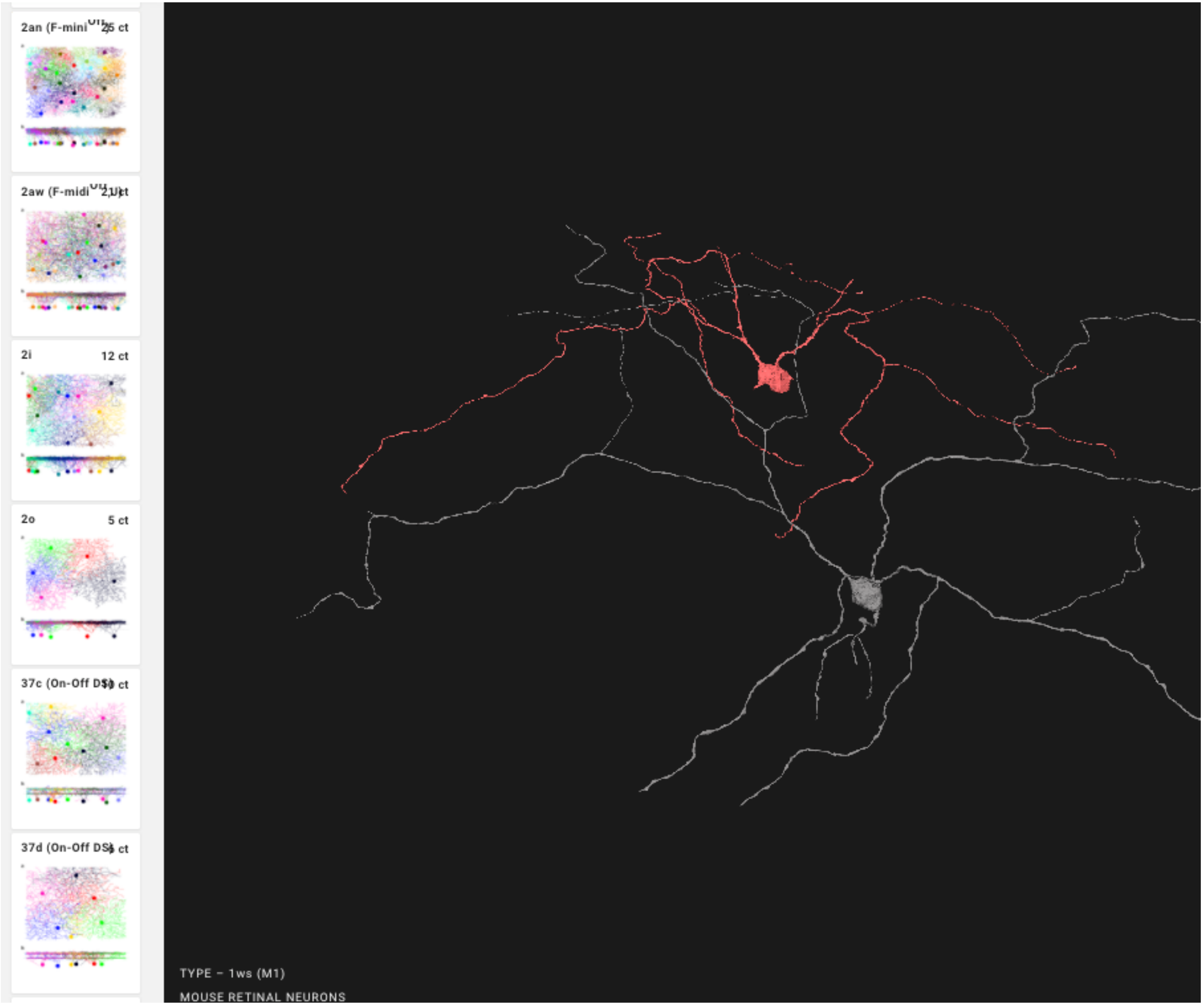
cell types like different filters in a filterbanks

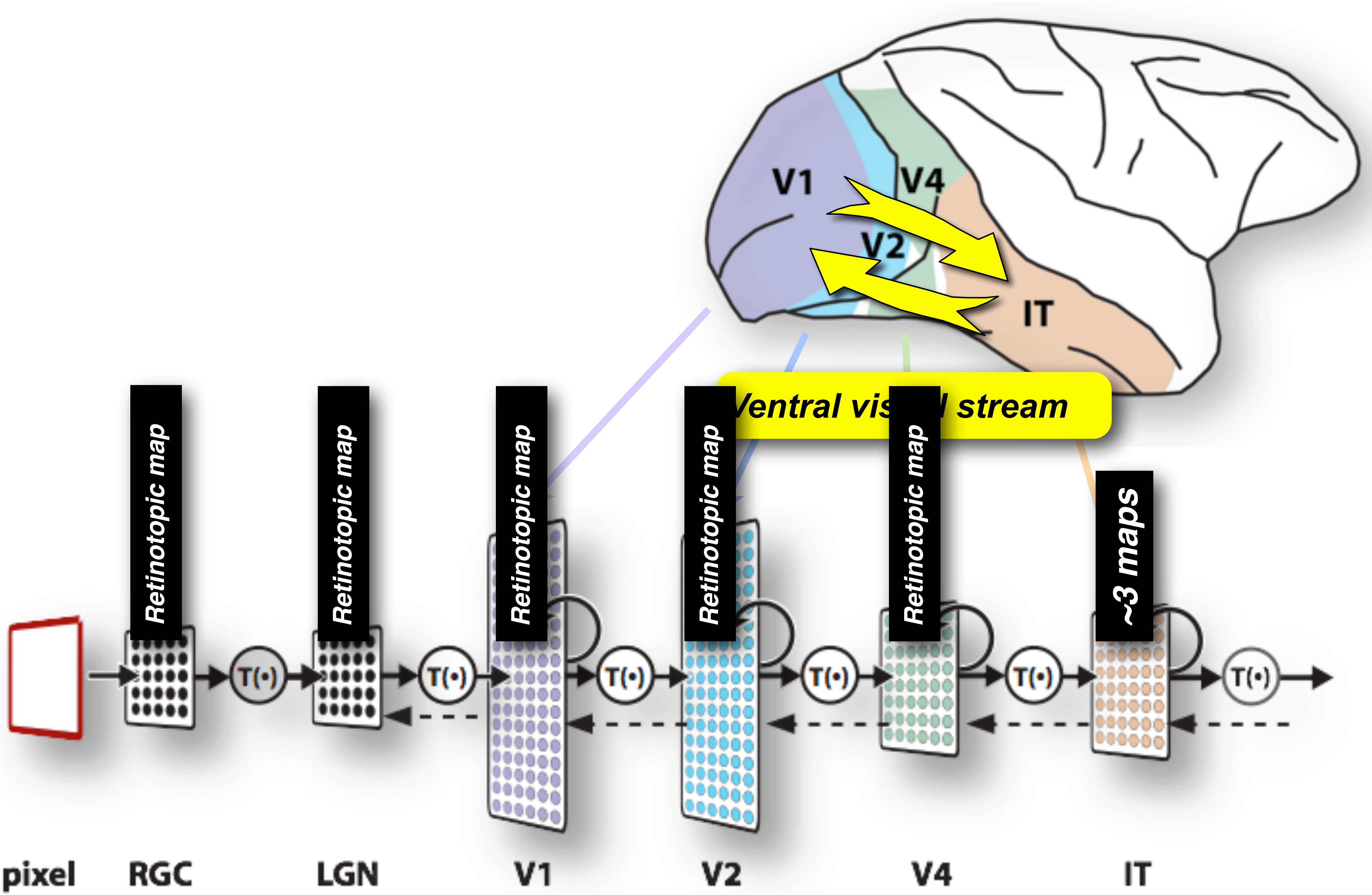
.

.

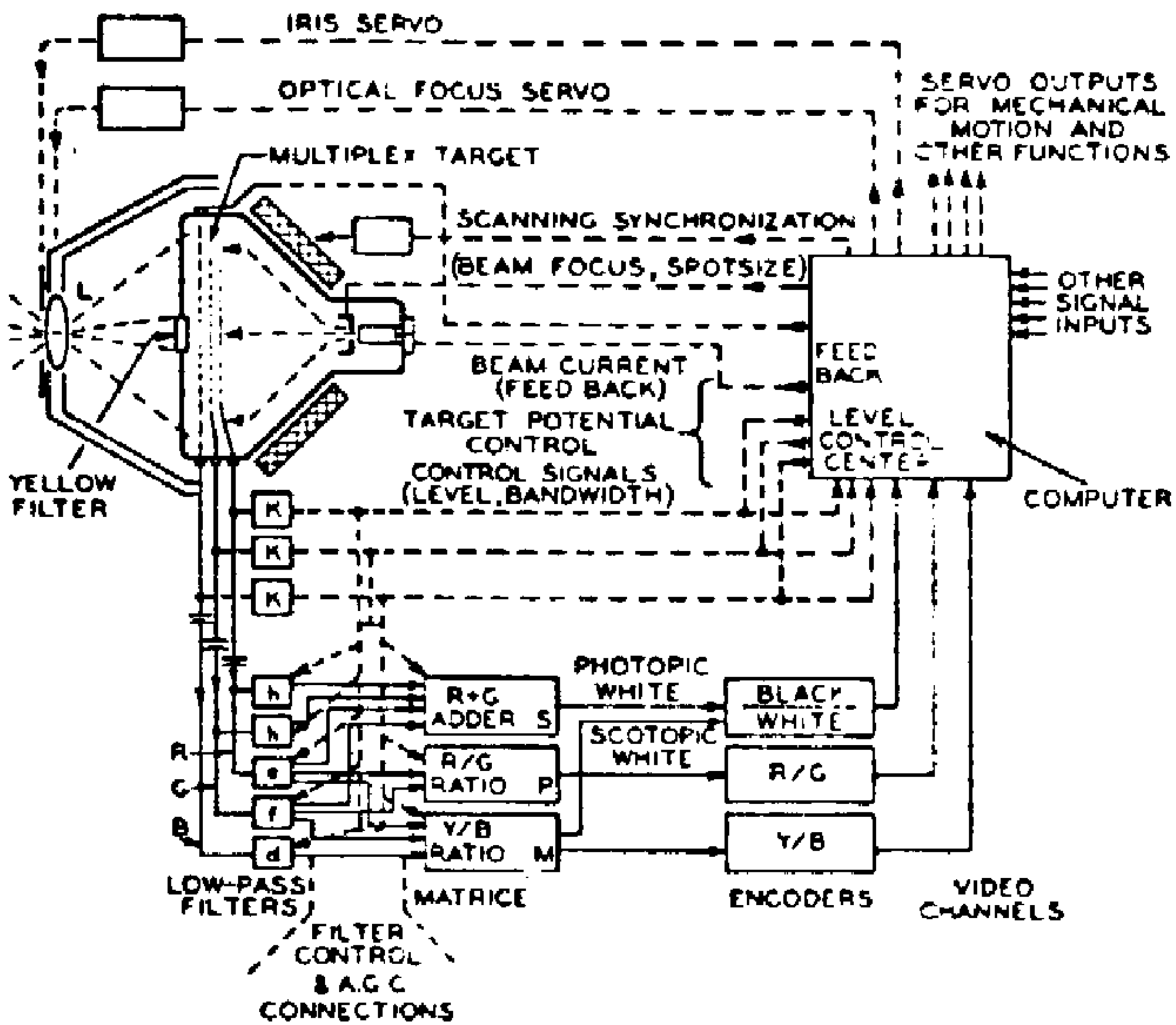
.

but which filters?




V1
V4
V2
IT
Ventral vis
l stream
Retinotopic map
Retinotopic map
Retinotopic map
Retinotopic map
Retinotopic map
~3 maps
T(•)
pixel
RGC
LGN
V1
V2
V4
IT

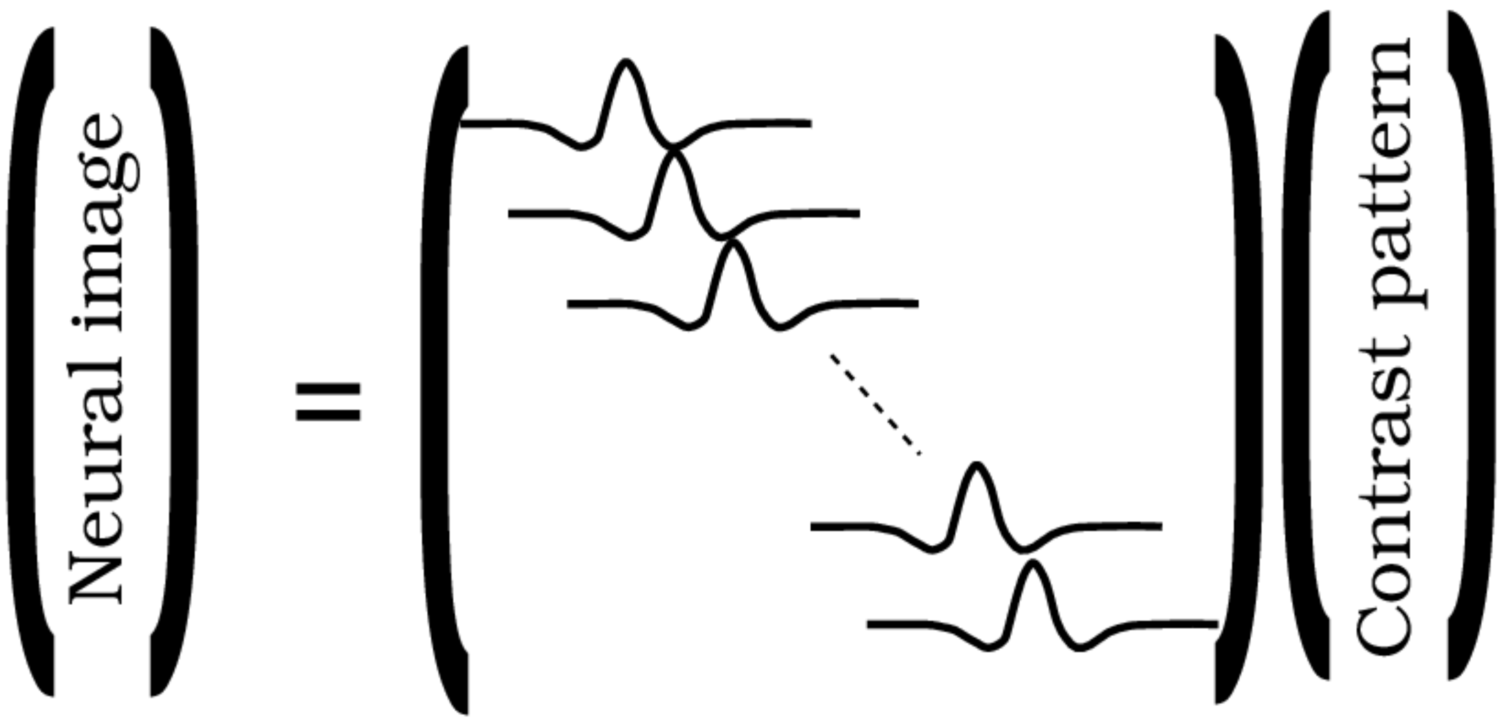
characterizing a *transfer function . . .*



Schade, 1956 from Wandell 1996

Origins in the Retina

characterizing a *transfer function* . . .

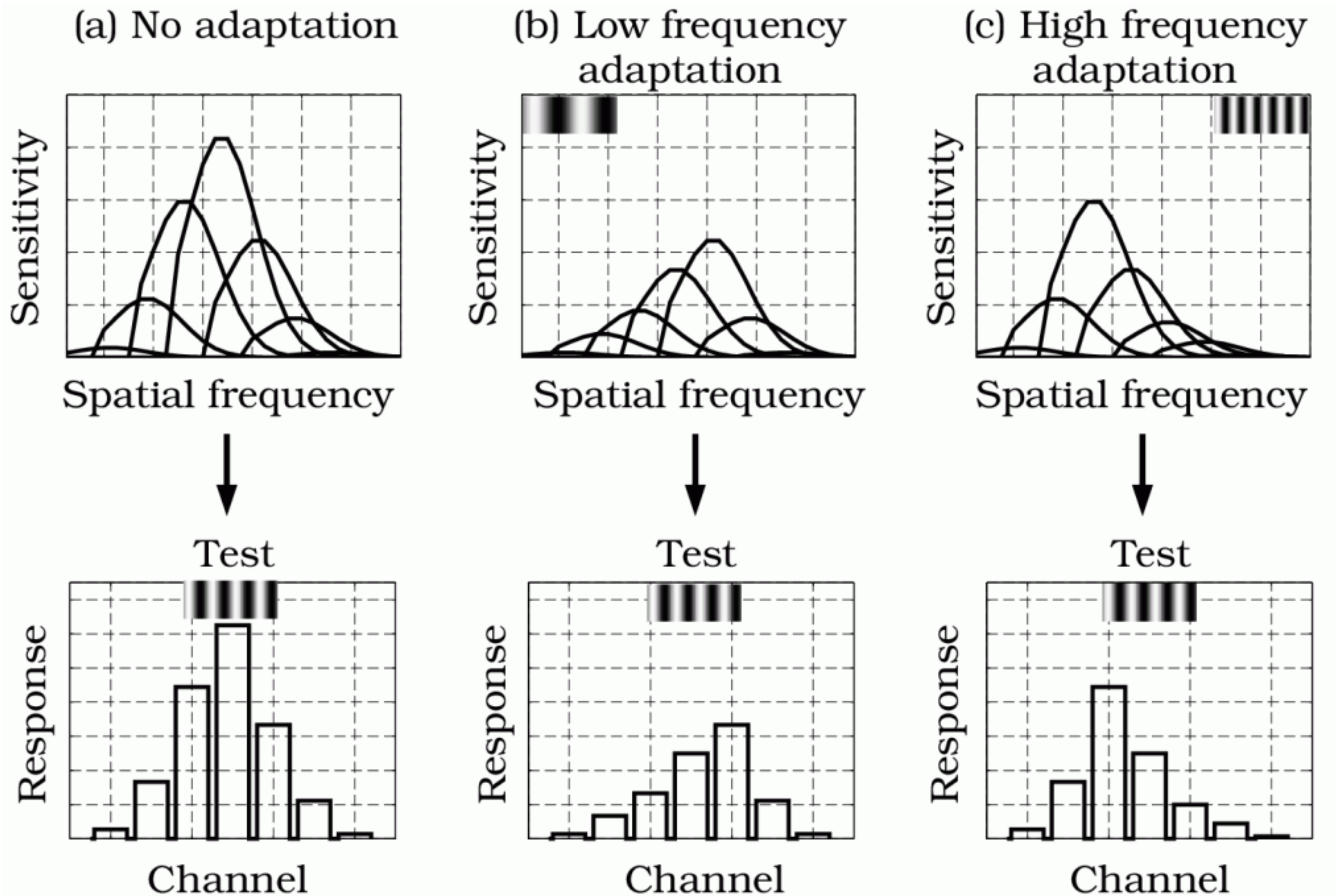


Christina Enroth-Cugell

John Robson

. . . and thus, presumably, doing linear systems (e.g fourier) analysis

Origins in the Retina



from Wandell 1996

Origins in the Retina

THE CONTRAST SENSITIVITY OF RETINAL GANGLION CELLS OF THE CAT

BY CHRISTINA ENROTH-CUGELL AND J. G. ROBSON*
From the Biomedical Engineering Center, Technological Institute, Northwestern University, Evanston, Illinois, U.S.A.† and the Department of Physiology, Northwestern University Medical School, Chicago, U.S.A.

(*Received* 19 *April* 1966)

Christina Enroth-Cugell

John Robson

1. Spatial summation within cat retinal receptive fields was studied by recording … responses of ganglion cells to grating patterns

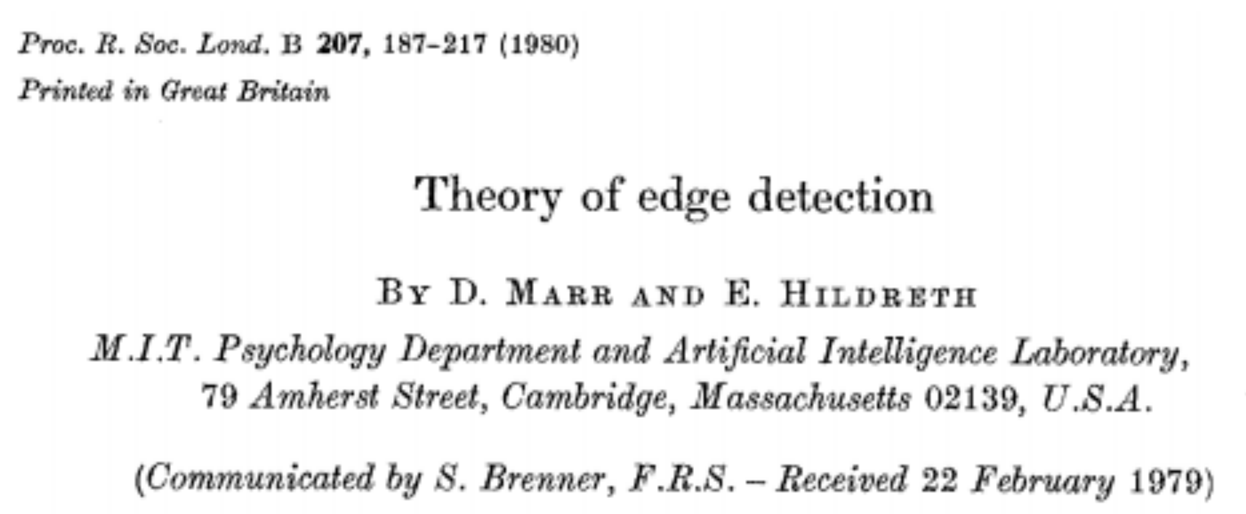
2. Summation over the receptive fields of some cells (X-cells) was found to be **approximately linear**, while for other cells (Y-cells) summation was **very non-linear**.

3. The mean discharge frequency of Y-cells … was greatly increased when grating patterns drifted across their receptive fields.

4. In X-cells … it was found that the contrast sensitivity function, **could be satisfactorily described by the difference of two Gaussian functions**.

5. This finding supports the hypothesis that the sensitivities of the antagonistic centre and surround summating regions of ganglion cell receptive fields fall off as Gaussian functions of the distance from the field centre.

Old-School CV: Marr-Hildreth's "Laplacian of Gaussians"

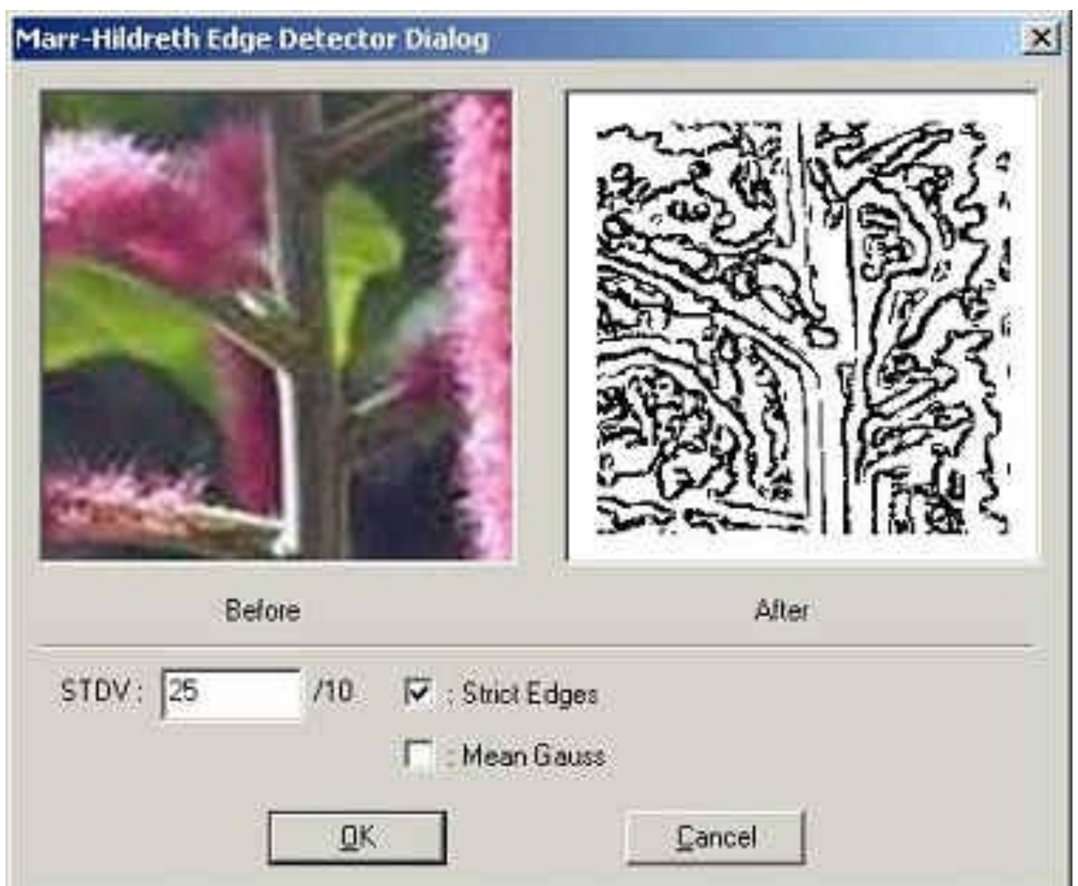


Proc. R. Soc. Lond. B **207**, 187–217 (1980)
Printed in Great Britain

Theory of edge detection

BY D. MARR AND E. HILDRETH
M.I.T. Psychology Department and Artificial Intelligence Laboratory,
79 *Amherst Street, Cambridge, Massachusetts* 02139, *U.S.A.*

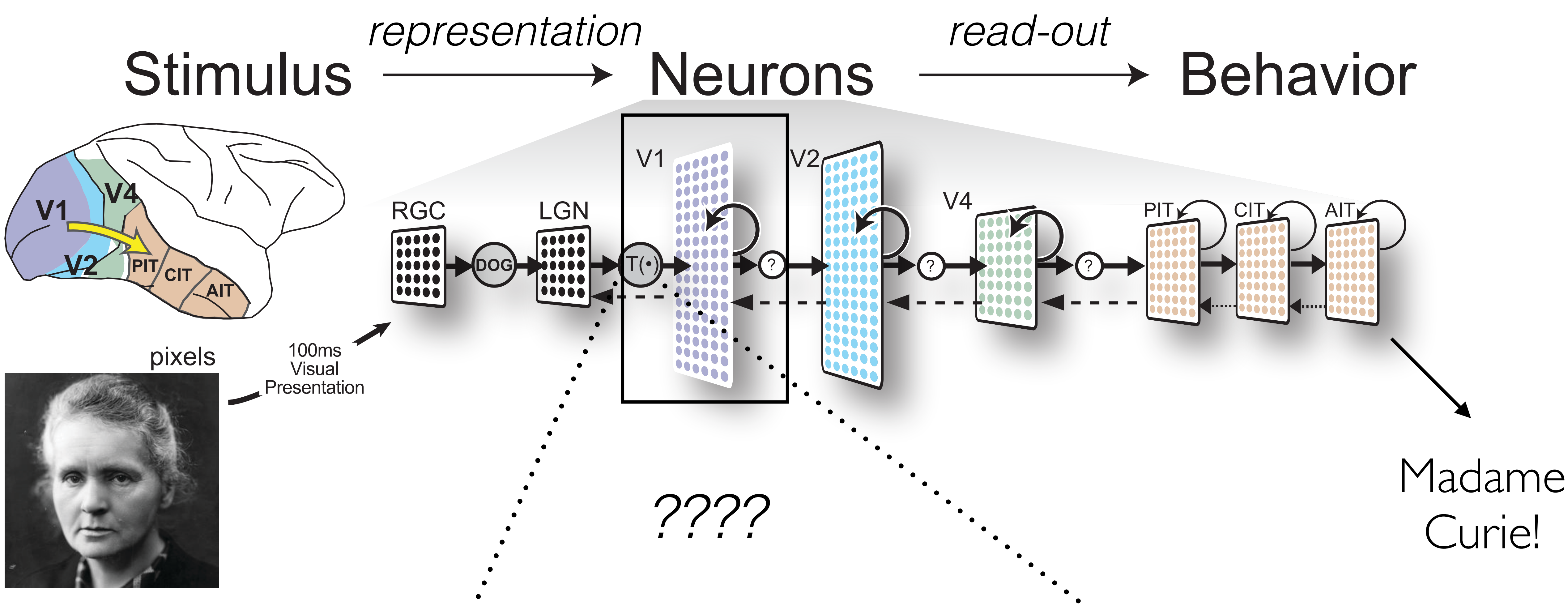
(*Communicated by S. Brenner, F.R.S. – Received* 22 *February* 1979)



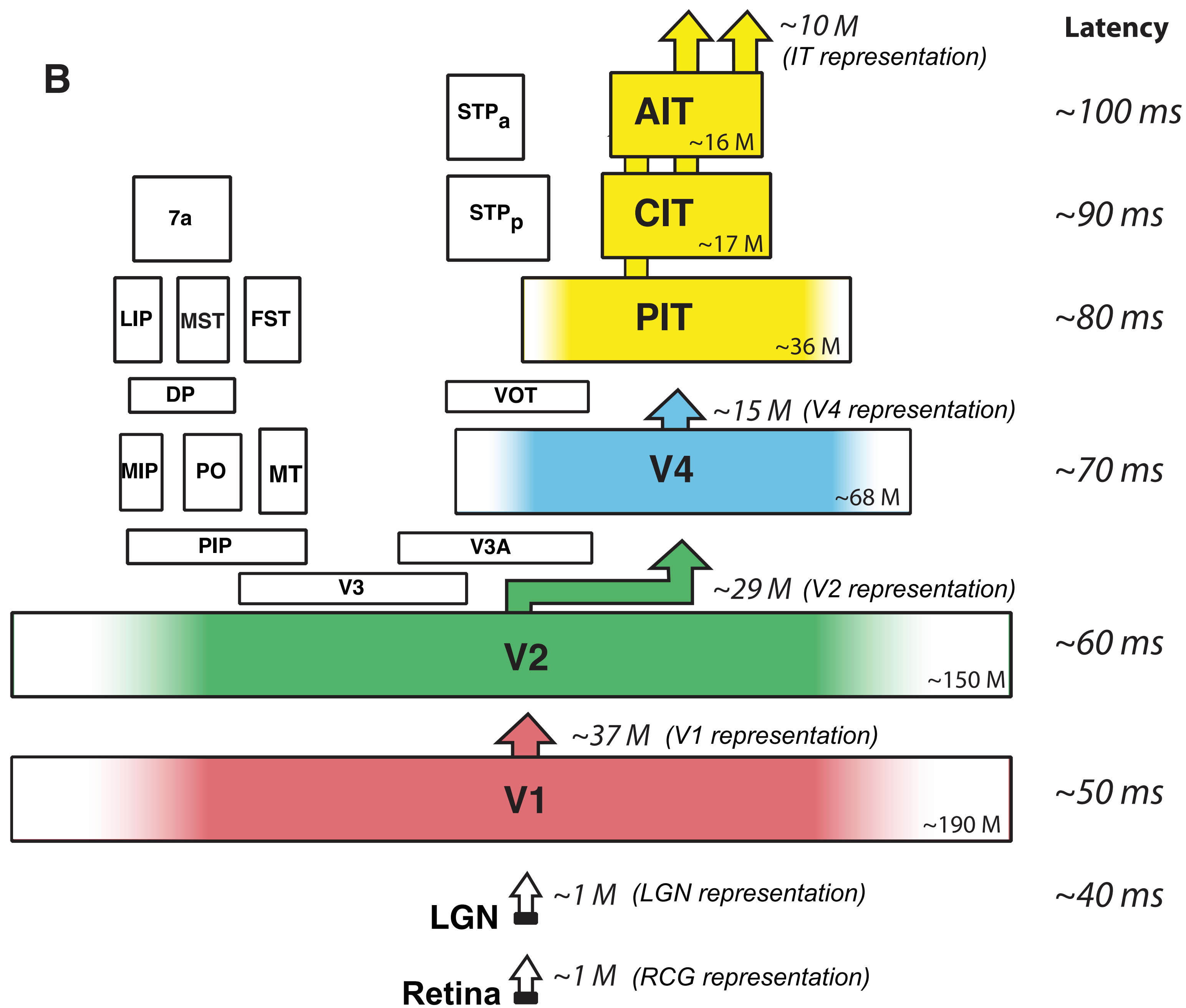
$$\vec{\nabla}^2 G(x, y) * Im(x, y)$$

$$\sim DoG$$

Gabors in V1


representation
read-out
Stimulus
Neurons
Behavior
V1
V4
V2
PIT
CIT
AIT
pixels
100ms
Visual
Presentation
RGC
DOG
LGN
T(•)
V1
V2
V4
PIT
CIT
AIT
????
Madame
Curie!

You are here.



Adapted from DiCarlo et al. 2012

Gabors in V1

J. Physiol. (1962), **160**, *pp.* 106–154
With 2 plates and 20 text-figures
Printed in Great Britain

RECEPTIVE FIELDS, BINOCULAR INTERACTION AND FUNCTIONAL ARCHITECTURE IN THE CAT'S VISUAL CORTEX

BY D. H. HUBEL AND T. N. WIESEL

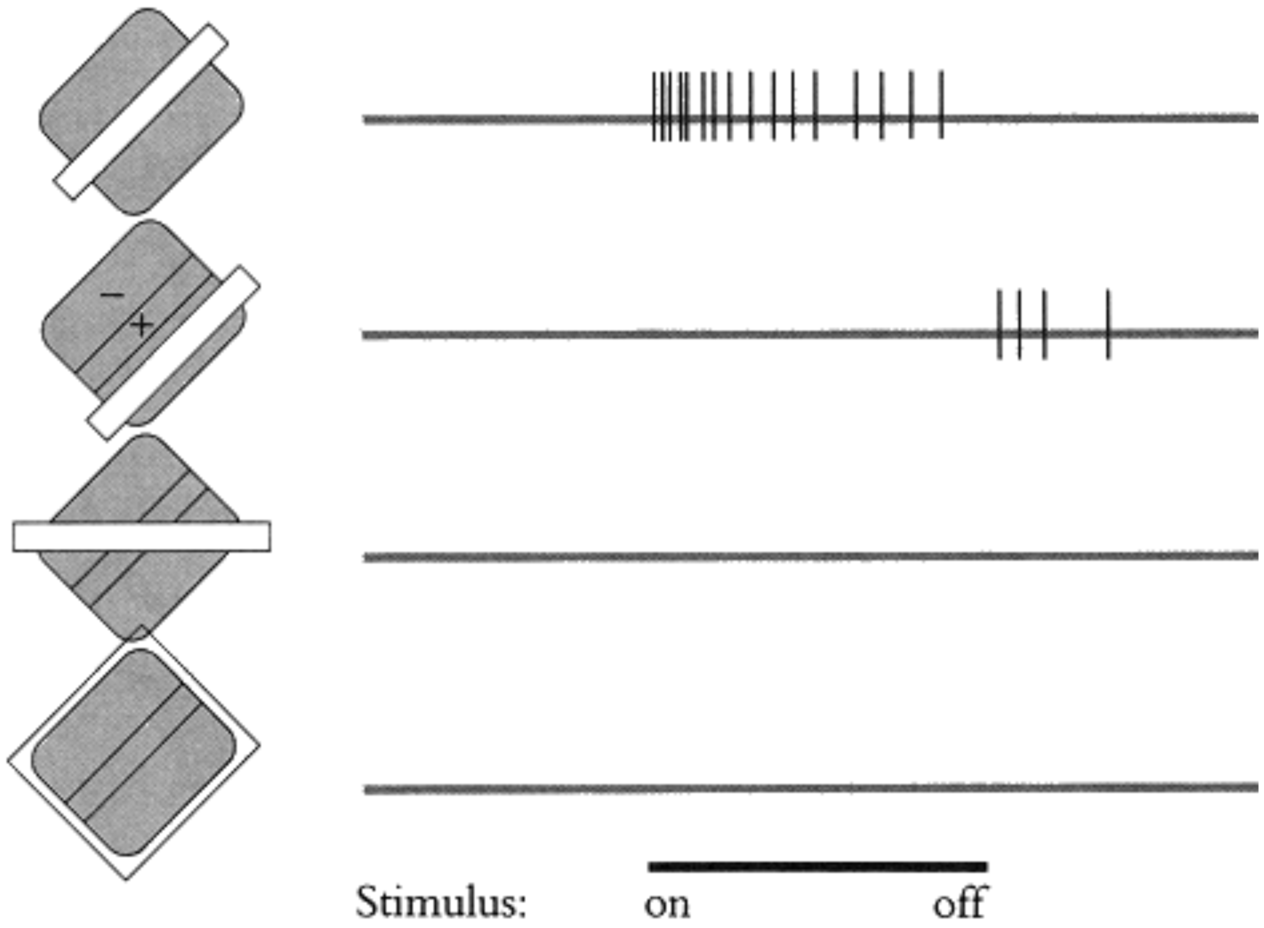
From the Neurophysiology Laboratory, Department of Pharmacology Harvard Medical School, Boston, Massachusetts, U.S.A.

(*Received* 31 *July* 1961)

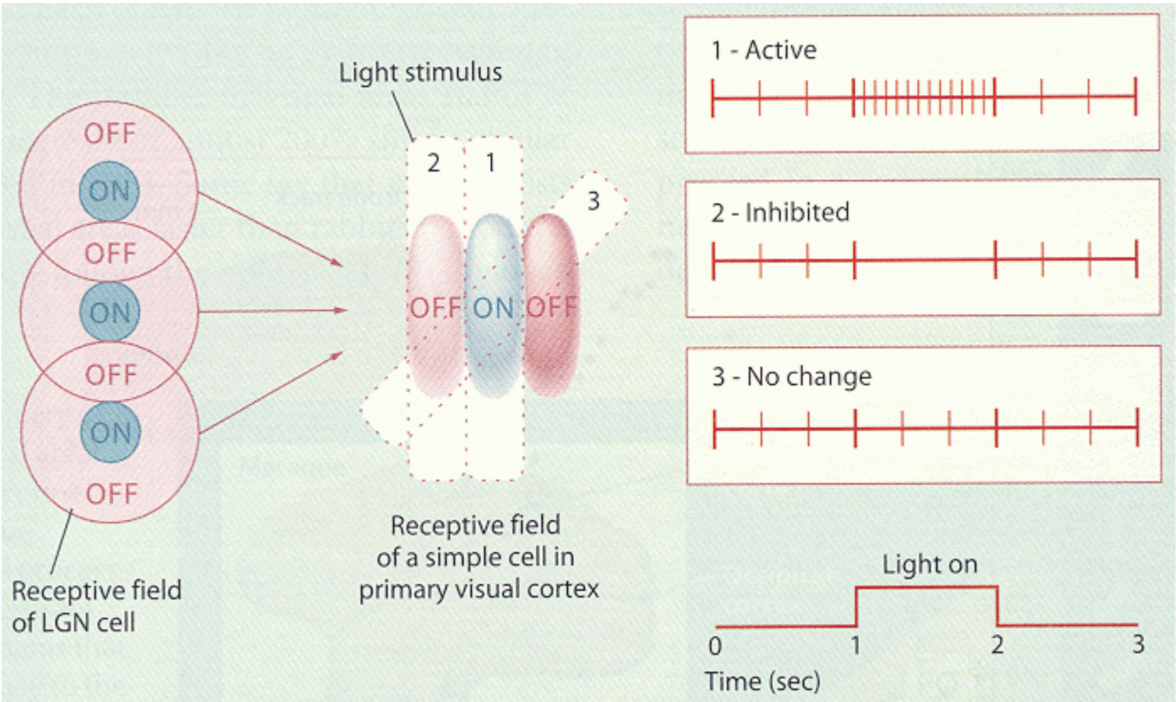
PART I

ORGANIZATION OF RECEPTIVE FIELDS IN CAT'S VISUAL CORTEX: PROPERTIES OF 'SIMPLE' AND 'COMPLEX' FIELDS

Gabors in V1

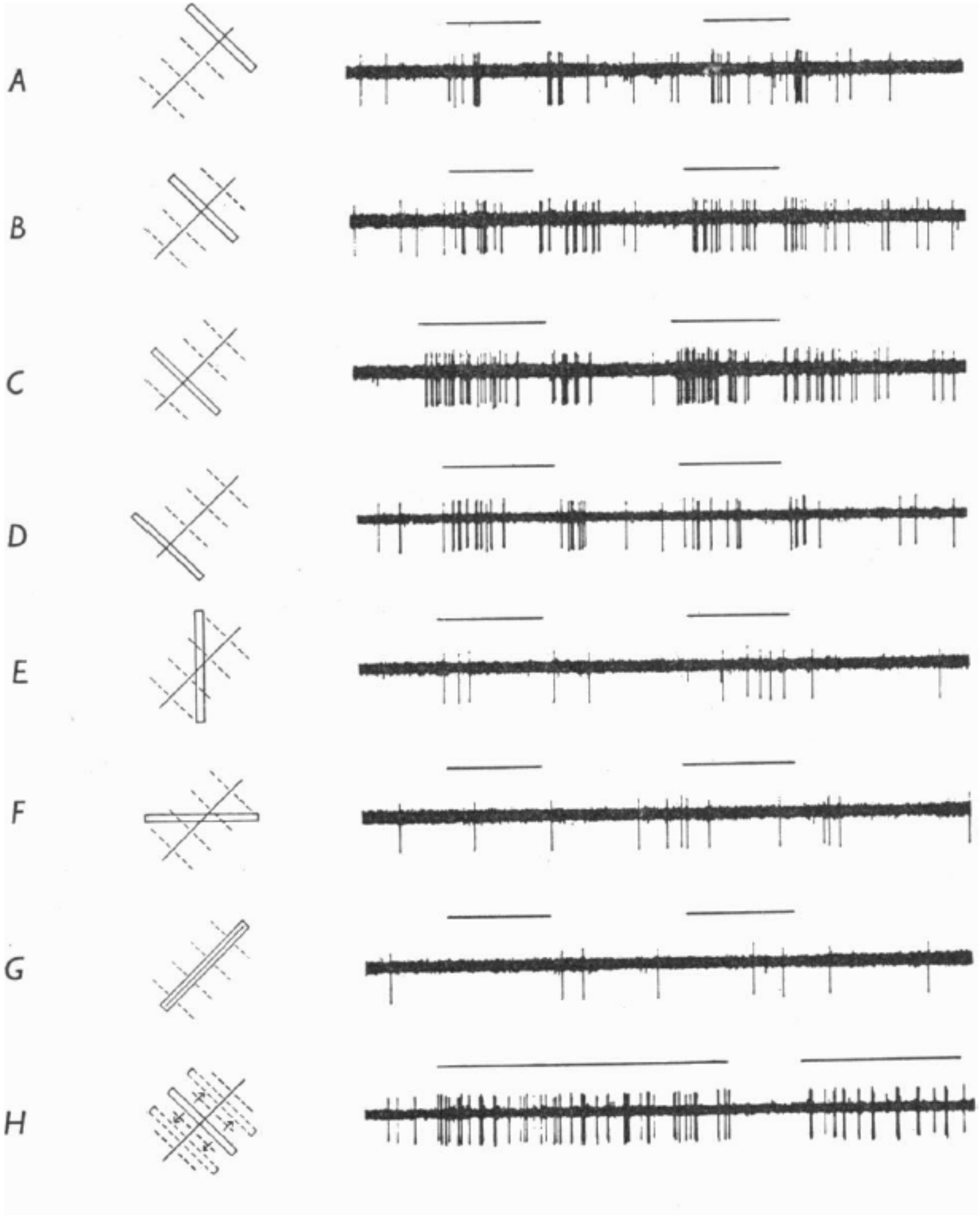


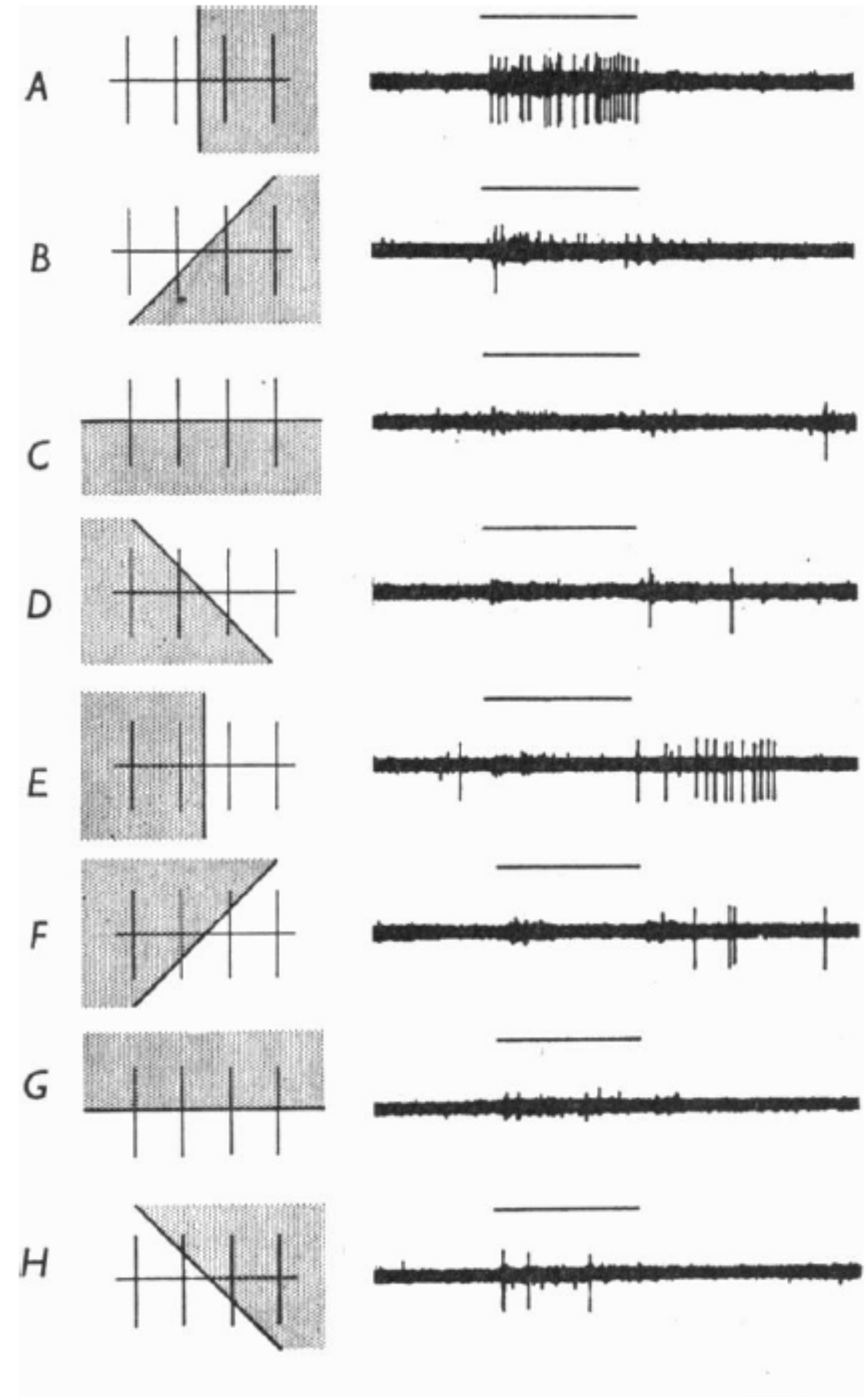
−
+
Stimulus:
on
off



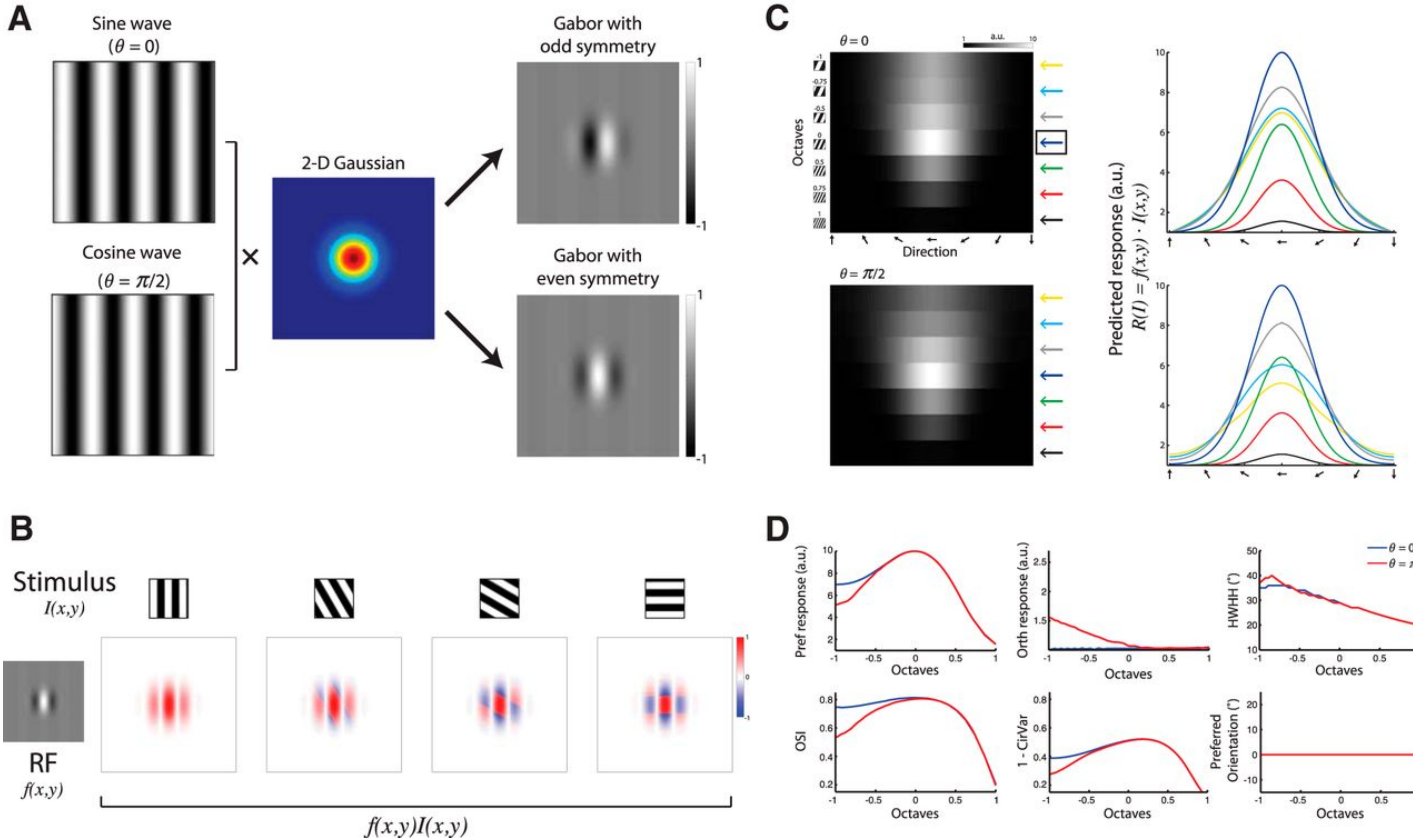
Light stimulus
OFF
ON
OFF
ON
OFF
ON
OFF
2
1
3
OFF ON OFF
Receptive field of LGN cell
Receptive field of a simple cell in primary visual cortex
1 - Active
2 - Inhibited
3 - No change
Light on
0
1
2
3
Time (sec)

Gabors in V1



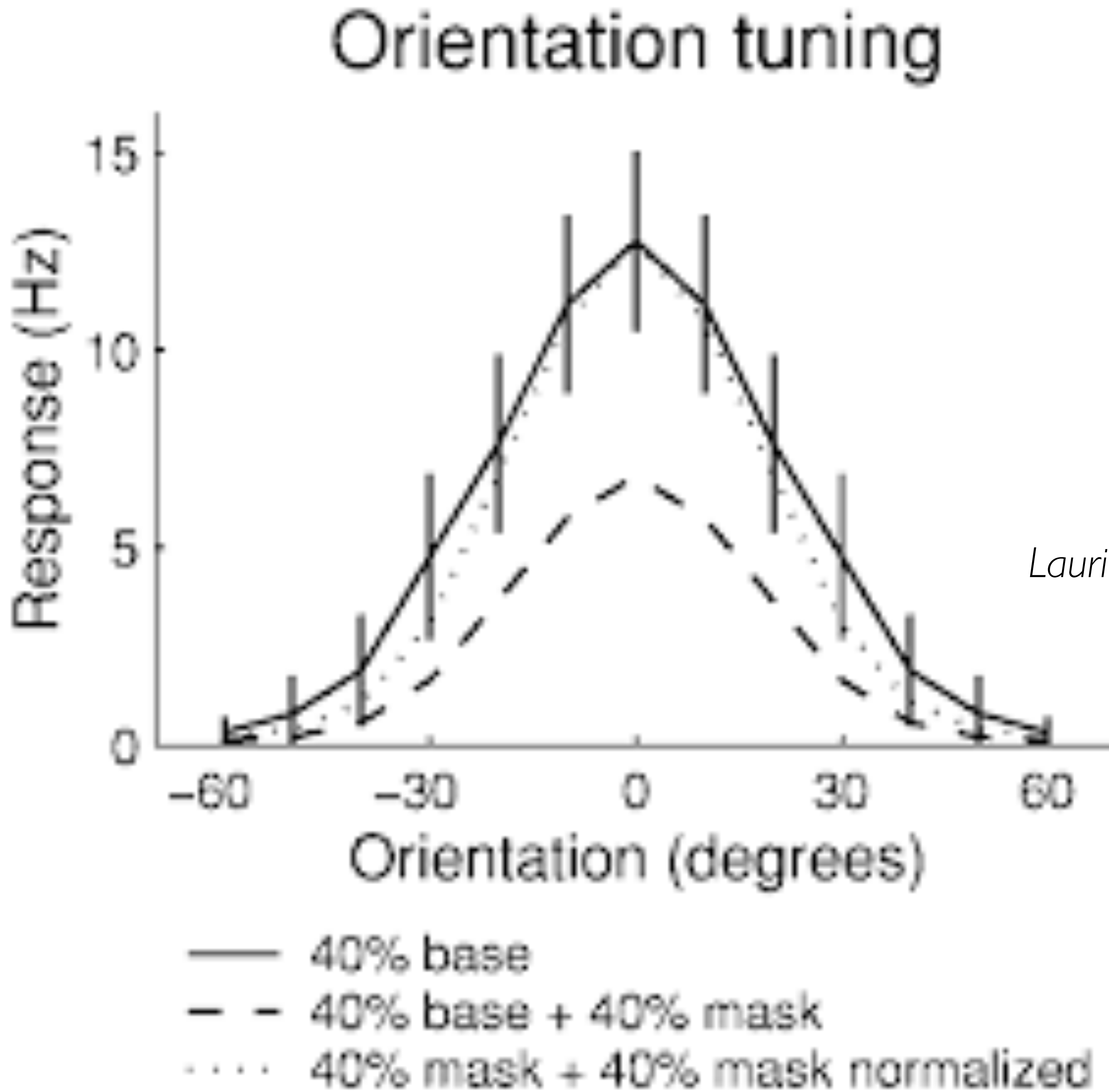


Orientation Tuning Curves



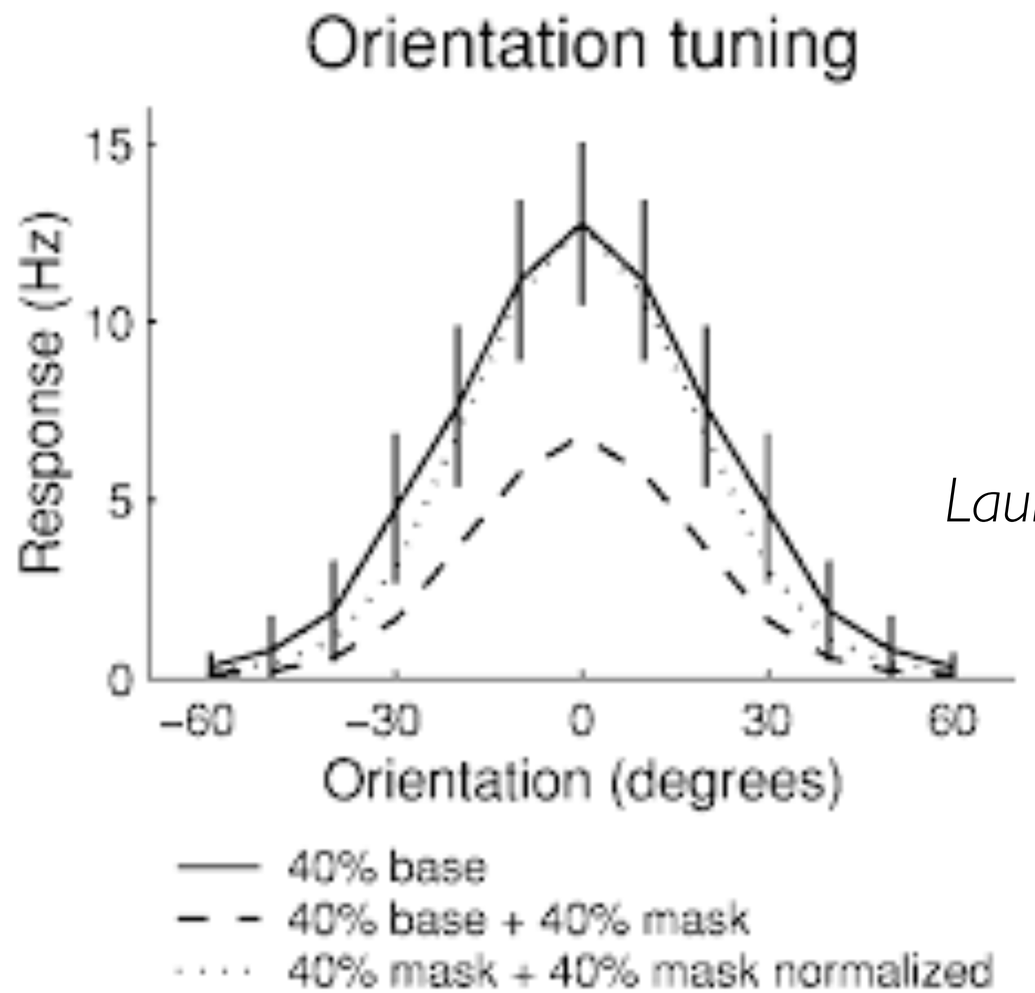
from Ayzenshtat et al (2016)

Orientation Tuning Curves



Lauritzen et al 2001

Orientation Tuning Curves

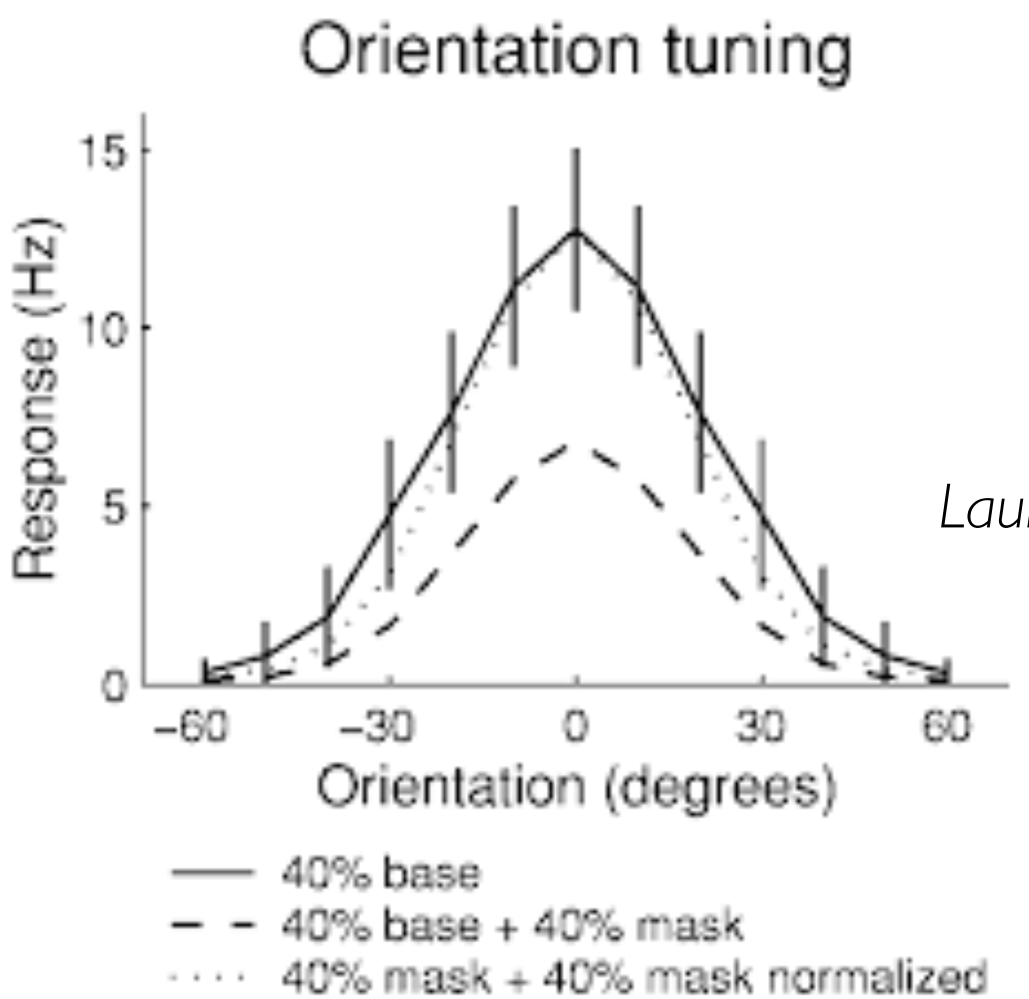


Lauritzen et al 2001

$$\textbf{Circular Variance} = 1 - \frac{\sum_k r_k \mathbf{e}^{2i\theta_k}}{\sum_k r_k}$$

r_k = neuron **r**'s response to stimulus with pure orientation ***k***

Orientation Tuning Curves



Lauritzen et al 2001

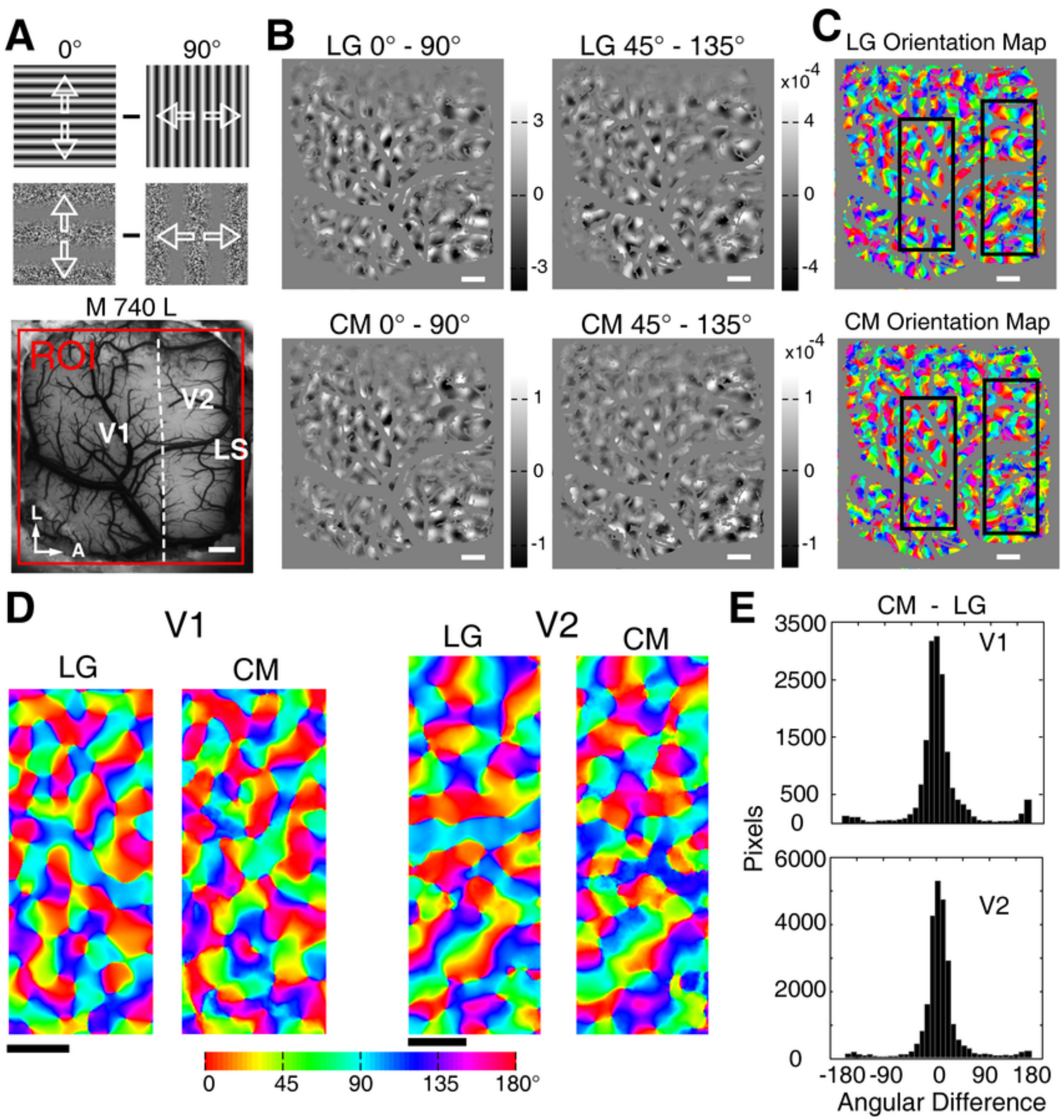
to make it "circular"

$$\textbf{Circular Variance} = 1 - \frac{\sum_k r_k \boxed{\mathbf{e}^{2i\theta_k}}}{\boxed{\sum_k r_k}}$$

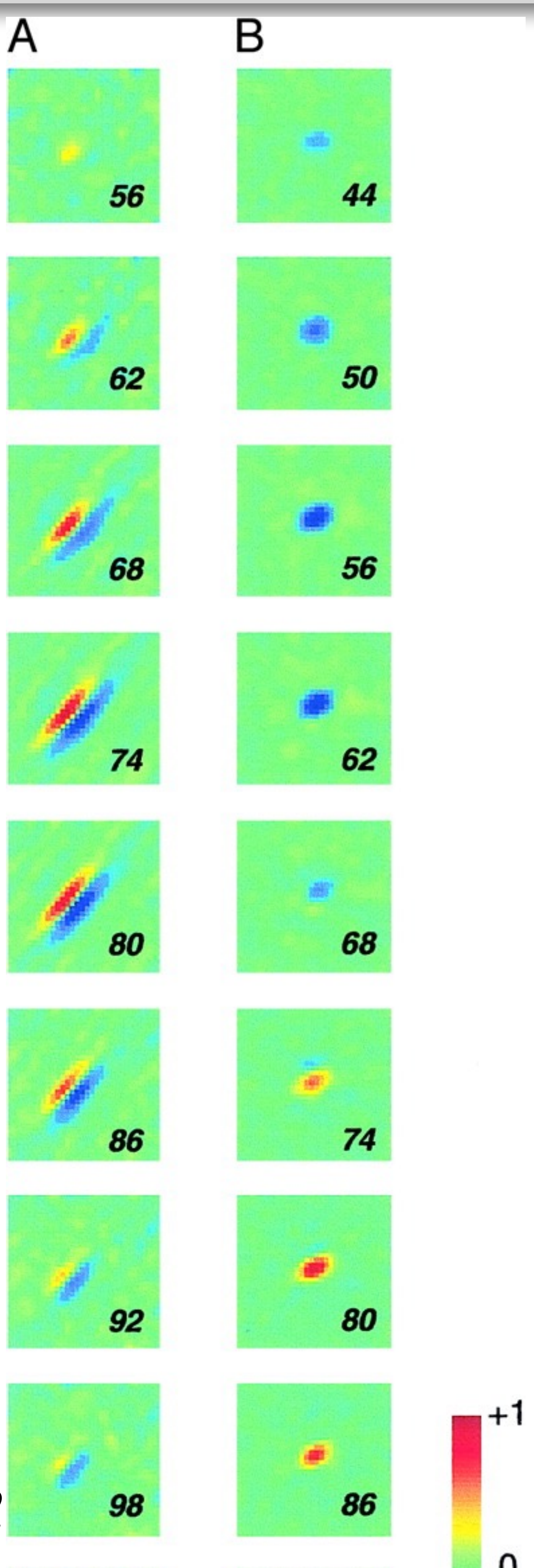
total response

r_k = neuron **r**'s response to stimulus with pure orientation ***k***

Gabors in V1



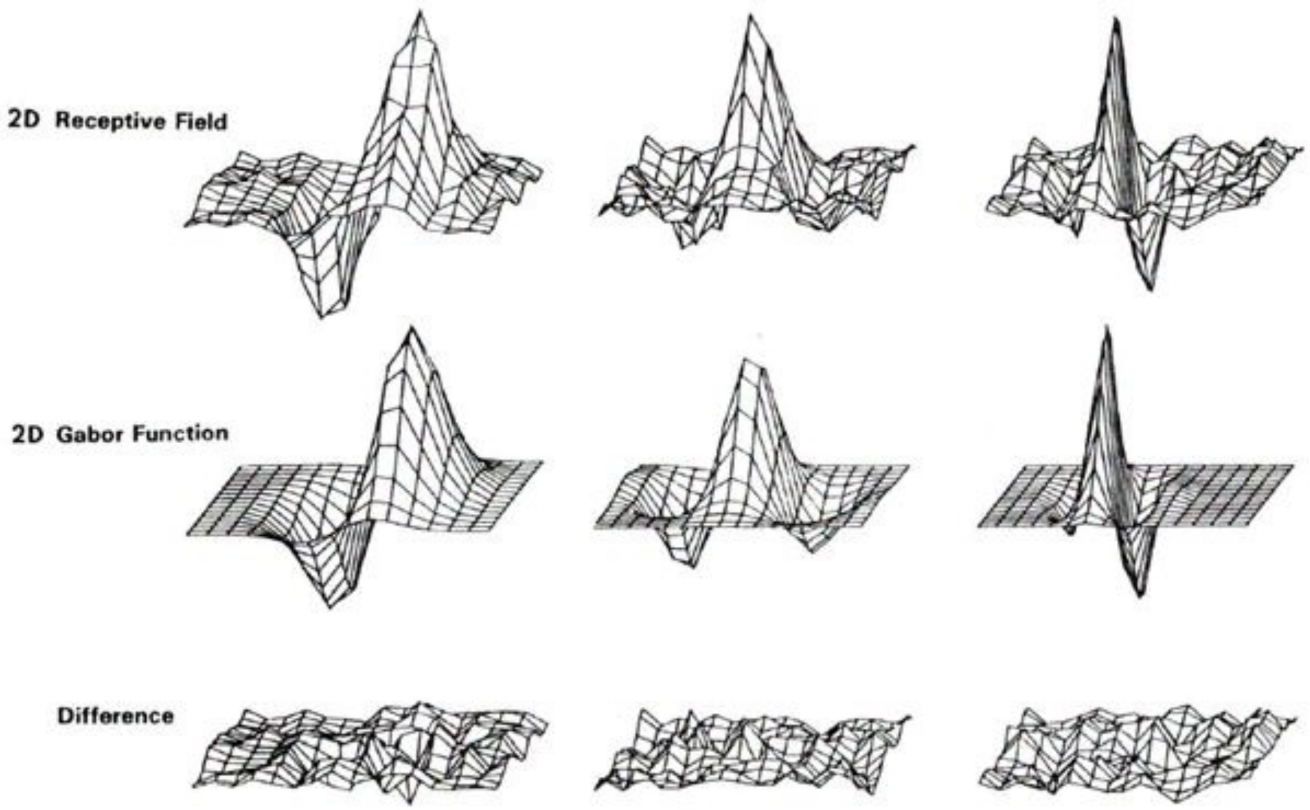
An et al 2015



Ringach 2002

Gabors in V1

Simple V1 cells Daugman, 1985



Receptive fields in primary visual cortex (Jones and Palmer, 1987)

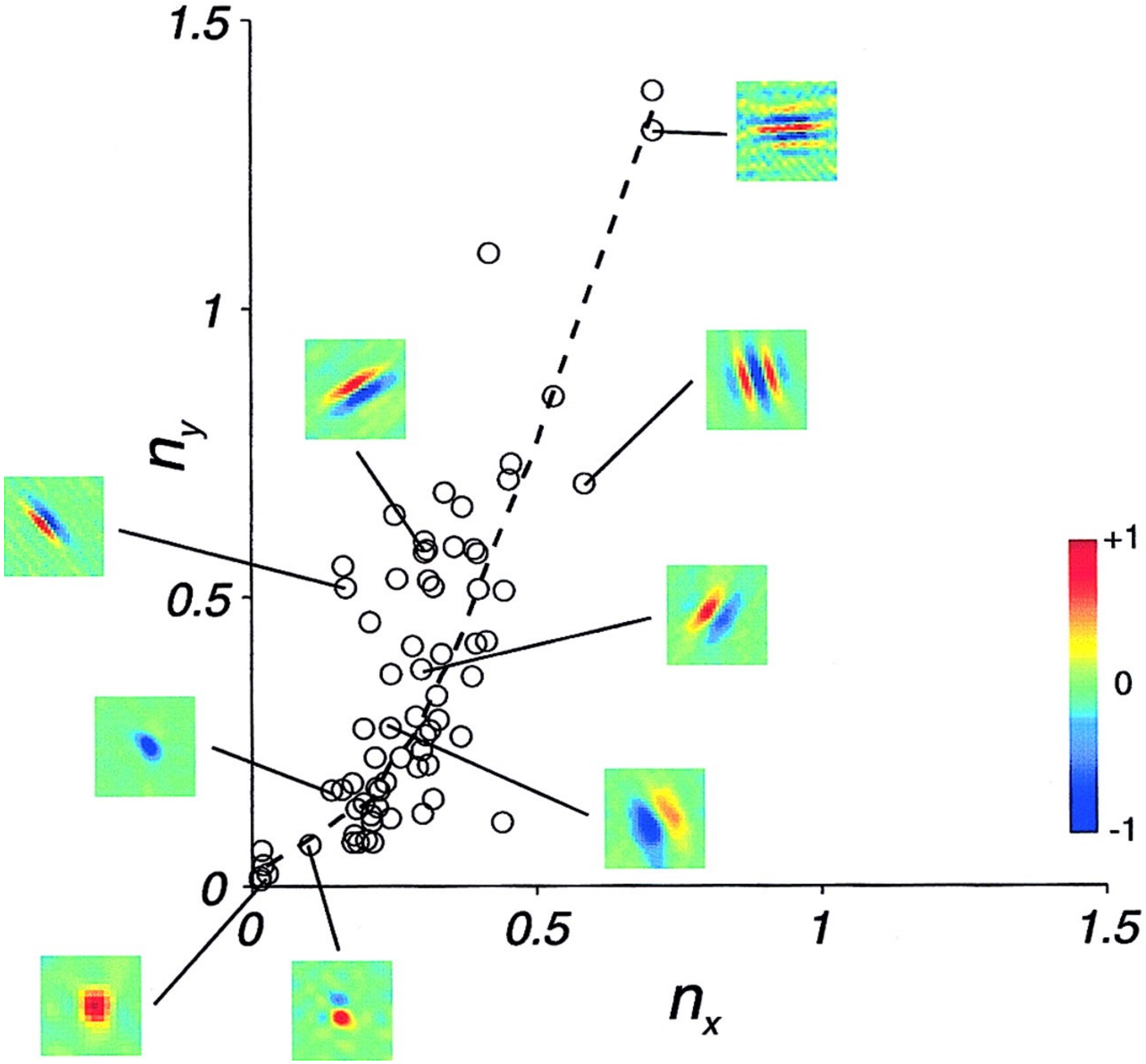
Gabor wavelets: localized sine and cosine waves

$$G(x) \propto \exp\{-\frac{1}{2}[\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2}]\}e^{ix_1}$$

Transation, rotation, dilation of the above function

Gabors in V1

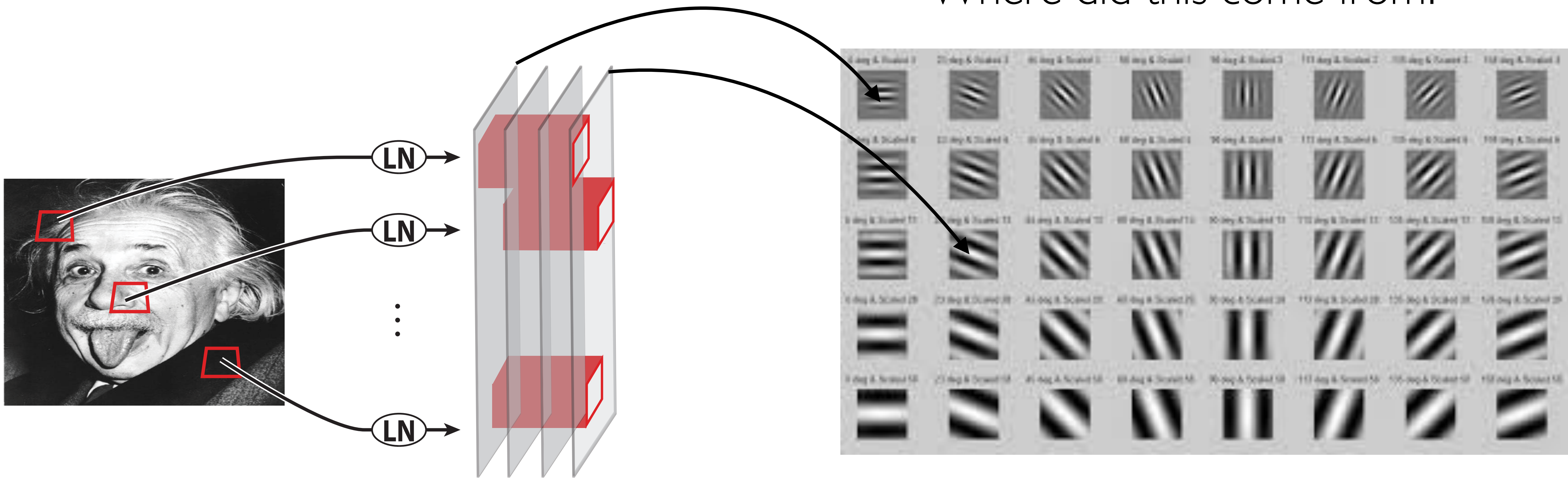
There is a frequency-orientation relationship:



from Ringach 2002

Models of V1

Where did this come from?



Two strategies to find the correct parameters.

less normative theory ⟷ *more normative theory*

1. Fit neural data

2. Solve a high-level ecological task

…

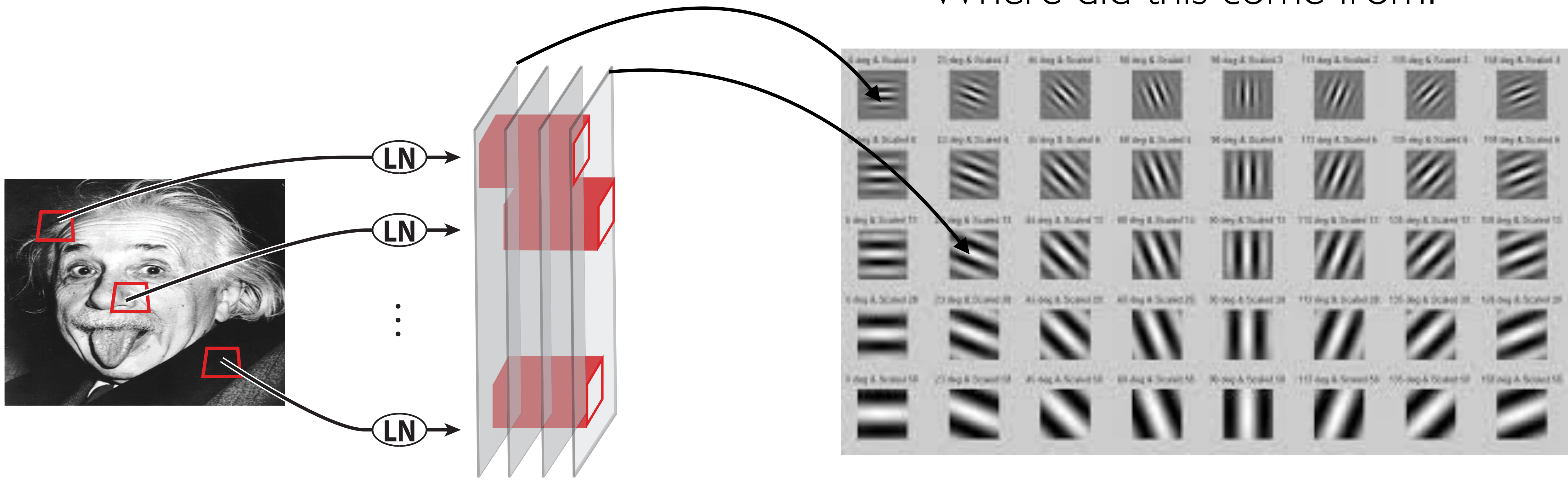
compare to neural data

and

Turing Test

Models of V1

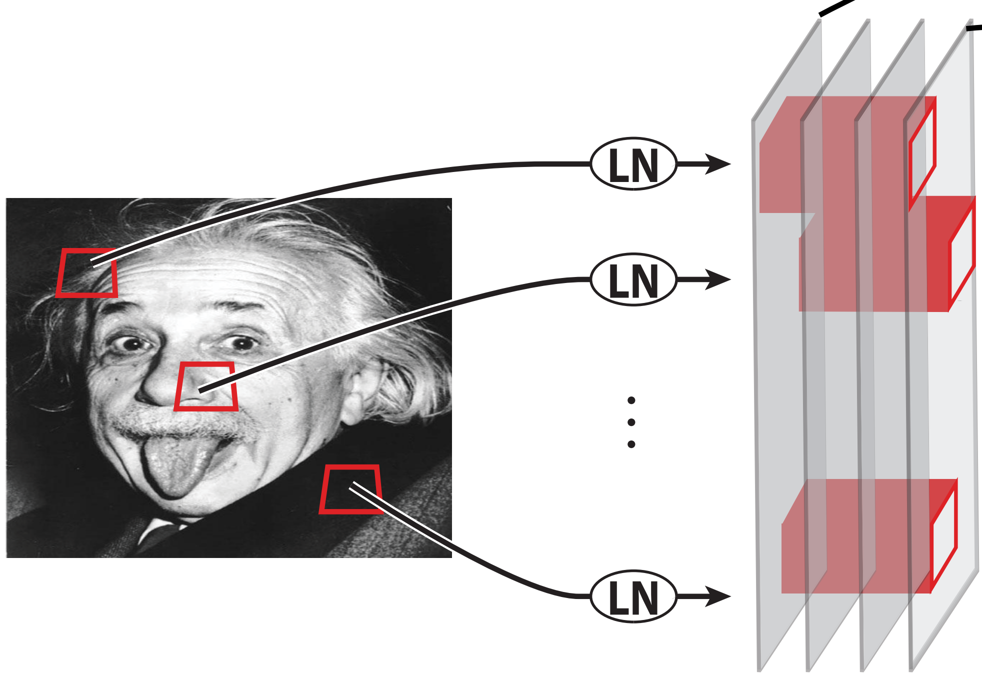
Where did this come from?



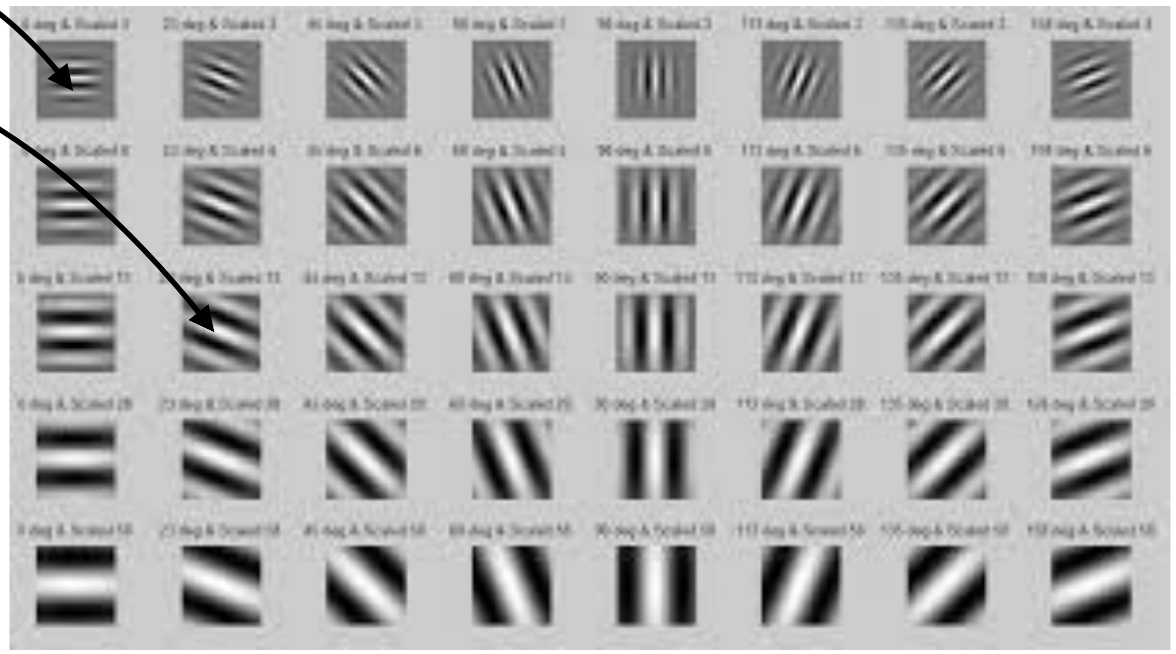
(1) "Hubel and Wiesel's Intuition"
~1970s and formalized later

→ e.g. there is a "fixed basis set" that just "makes sense" if we're smart enough

Models of V1



Where did this come from?



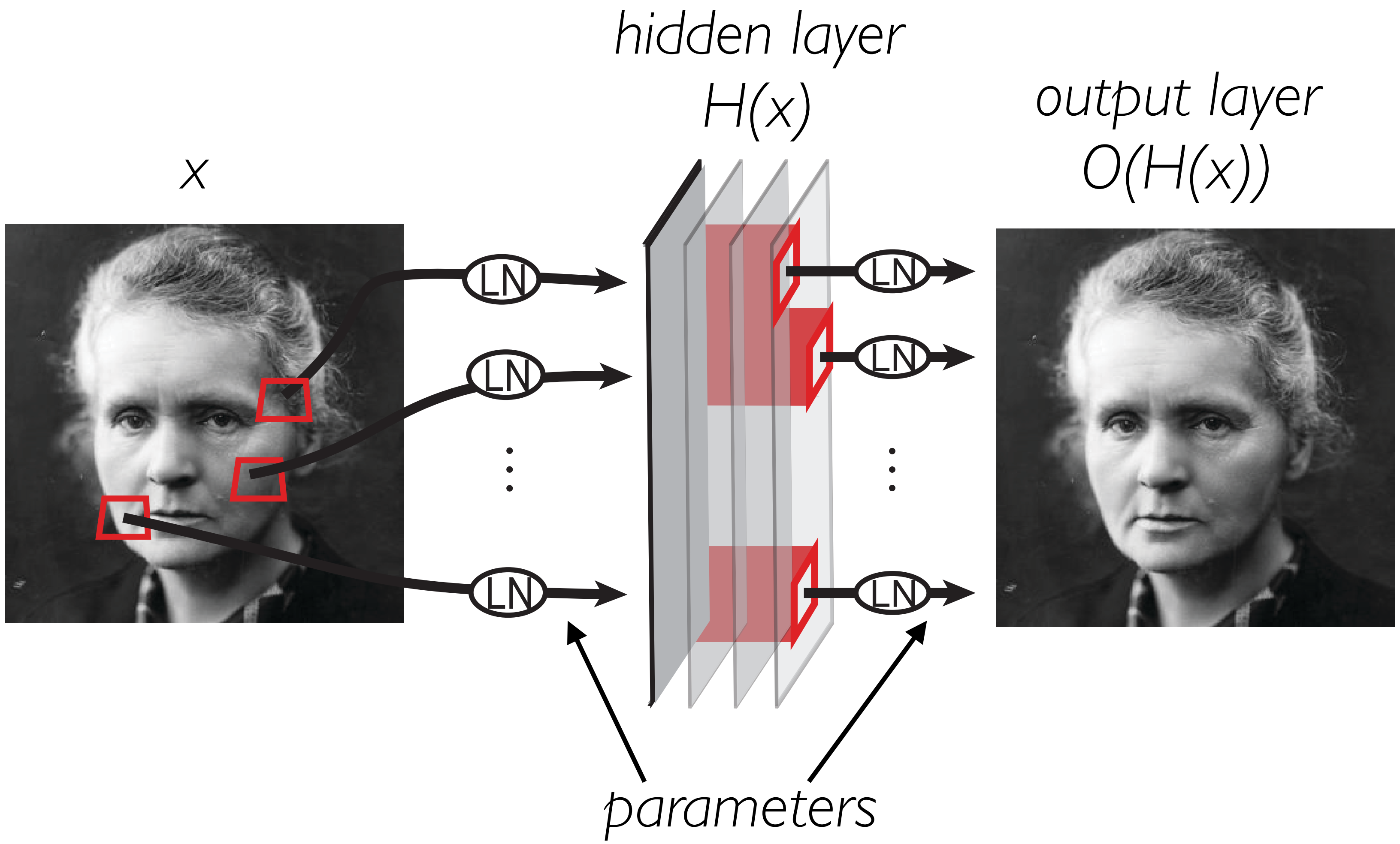
(1) "Hubel and Wiesel's Intuition"
~1970s and formalized later

→ e.g. there is a "fixed basis set" that just "makes sense" if we're smart enough

(2) Sparse Coding Foldiak, Olshausen, mid 1990s

→neurons have to represent their environment, as efficiently as possible

Models of V1

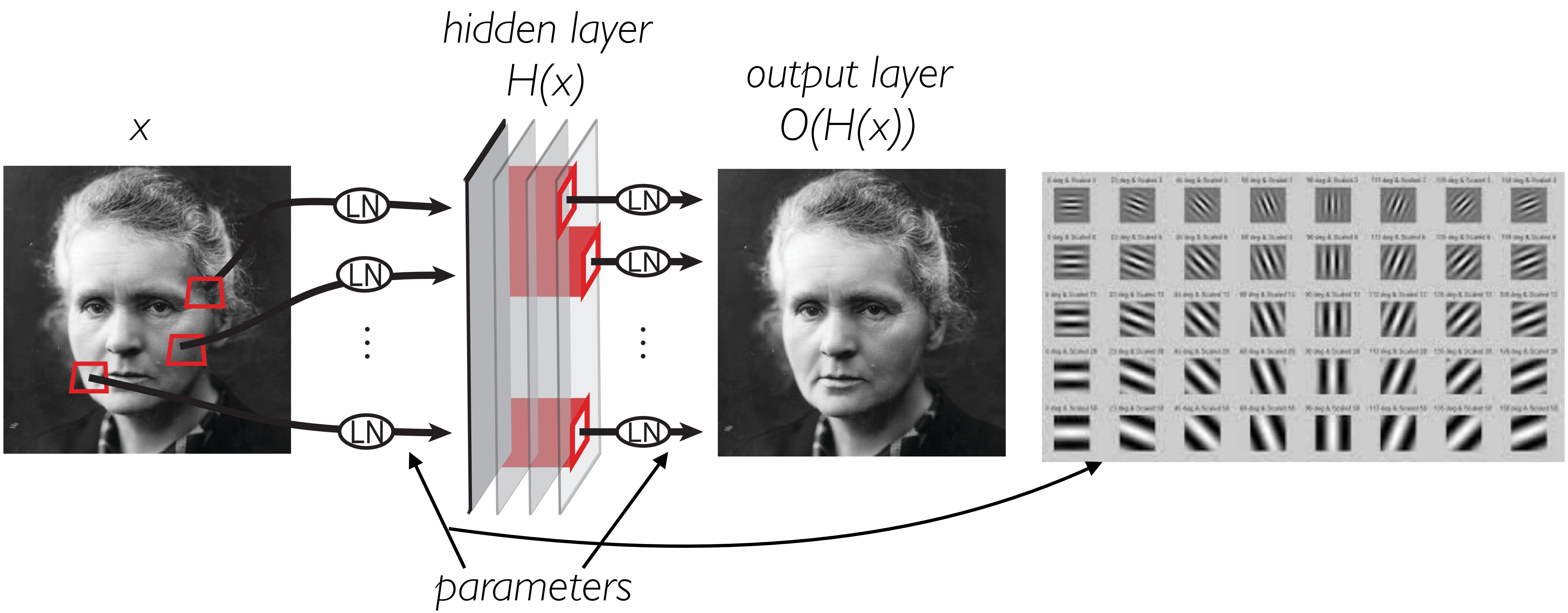


$$L(x) = |x - O(H(x))|^2 + \lambda \cdot |H(x)|$$

(2) Sparse Coding Foldiak, Olshausen, mid 1990s

→neurons have to represent their environment, as efficiently as possible

Models of V1

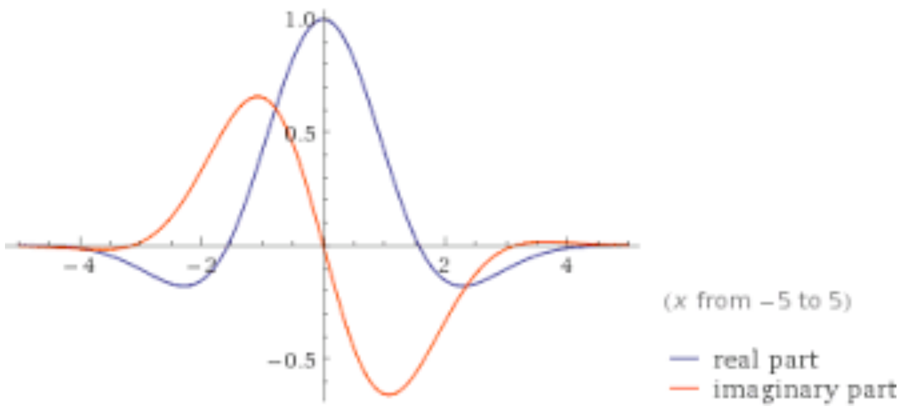


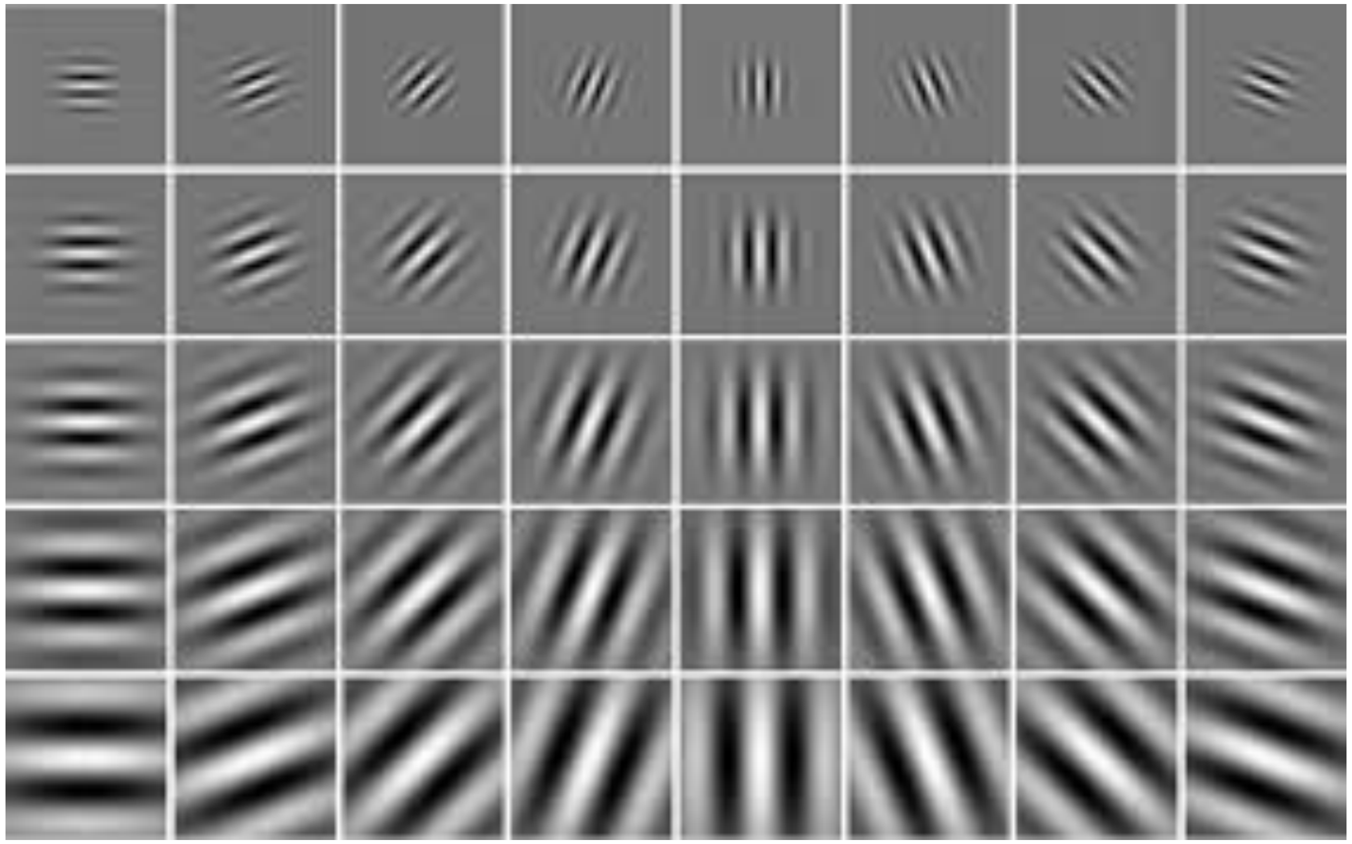
$$L(x) = |x - O(H(x))|^2 + \lambda \cdot |H(x)|$$

(2) Sparse Coding Foldiak, Olshausen, mid 1990s

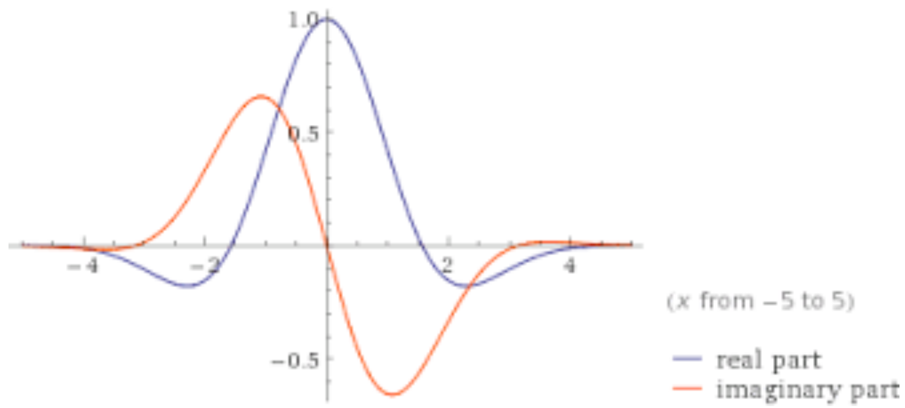
→neurons have to represent their environment, as efficiently as possible

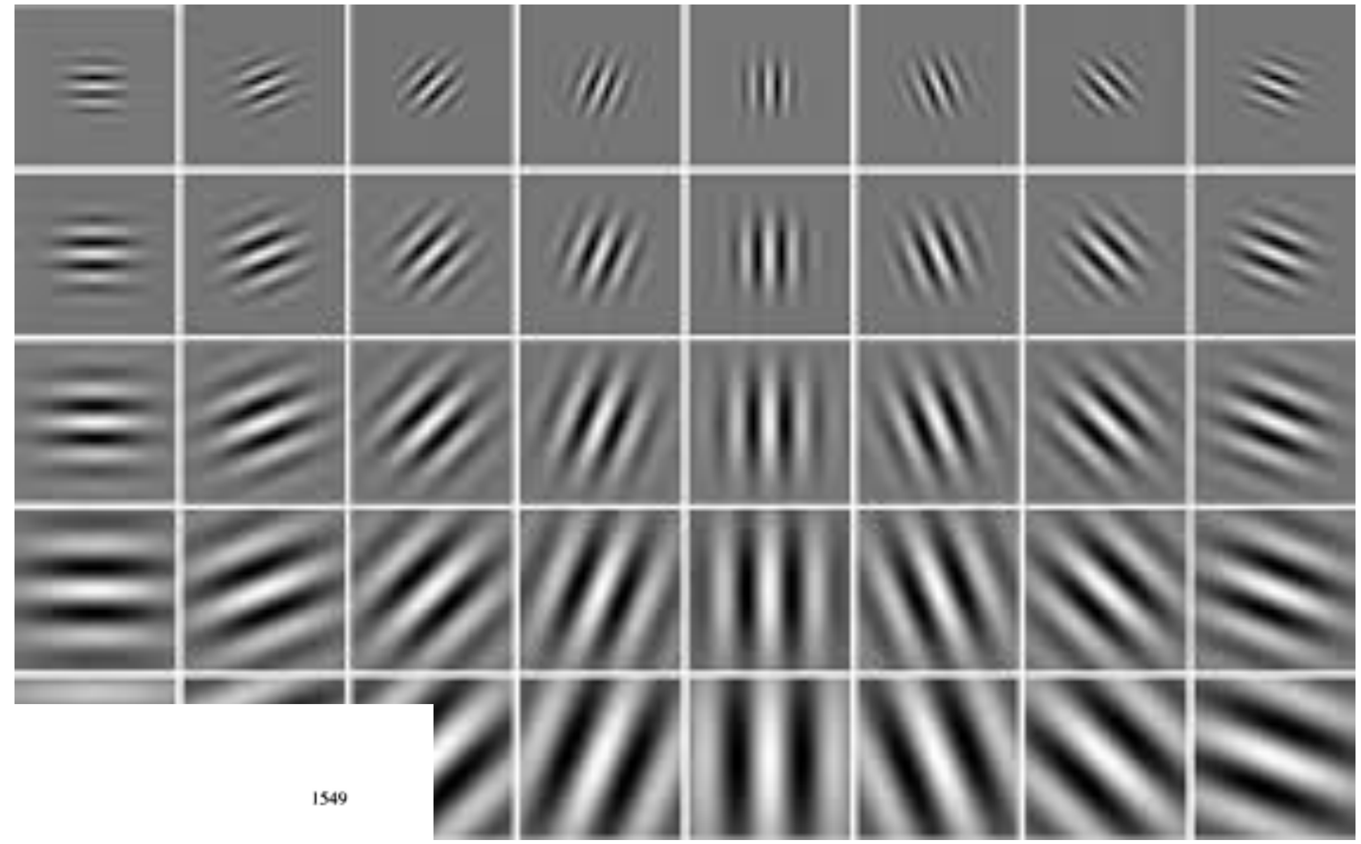
Old-School CV part 2: The Wavelet Wave





Old-School CV part 2: The Wavelet Wave





IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 4, NO. 11, NOVEMBER 1995 1549

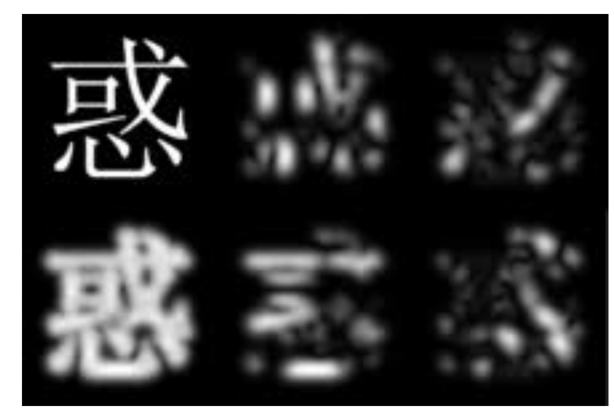
Texture Classification and Segmentation Using Wavelet Frames

Michael Unser, *Senior Member, IEEE*

Abstract—**This paper describes a new approach to the characterization of texture properties at multiple scales using the wavelet transform. The analysis uses an overcomplete wavelet decomposition, which yields a description that is translation invariant. It is shown that this representation constitutes a tight frame of l_2 and that it has a fast iterative algorithm. A texture is characterized by a set of channel variances estimated at the output of the corresponding filter bank. Classification experiments with 12 Brodatz textures indicate that the discrete wavelet frame (DWF) approach is superior to a standard (critically sampled) wavelet transform feature extraction. These results also suggest that this approach should perform better than most traditional single resolution techniques (co-occurrences, local linear transform, and the like). A detailed comparison of the classification performance of various orthogonal and biorthogonal wavelet transforms is also provided. Finally, the DWF feature extraction technique is incorporated into a simple multicomponent texture segmentation algorithm, and some illustrative examples are presented.**

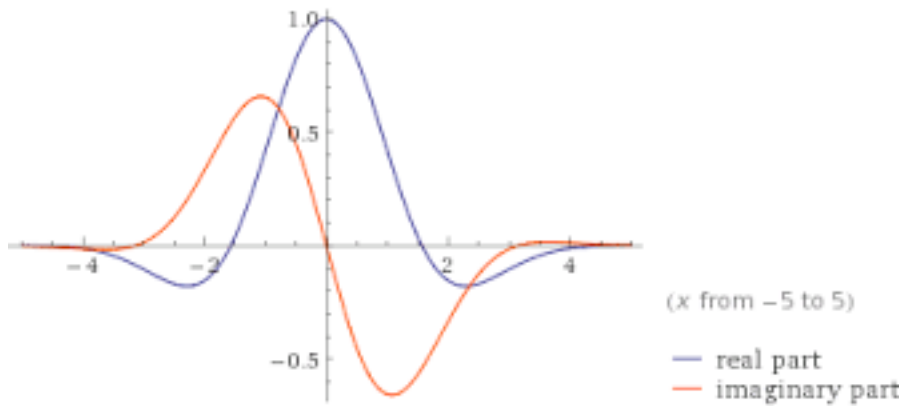
reversible, which limits their applicability for texture synthesis. Most of these problems can be avoided if one uses the wavelet transform, which provides a precise and unifying framework for the analysis and characterization of a signal at different scales [16]–[19]. The use of a pyramid-structured wavelet transform for texture analysis was first suggested in the pioneering work of Mallat [19]. This initial proposal has been followed by several studies on texture classification with a particular attention to the use of wavelet packets [20], [21], which constitute a multiband extension of the pyramid-structured wavelet transform.

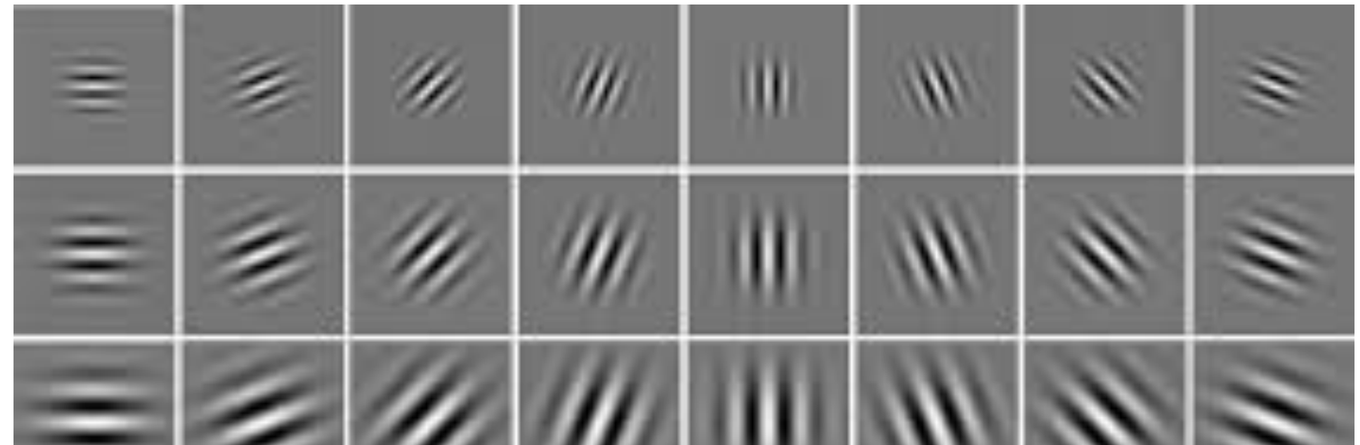
In this paper, a variation of the discrete wavelet transform is introduced for characterizing texture properties. This technique is applied to the problems of texture classification and segmentation. The present analysis method, which is described in Section II, uses an overcomplete wavelet decomposition (the discrete wavelet frame (DWF)) in which the

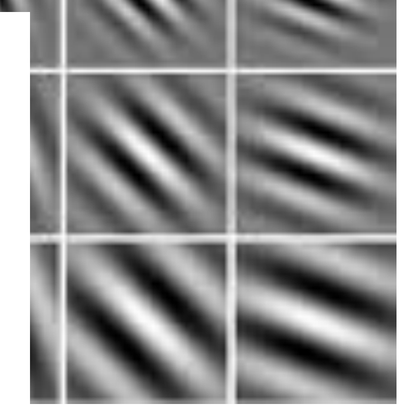


Many CV careers made on wavelets.

Old-School CV part 2: The Wavelet Wave







SIFT (Lowe, 1999)

Object Recognition from Local Scale-Invariant Features

David G. Lowe
Computer Science Department
University of British Columbia
Vancouver, B.C., V6T 1Z4, Canada
lowe@cs.ubc.ca

IEEE TRANSACTIONS ON IMAGE PROCE

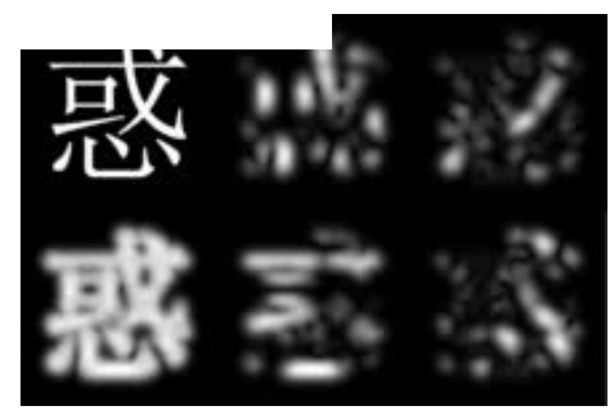
Segme

Michael Unser, *Senior Member, IEEE*

Abstract—This paper describes a new approach to the characterization of texture properties at multiple scales using the wavelet transform. The analysis uses an overcomplete wavelet decomposition, which yields a description that is translation invariant. It is shown that this representation constitutes a tight frame of l_2 and that it has a fast iterative algorithm. A texture is characterized by a set of channel variances estimated at the output of the corresponding filter bank. Classification experiments with 12 Brodatz textures indicate that the discrete wavelet frame (DWF) approach is superior to a standard (critically sampled) wavelet transform feature extraction. These results also suggest that this approach should perform better than most traditional single resolution techniques (co-occurrences, local linear transform, and the like). A detailed comparison of the classification performance of various orthogonal and biorthogonal wavelet transforms is also provided. Finally, the DWF feature extraction technique is incorporated into a simple multicomponent texture segmentation algorithm, and some illustrative examples are presented.

reversible, which limits their applicability for texture synthesis. Most of these problems can be avoided if one uses the wavelet transform, which provides a precise and unifying framework for the analysis and characterization of a signal at different scales [16]–[19]. The use of a pyramid-structured wavelet transform for texture analysis was first suggested in the pioneering work of Mallat [19]. This initial proposal has been followed by several studies on texture classification with a particular attention to the use of wavelet packets [20], [21], which constitute a multiband extension of the pyramid-structured wavelet transform.

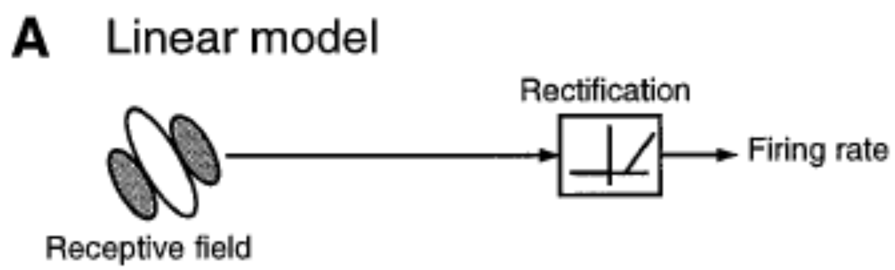
In this paper, a variation of the discrete wavelet transform is introduced for characterizing texture properties. This technique is applied to the problems of texture classification and segmentation. The present analysis method, which is described in Section II, uses an overcomplete wavelet decomposition

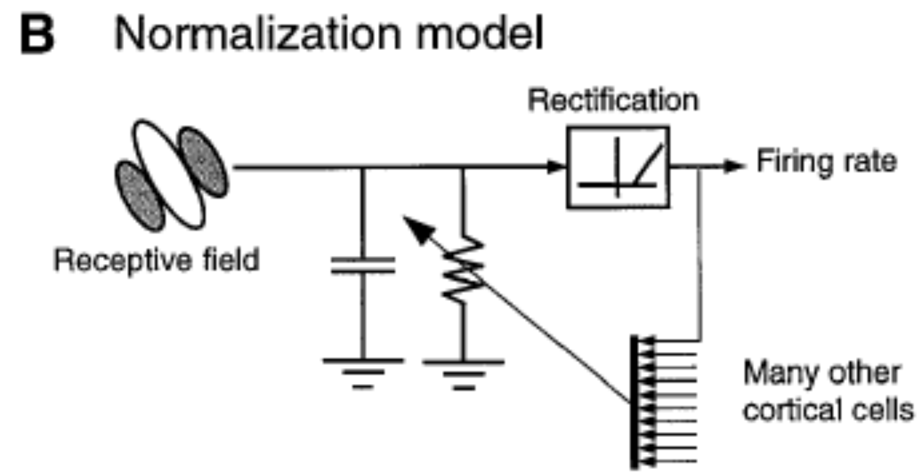


Many CV careers made on wavelets.

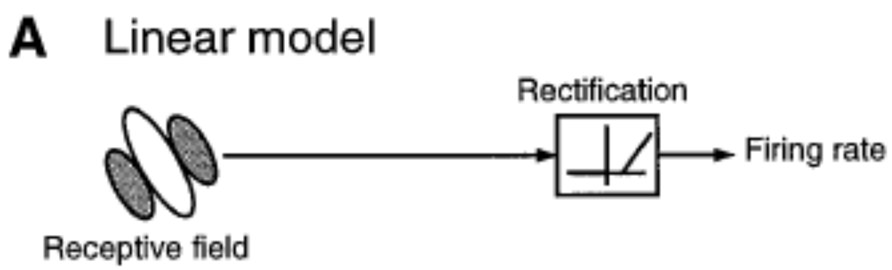


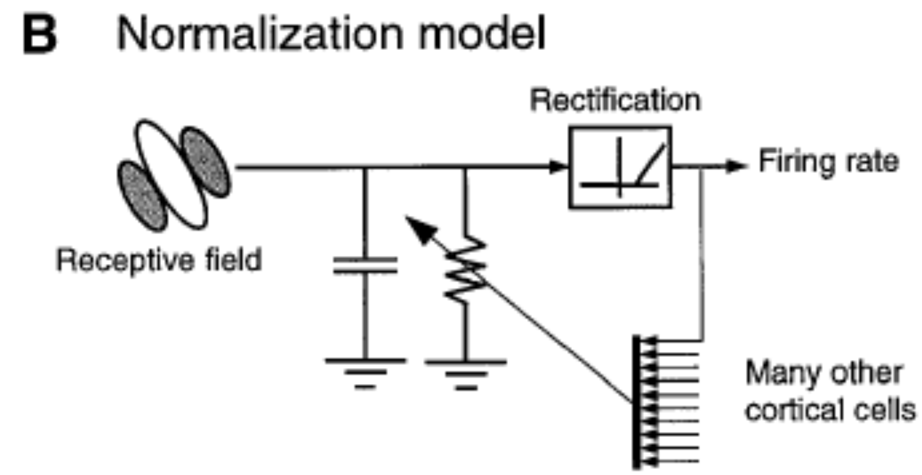
Carandini, Heeger and Movshon (1997)





Carandini, Heeger and Movshon (1997)





$$C\frac{dV}{dt} + gV = I$$

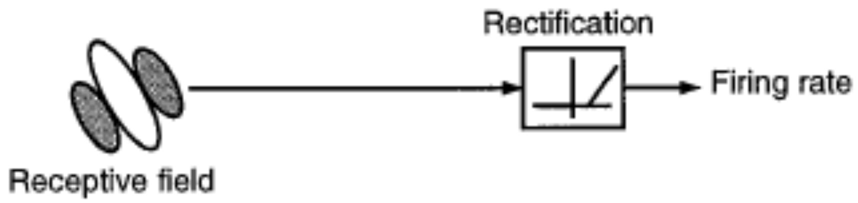
$$g = \frac{g_0}{\sqrt{1 - k \cdot \sum_{r \in R_x} r}}$$

$$R = max(0, V)$$

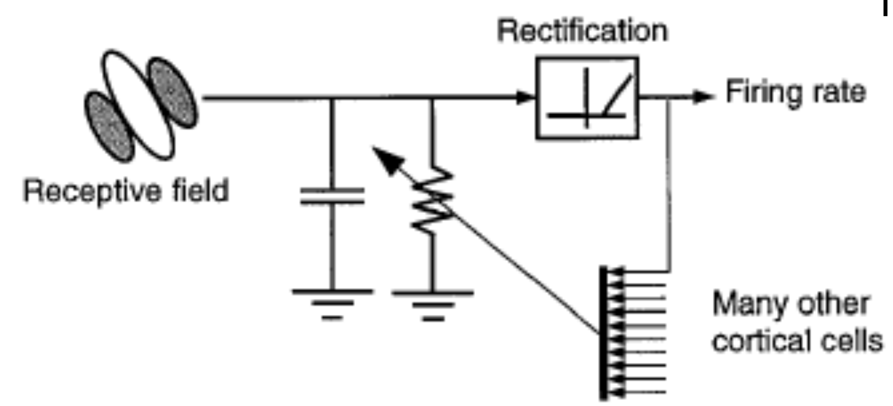


Carandini, Heeger and Movshon (1997)

A Linear model



B Normalization model



R

$$C\frac{dV}{dt} + gV = I$$

$$g = \frac{g_0}{\sqrt{1 - k \cdot \sum_{r \in R_x} r}}$$

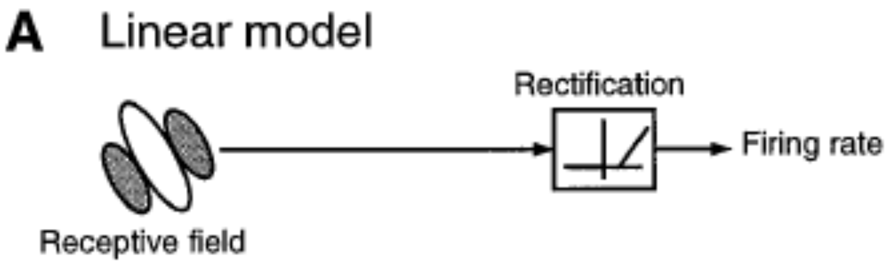
$$R = max(0, V)$$

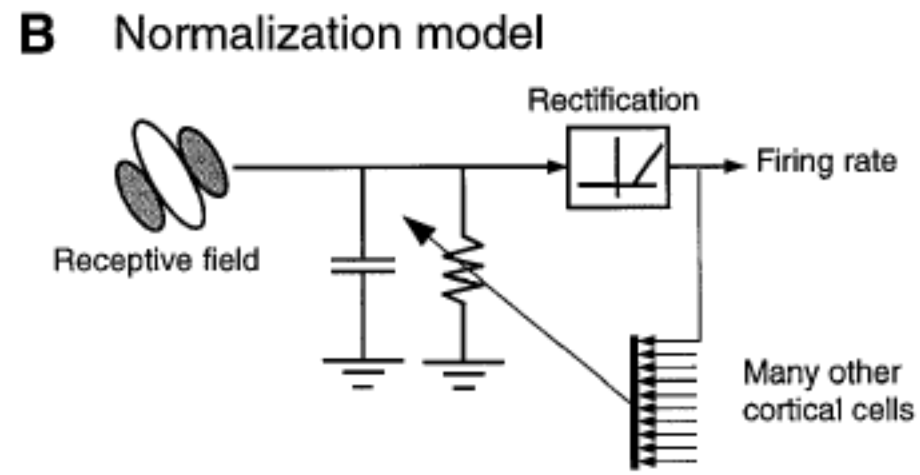
measure R from neural data

solve diff eq for equilibrium, estimate free parameters: C, k, g0



Carandini, Heeger and Movshon (1997)





y = R [W * x]

y = R [norm(W * x)]

OR: derive this expression: ——>
(basically)

$$norm(x) \sim \frac{x}{\left(\gamma + \alpha \cdot \sum_{r \in R_x} x_r^2\right)^{\beta}}$$

Carandini, Heeger and Movshon (1997)

y = R [W * x]

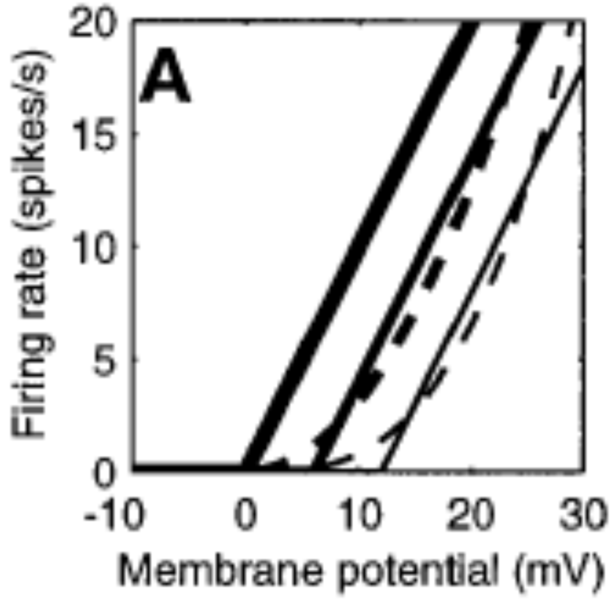
y = R [norm(W * x)]

OR: derive this expression: ——> (basically)

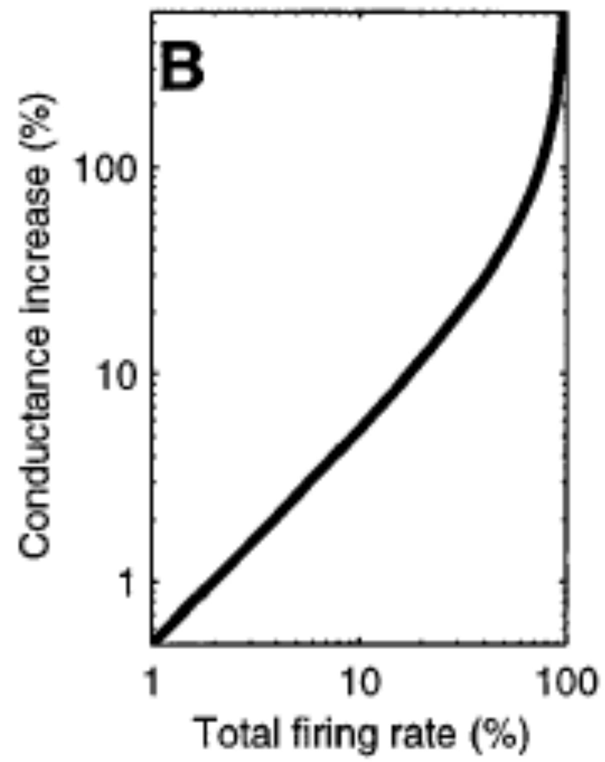
$$norm(x) \sim \frac{x}{\left(\gamma + \alpha \cdot \sum_{r \in R_x} x_r^2\right)^{\beta}}$$

NB: (1) derivation involves "reasonable" assumption that "the normalization pool to contain quadruples of cells with the same amplitude response but with phases 90° apart." (2) <u>**The above is how we now *define* local response normalization**</u>

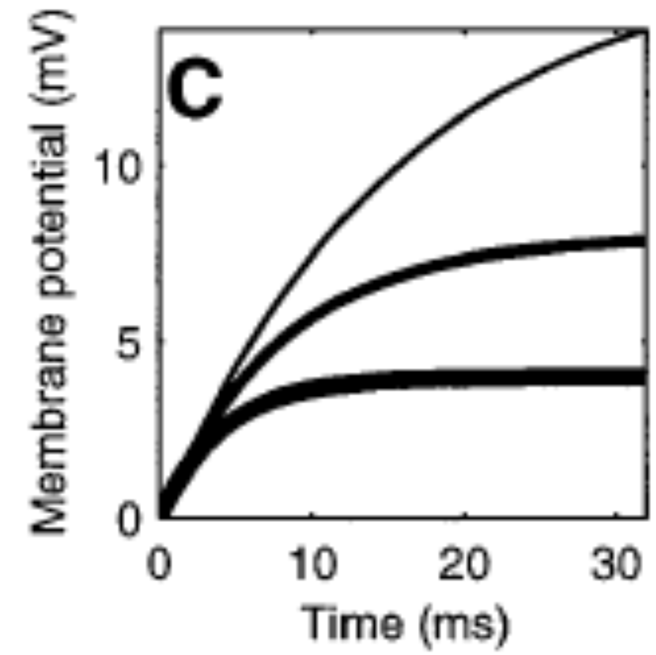
Interrelations and effects of the principal variables in the normalization model.



Relation between membrane potential and firing rate.



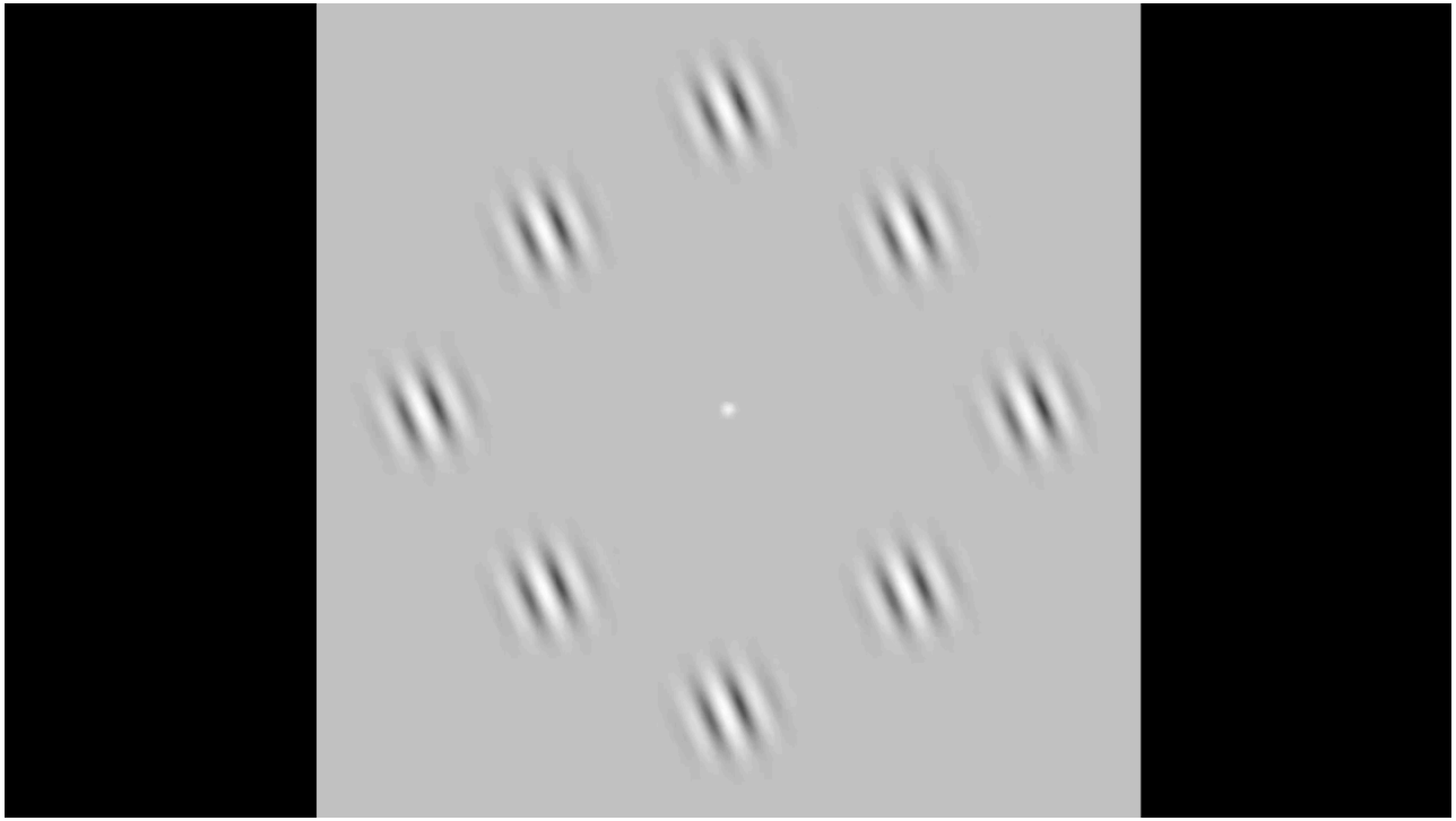
Relation between pool activity and membrane conductance.



Effects of conductance on the size and time course of the membrane potential.

Responses to drifting sine gratings of different contrasts

Carandini, Heeger and Movshon (1997)



The ***I*** in this equation:

$$C\frac{dV}{dt} + gV = I$$

is a sinusoid

neural response ~ A * sin (w t + d)

A = amplitude of cell

w = frequency of the drift

d = phase of cell

A, d are fit to the data

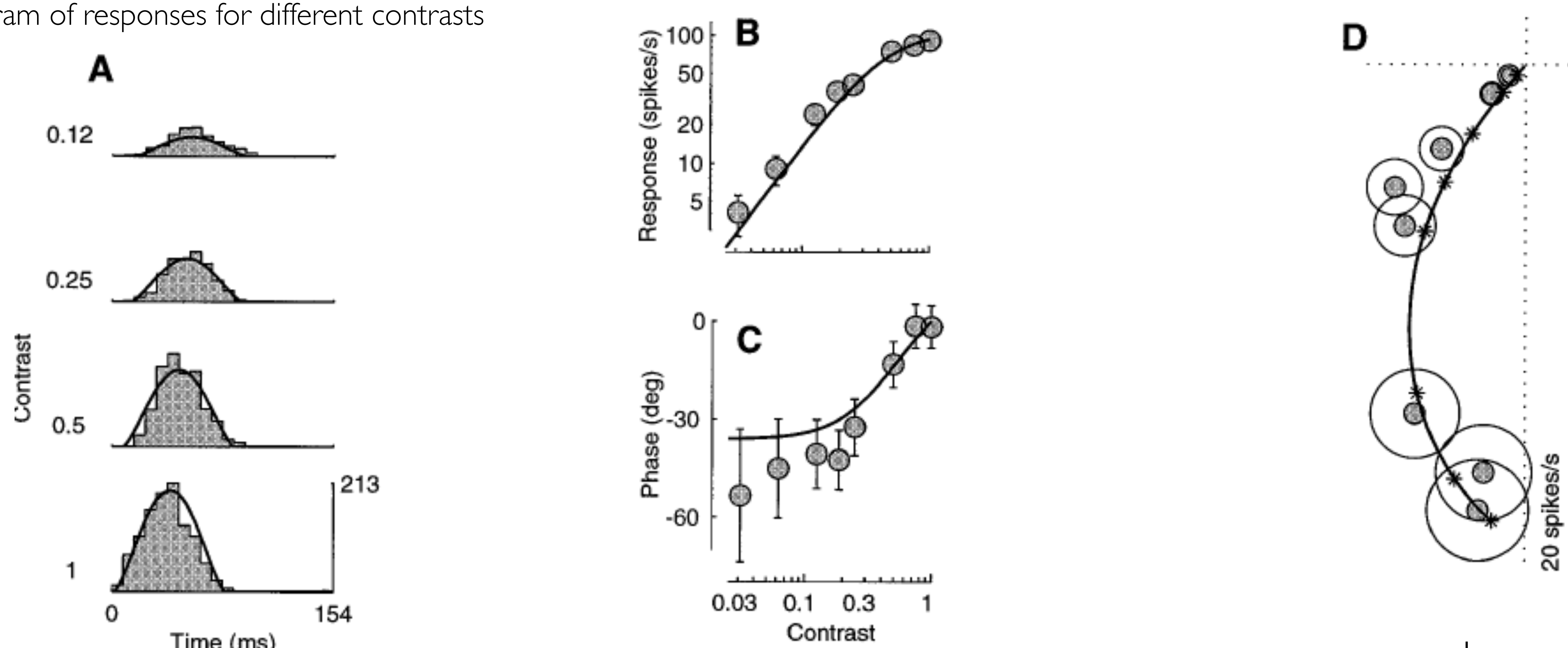
Responses to drifting sine gratings of different contrasts

Carandini, Heeger and Movshon (1997)

What functions are A and d of a stimulus parameter — contrast?

the parameters **C, k, A, d** (basically) are fit to the neuron over a bunch of stimulus conditions

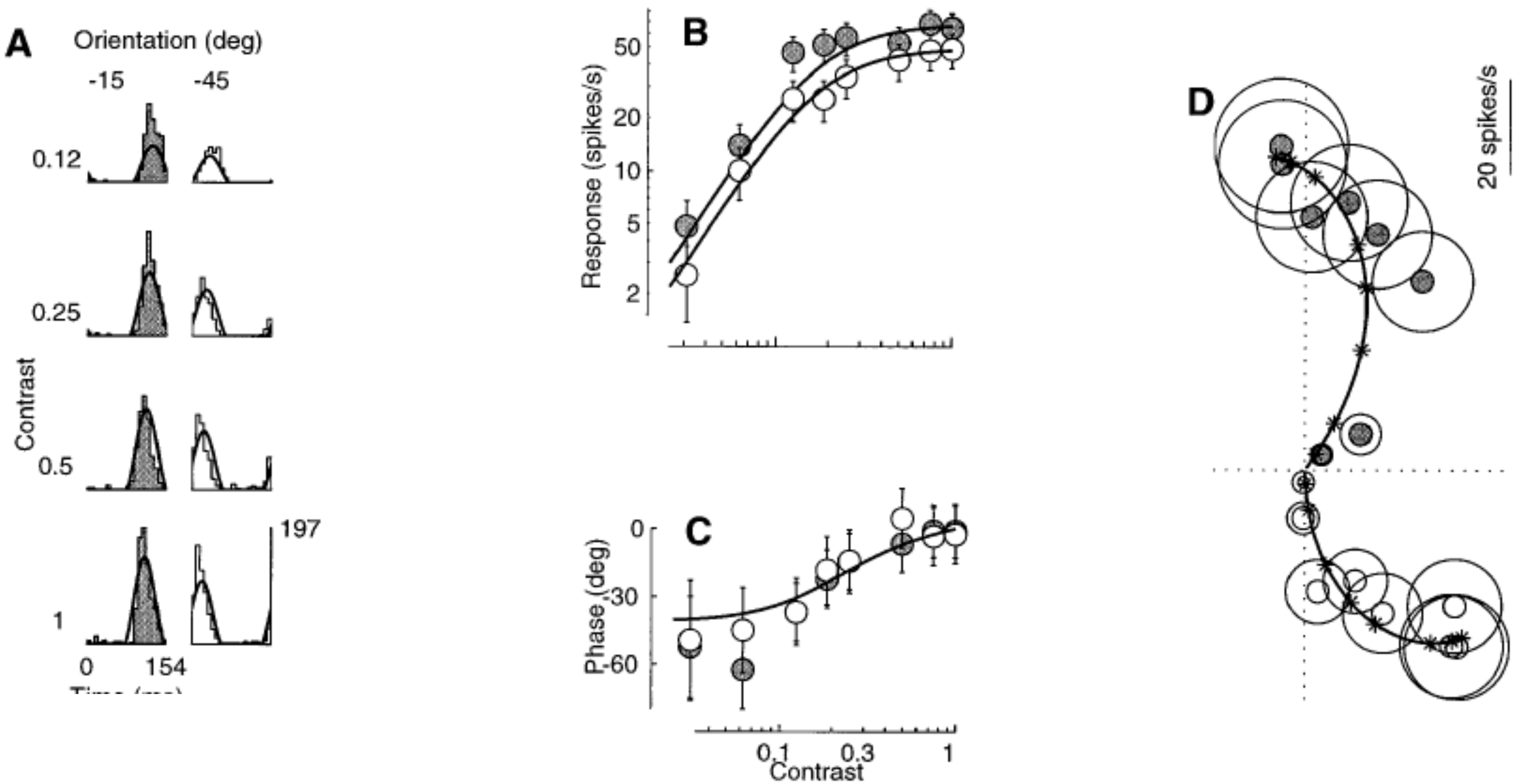
histogram of responses for different contrasts



response vs phase

each point is response to a different sinusoid

Now as a function of grating orientation

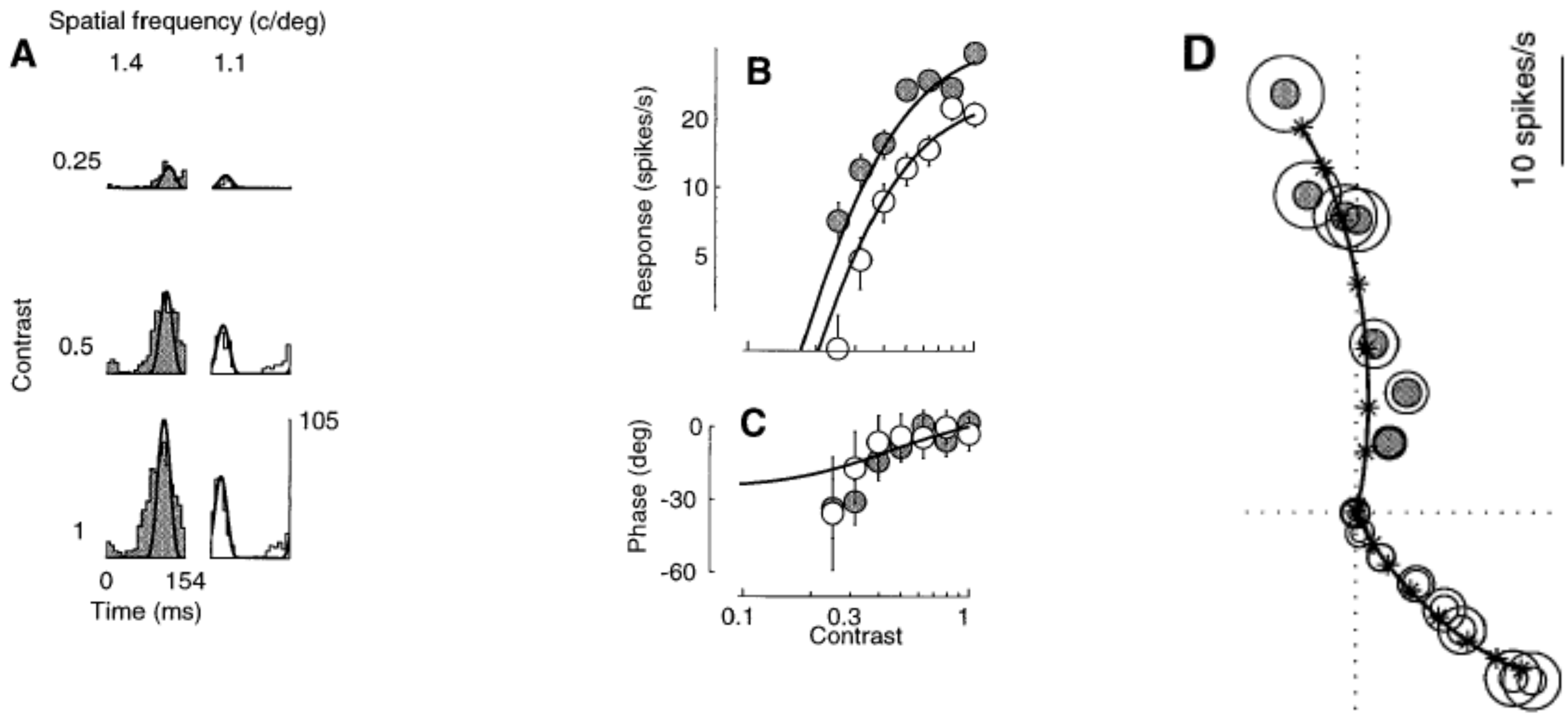


response vs phase

gray = -15deg, white = -45deg

Now as a function of spatial frequency

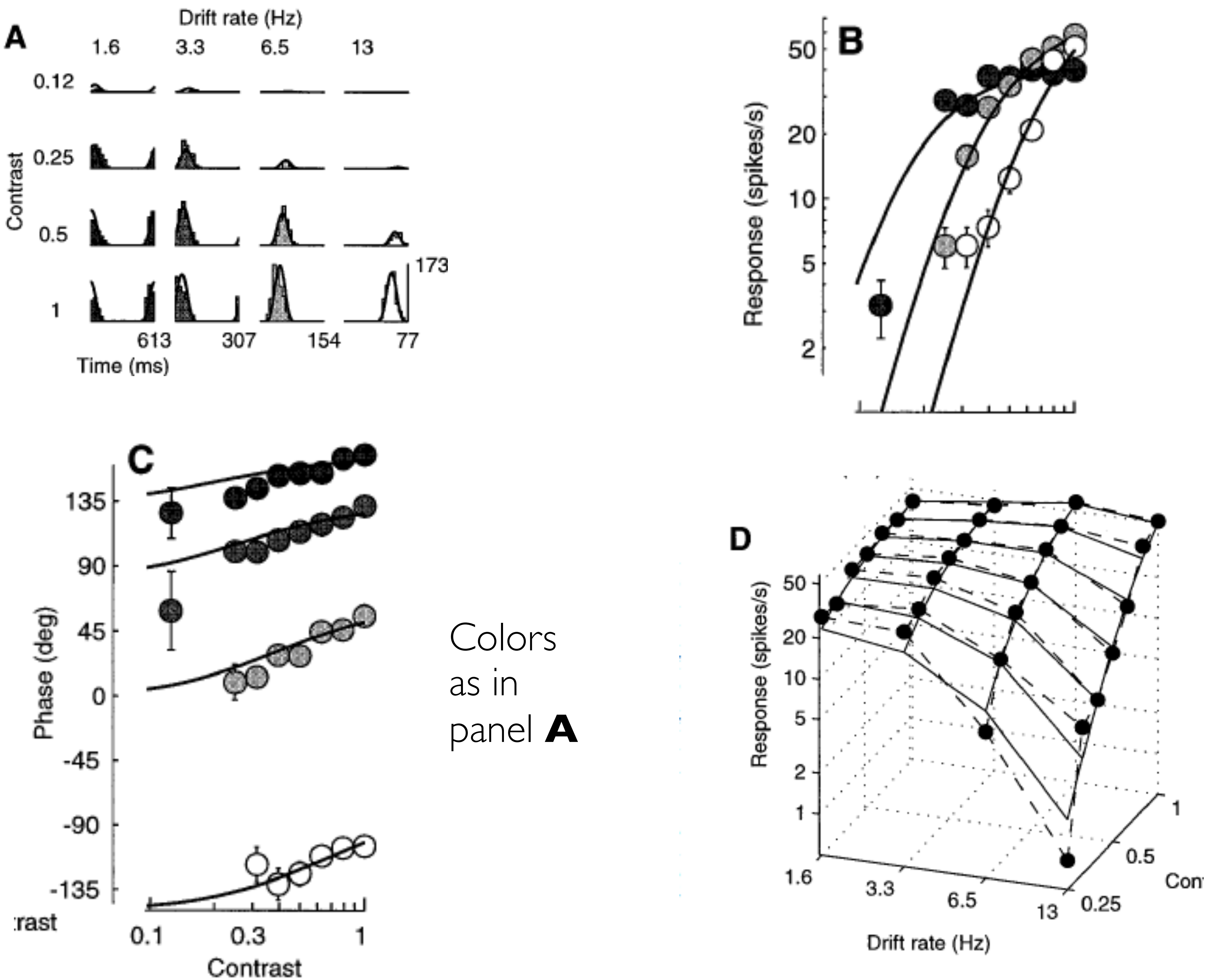
Carandini, Heeger and Movshon (1997)



gray = 1.4 cyc/deg, white = 1.1 cyc/deg

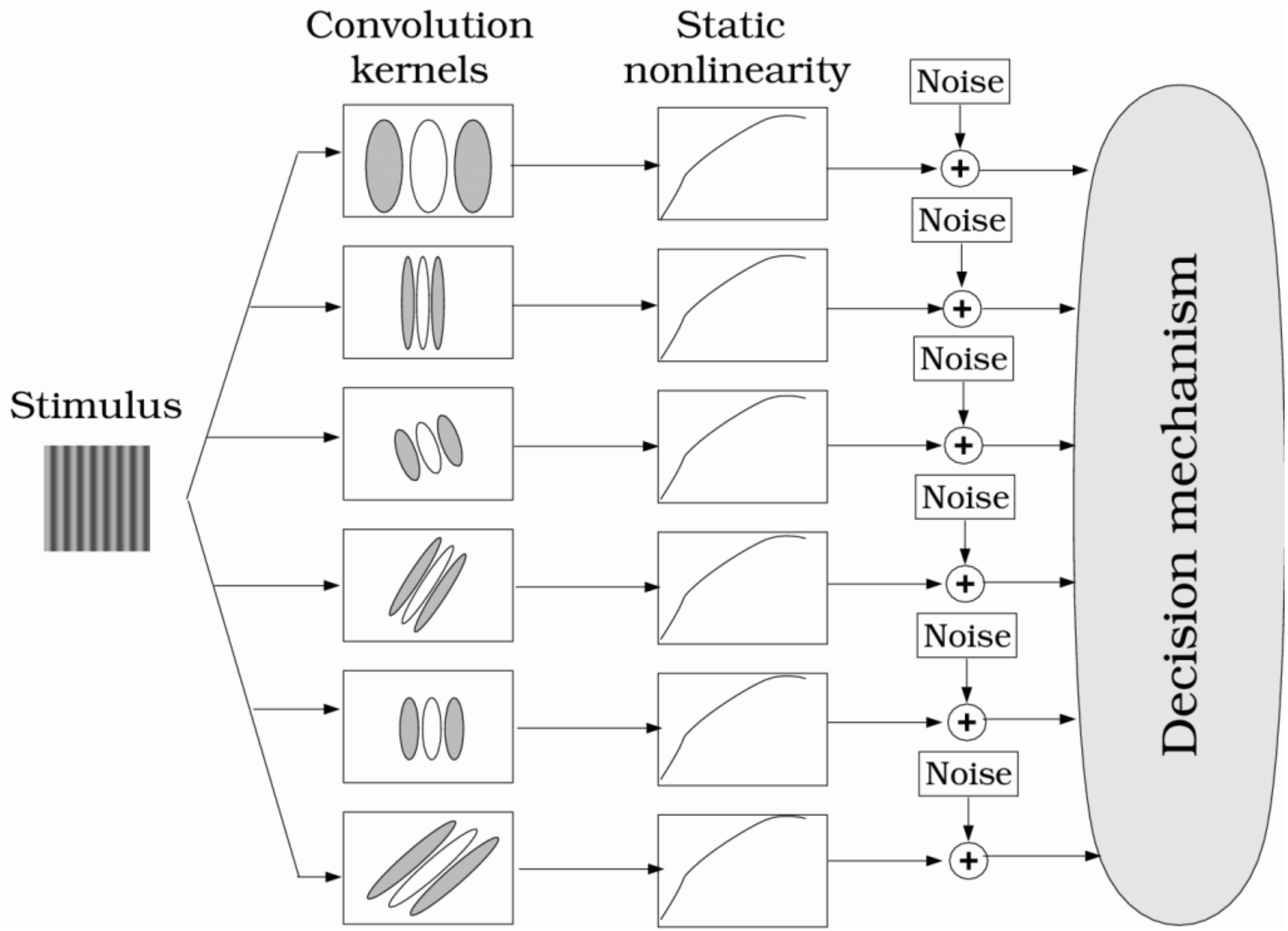
Now as a function of temporal frequency

Carandini, Heeger and Movshon (1997)



Colors as in panel **A**

Goal: Predictive Model of Ventral Stream

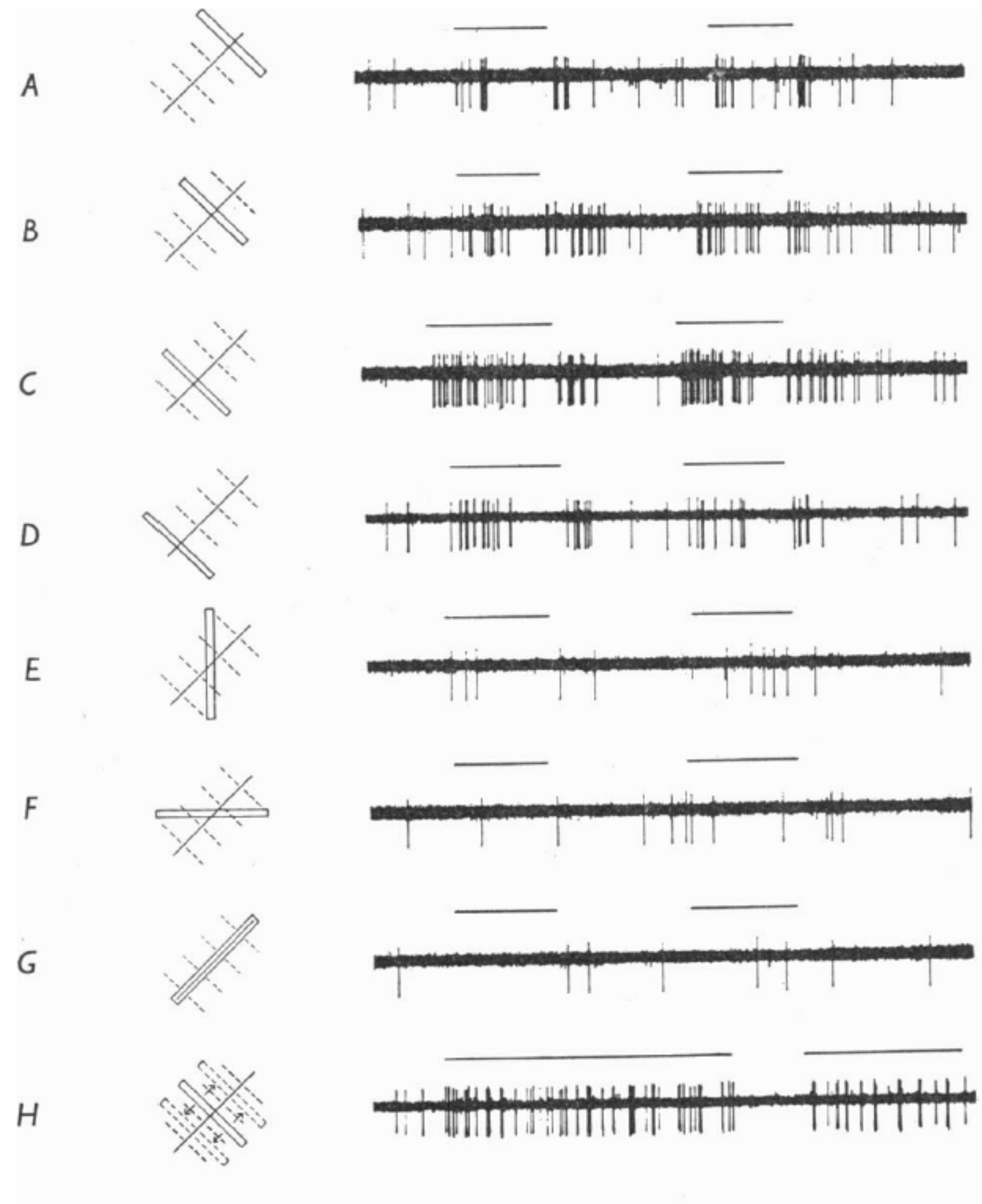


from Wandell 1996

Pooling

More generally, it was realized in computer vision that ***pooling*** was a good idea.

recall Hubel & Wiesel's complex cell >>



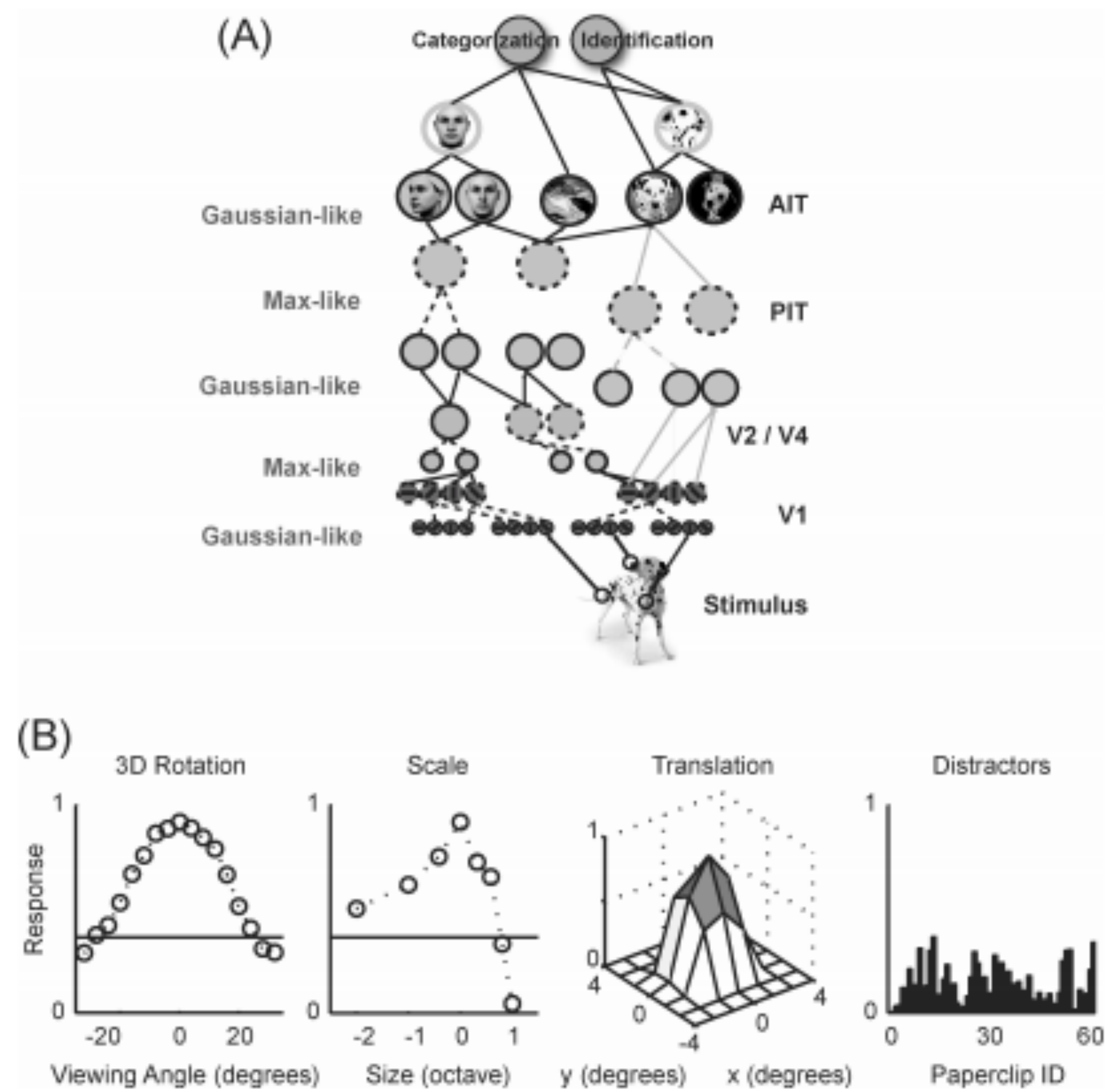
$$y = \left(\frac{1}{|N_r|} \sum_{i \in N_r} x_i^p \right)^{1/p}$$

*(Actually, if you're running a CNN, you basically *have* to do pooling + downsampling, for memory reasons.)*

Pooling

More generally, it was realized in computer vision that ***pooling*** was a good idea.

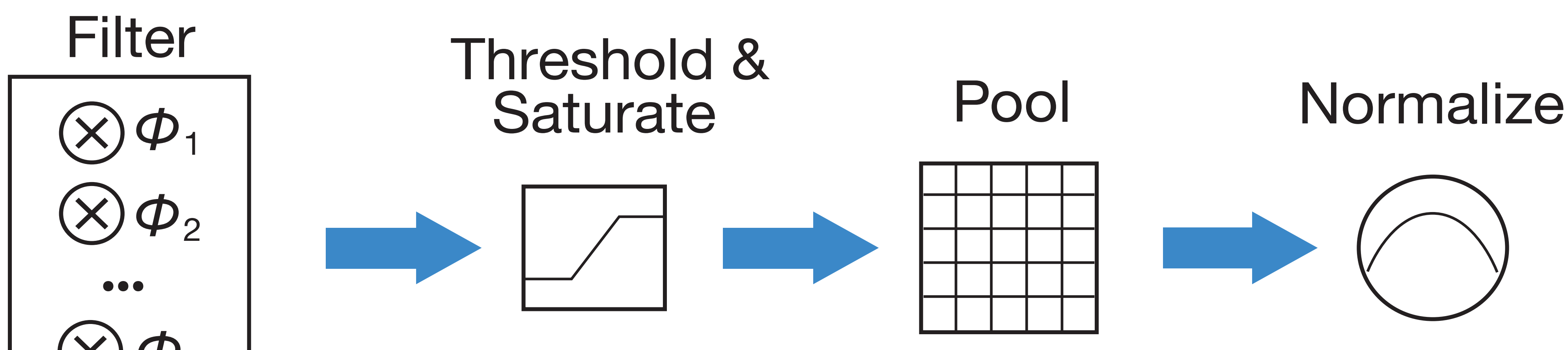
$$y = \left(\frac{1}{|N_r|} \sum_{i \in N_r} x_i^p \right)^{1/p}$$



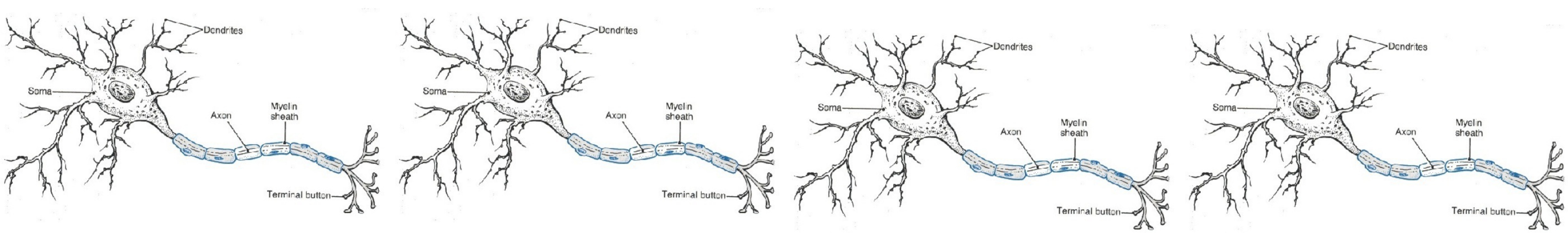
from Kouh and Poggio (2008)

*(Actually, if you're running a CNN, you basically *have* to do pooling + downsampling, for memory reasons.)*

- Linear-Nonlinear neurally-plausible **basic operations** within layer

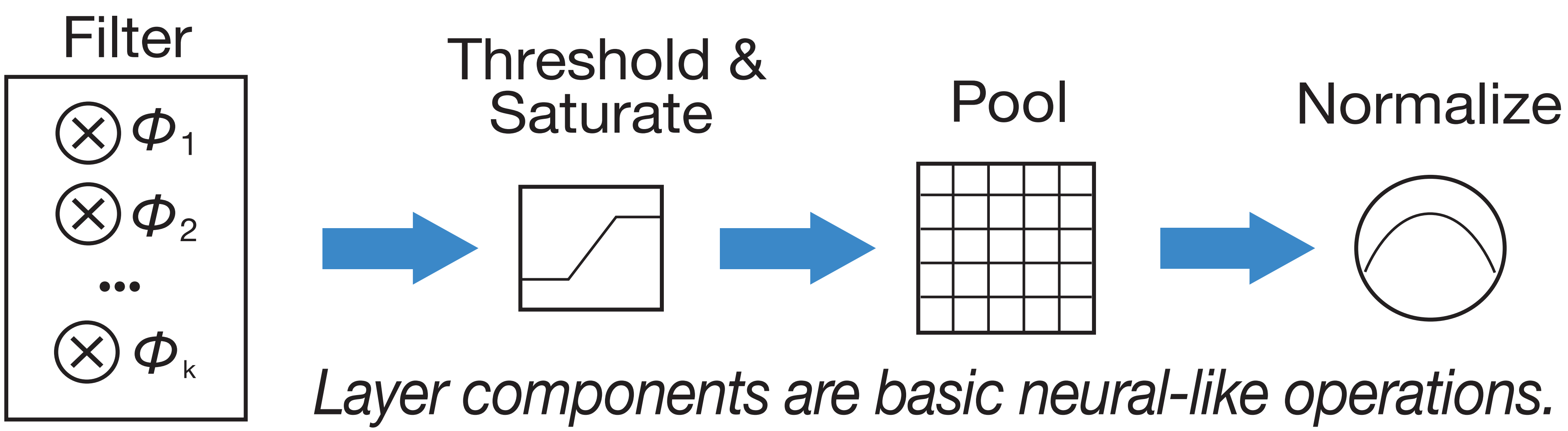


Layer components are basic neural-like operations.



Linear-Nonlinear Operations

- Linear-Nonlinear neurally-plausible **basic operations** within layer



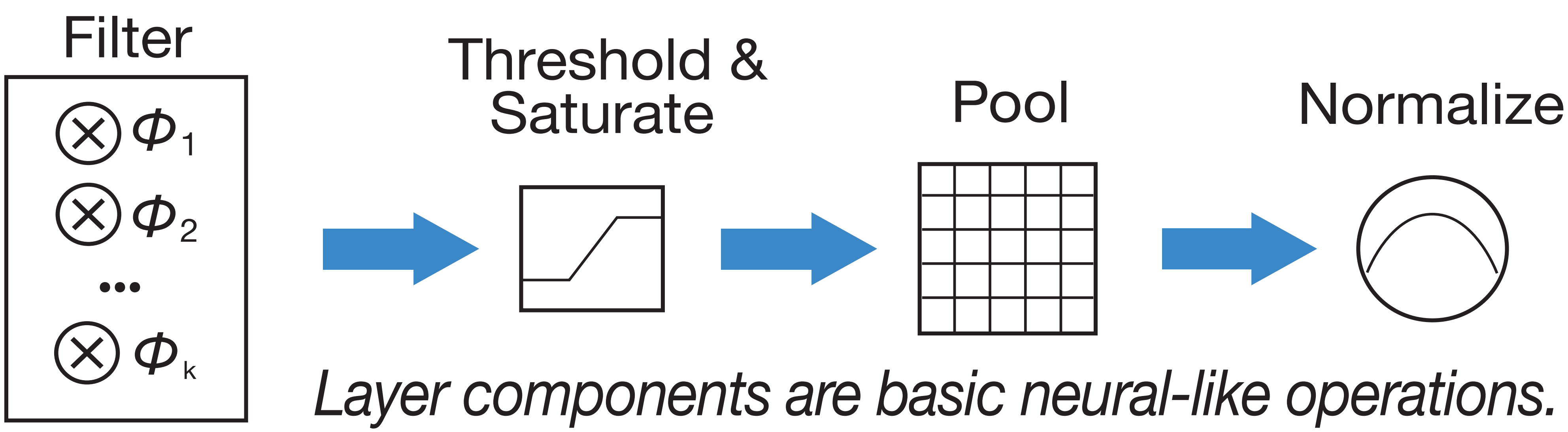
Layer components are basic neural-like operations.

neuro:	synaptic weights patterns
data:	untangling through dimension expansion

Hubel and Wiesel (1965-1975) Lecun (2004), Carandini et. al (2005), Lennie & Movshon (2005) , DiCarlo (2012)

Linear-Nonlinear Operations

- Linear-Nonlinear neurally-plausible **basic operations** within layer



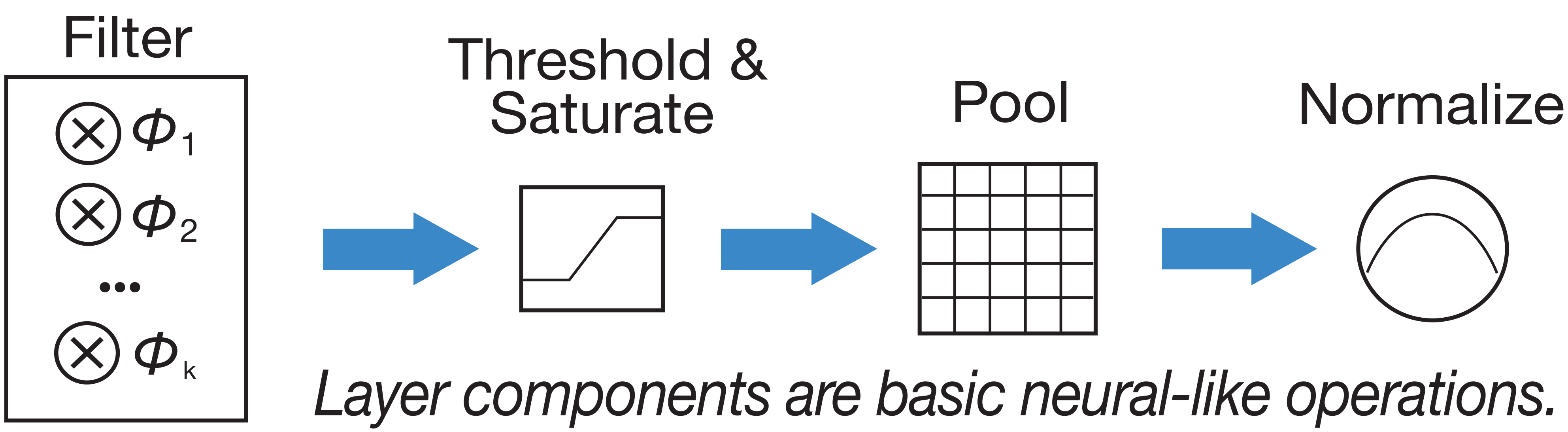
Layer components are basic neural-like operations.

	Filter	Threshold & Saturate
neuro:	synaptic weights patterns	single-unit activations
data:	untangling through dimension expansion	"AND" operation by limiting dynamic range

Hubel and Wiesel (1965-1975) Lecun (2004), Carandini et. al (2005), Lennie & Movshon (2005) , DiCarlo (2012)

Linear-Nonlinear Operations

- Linear-Nonlinear neurally-plausible **basic operations** within layer

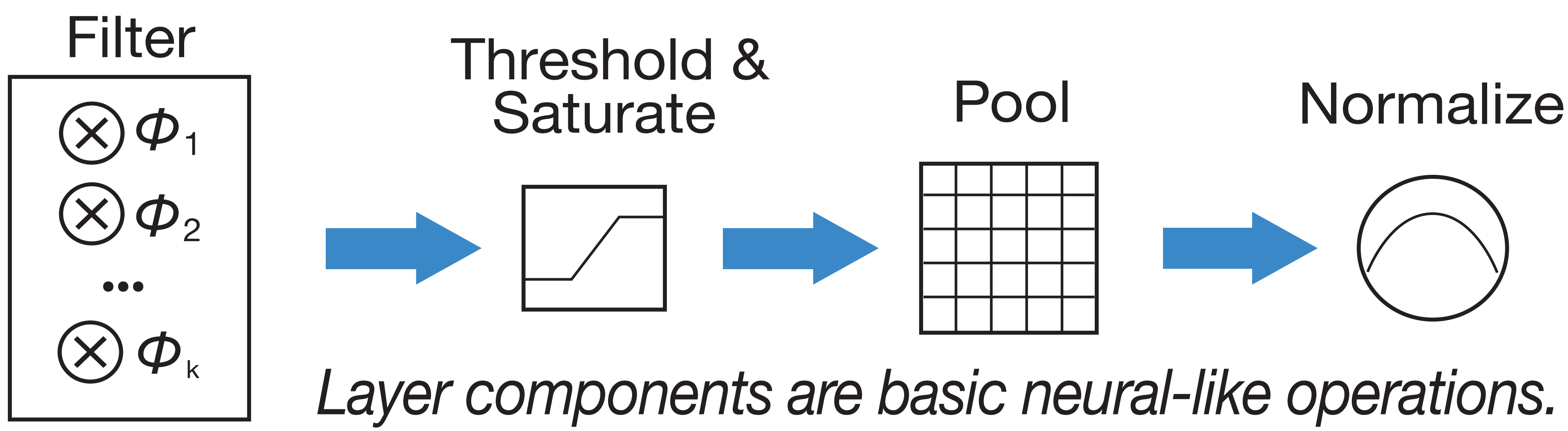


neuro:	synaptic weights patterns	single-unit activations	complex cells
data:	untangling through dimension expansion	"AND" operation by limiting dynamic range	adding robustness by dimension reduction

Hubel and Wiesel (1965-1975) Lecun (2004), Carandini et. al (2005), Lennie & Movshon (2005) , DiCarlo (2012)

Linear-Nonlinear Operations

- Linear-Nonlinear neurally-plausible **basic operations** within layer



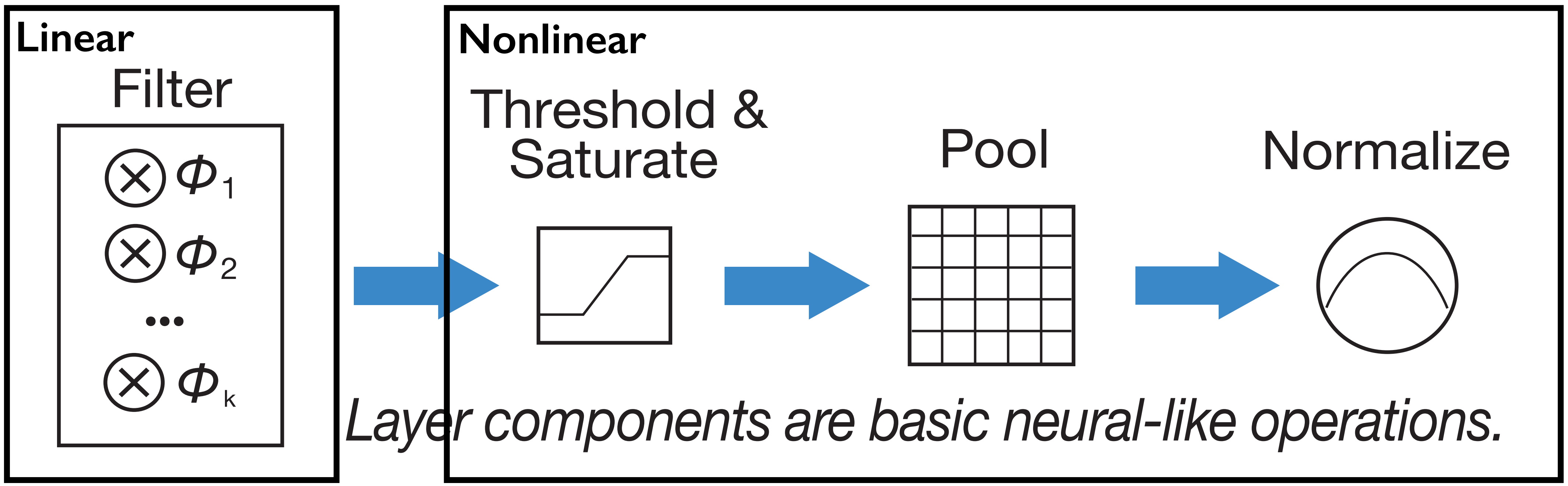
Layer components are basic neural-like operations.

neuro:	synaptic weights patterns	single-unit activations	complex cells	competitive inhibition
data:	untangling through dimension expansion	"AND" operation by limiting dynamic range	adding robustness by dimension reduction	put results back into standard range

Hubel and Wiesel (1965-1975) Lecun (2004), Carandini et. al (2005), Lennie & Movshon (2005) , DiCarlo (2012)

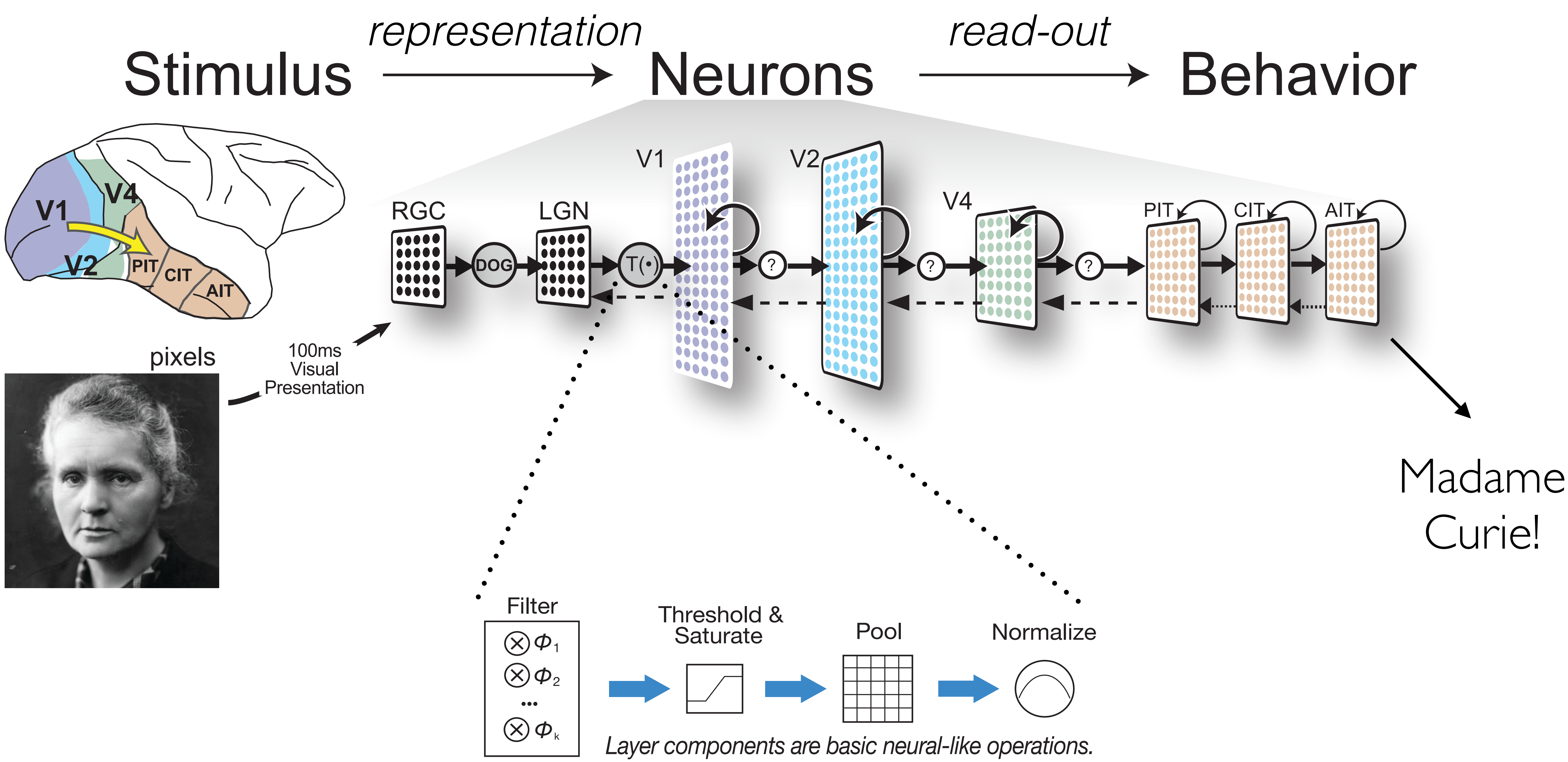
Linear-Nonlinear Operations

- Linear-Nonlinear neurally-plausible **basic operations** within layer



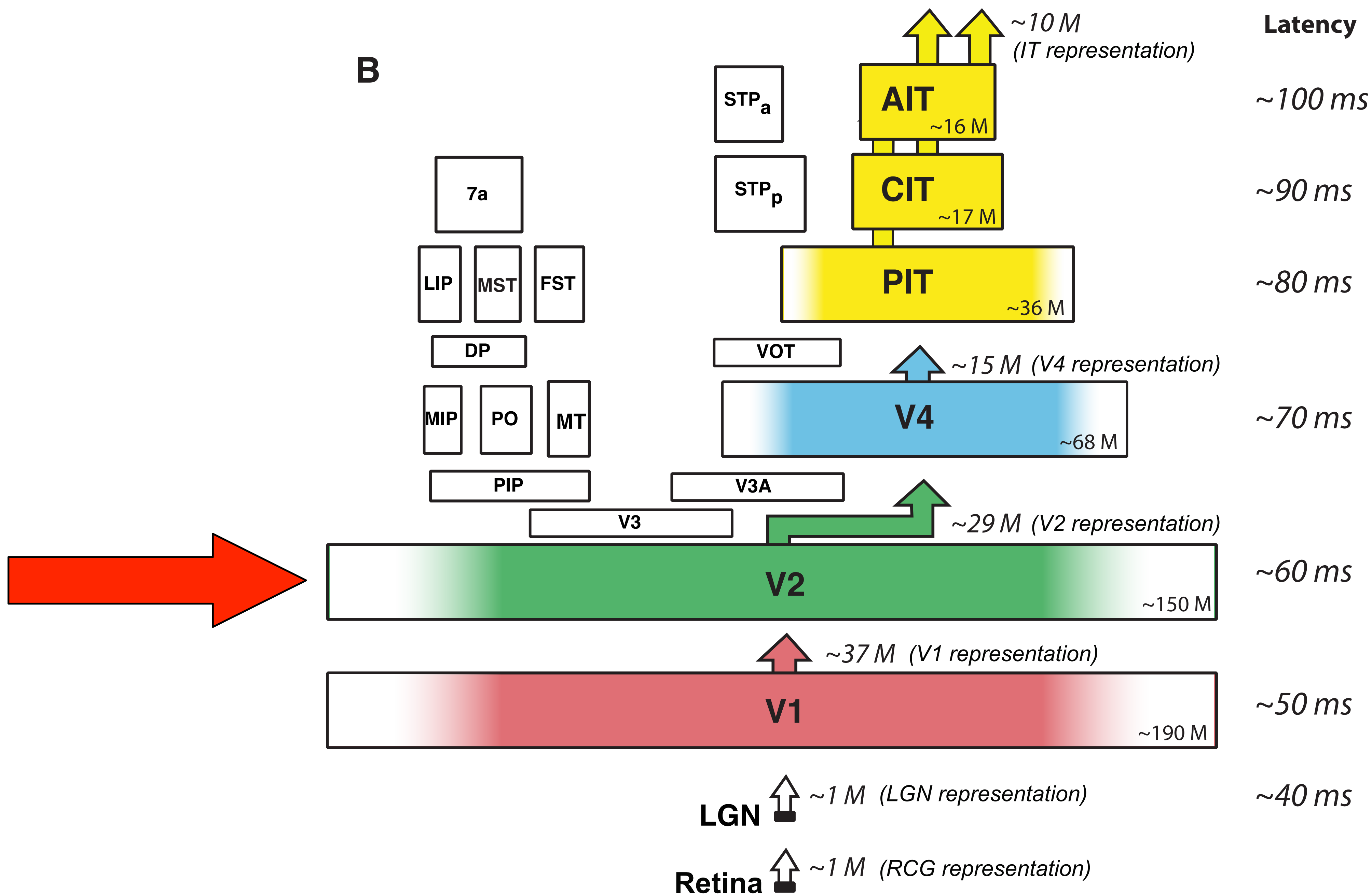
neuro:	synaptic weights patterns	single-unit activations	complex cells	competitive inhibition
data:	untangling through dimension expansion	"AND" operation by limiting dynamic range	adding robustness by dimension reduction	put results back into standard range

Hubel and Wiesel (1965-1975) Lecun (2004), Carandini et. al (2005), Lennie & Movshon (2005) , DiCarlo (2012)



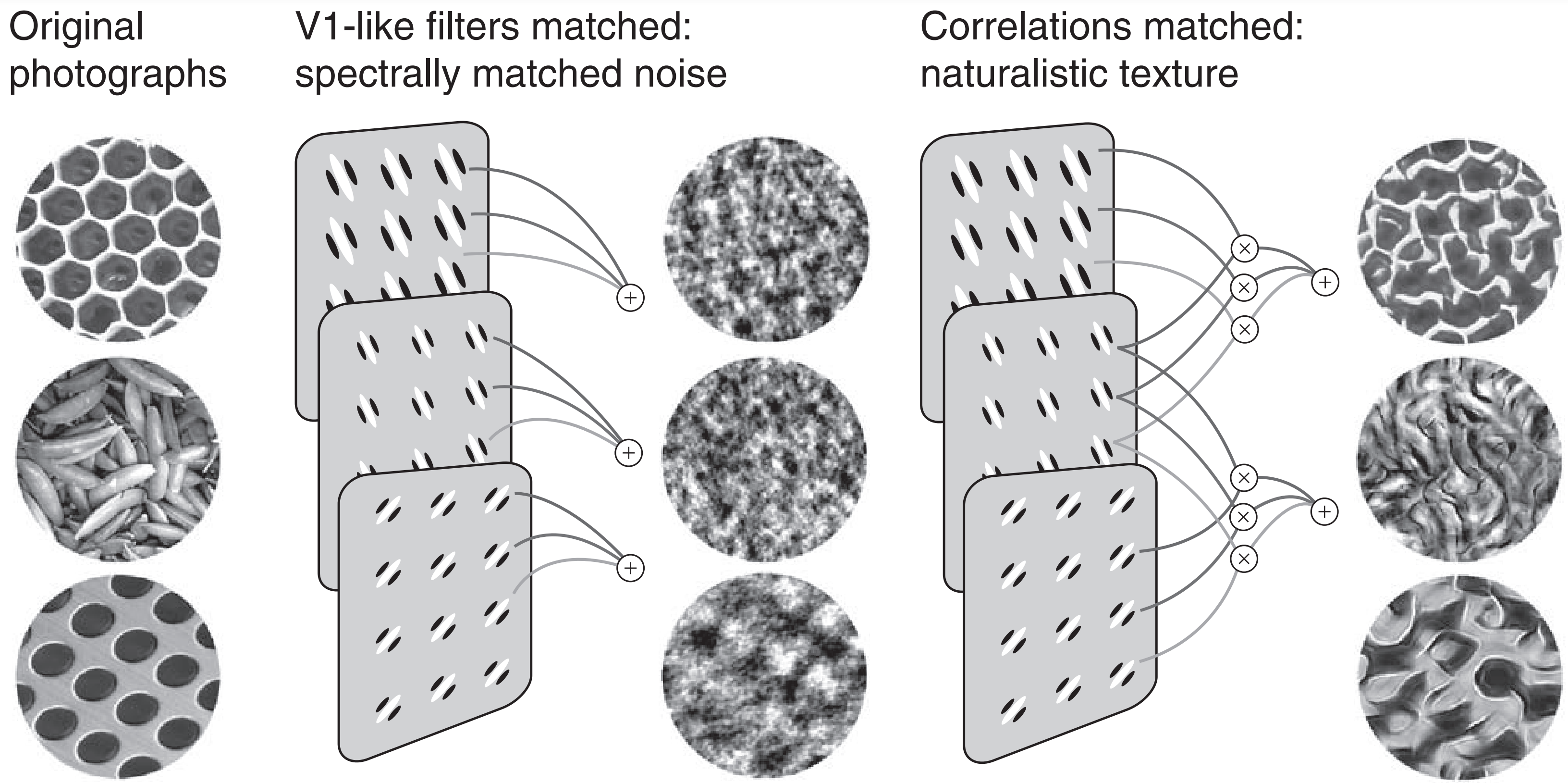
Linear-Nonlinear neurally-plausible **basic operations** within layer

You are here.



Adapted from DiCarlo et al. 2012

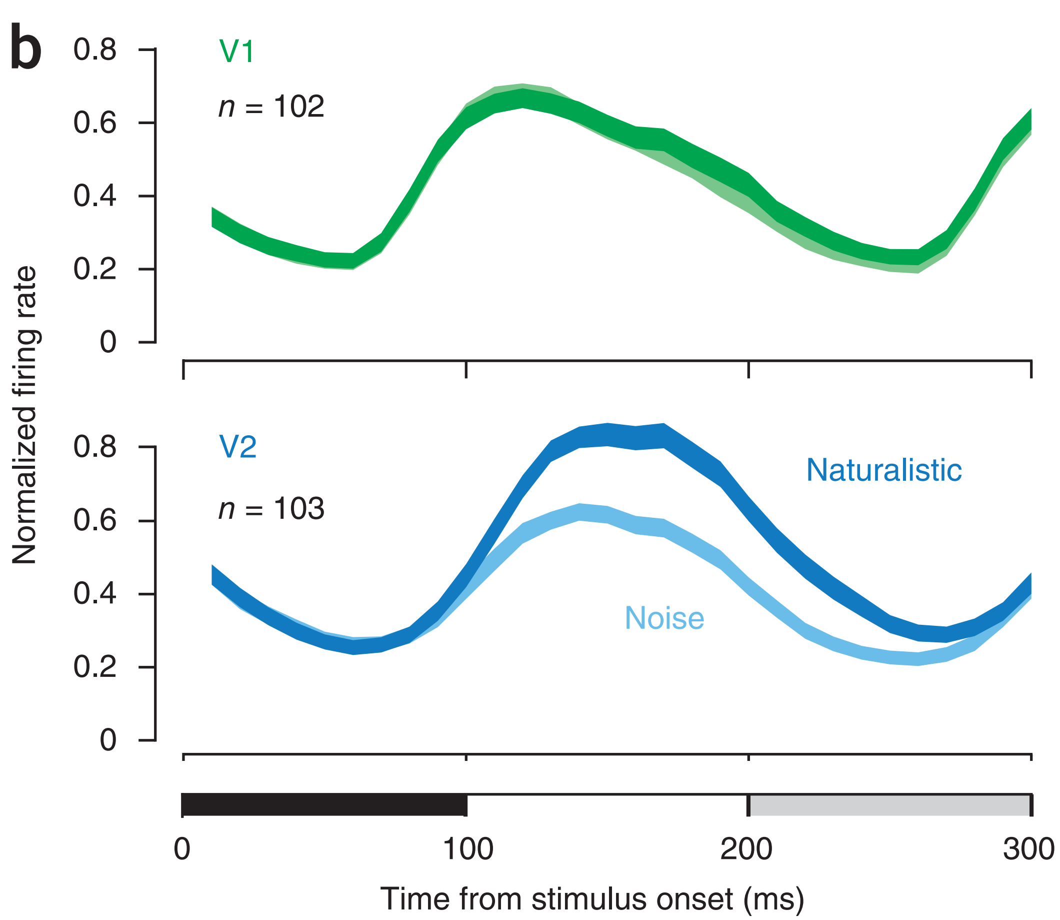
Area V2 (first cortical area after V1):




Eero Simoncelli


Tony Movshon


Jeremy Freeman



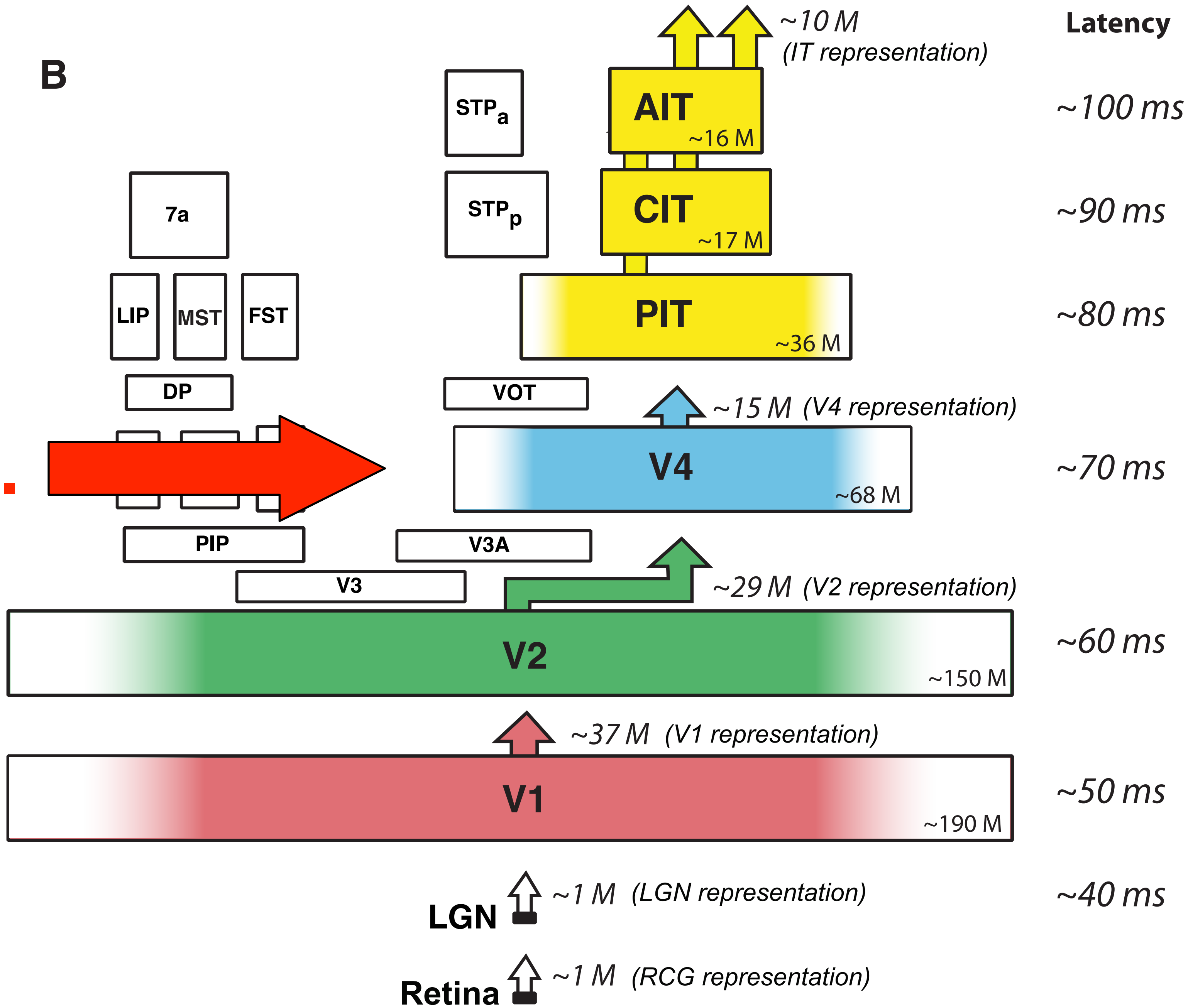
Adapted from Freeman, Ziemba, Heeger, Simoncelli, & Movshon, Nature Neuro (2013)

Interpretation:

- V2 neurons apply "and-like" operators on V1 outputs
- those "ands" are tuned toward natural co-occurring V1 statistics

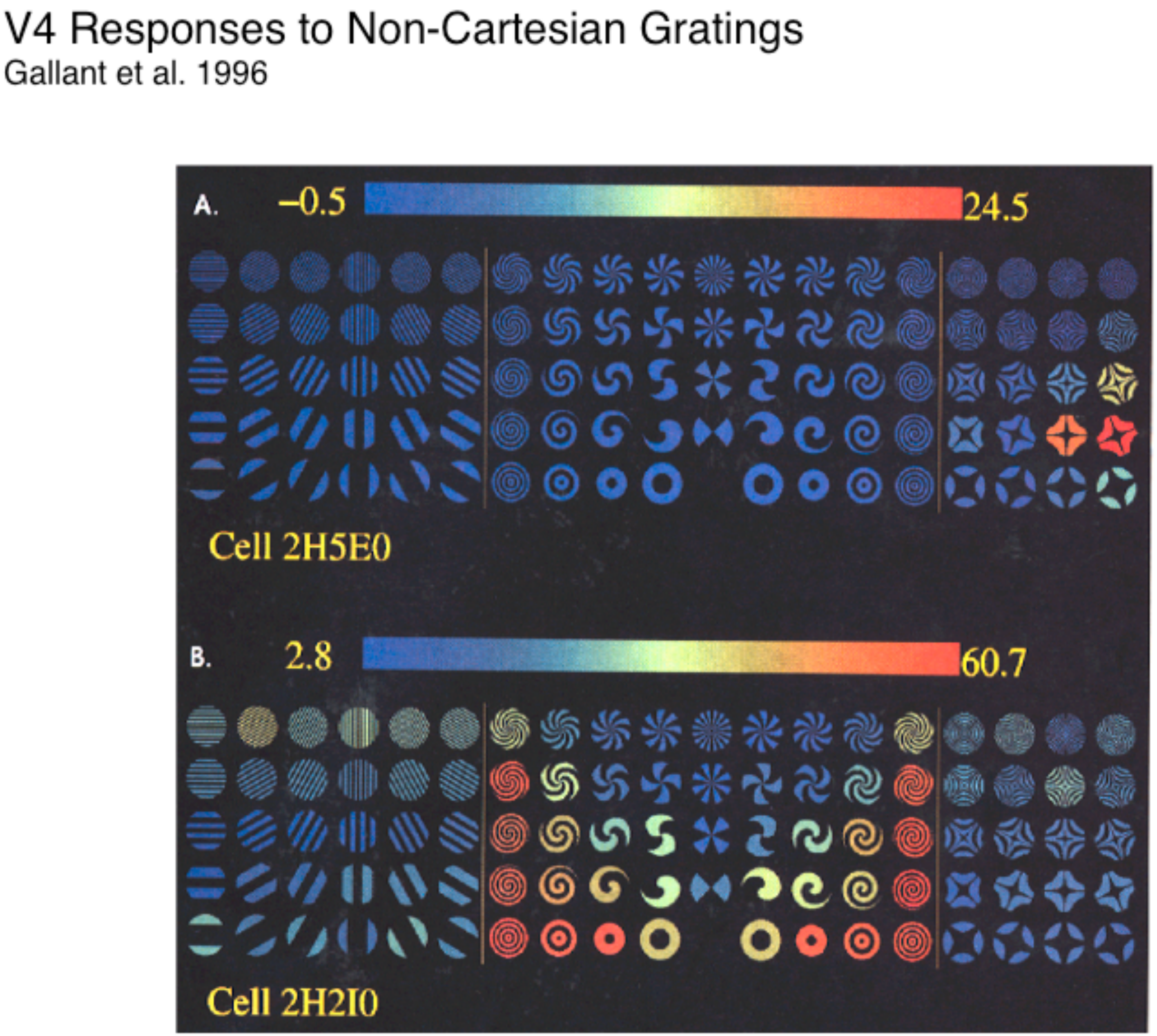
So, maybe a <u>hierarchically-built</u> sparse auto-encoding in a 2-layer model with max pooling??

You are here.



Adapted from DiCarlo et al. 2012

Area V4 (cortical area after V2):



Jack Gallant

Area V4 (cortical area after V2):

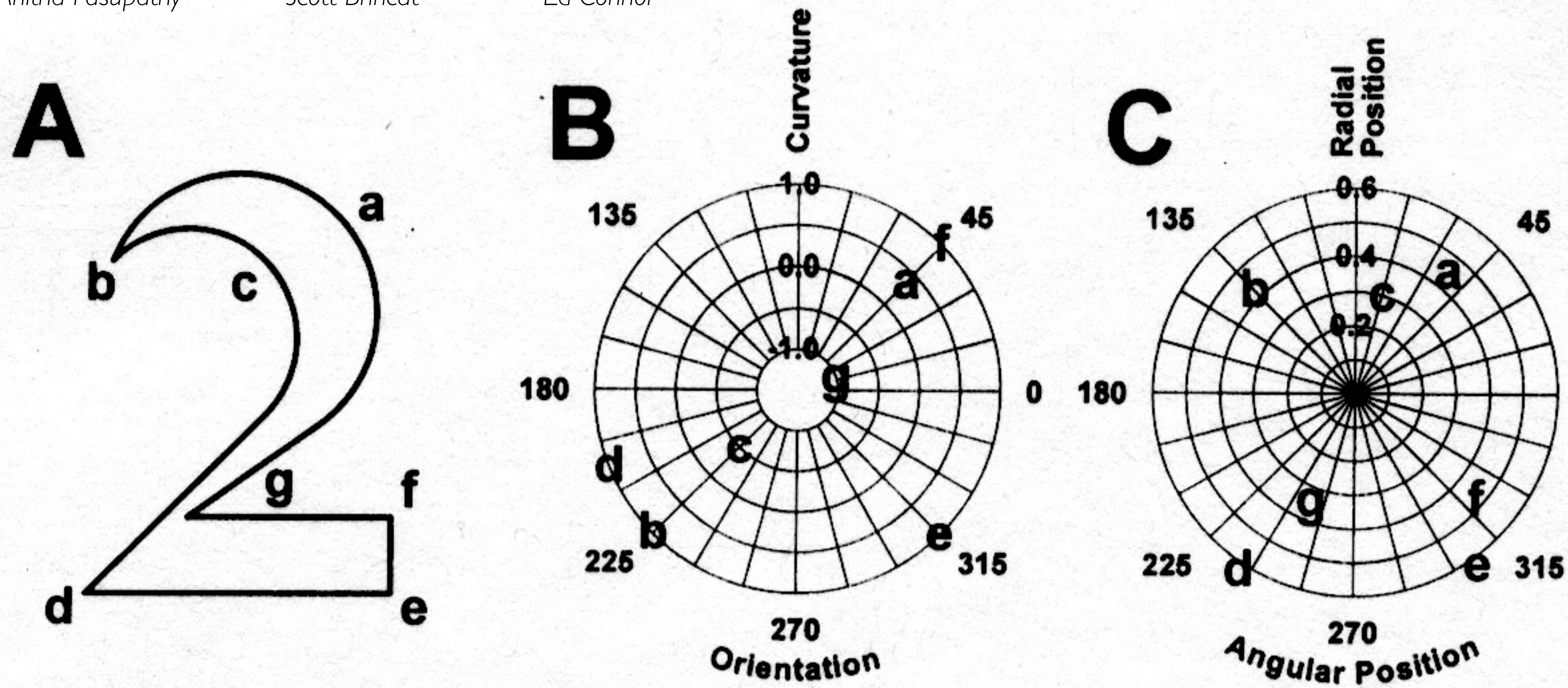
Anitha Pasupathy

Scott Brincat

Ed Connor

<u>Make a basis for shapes</u>:
each shape = set of curved elements
each element = (ang position, curvature)

<u>Hypothesis</u>:
V4 neurons are tuned in this basis



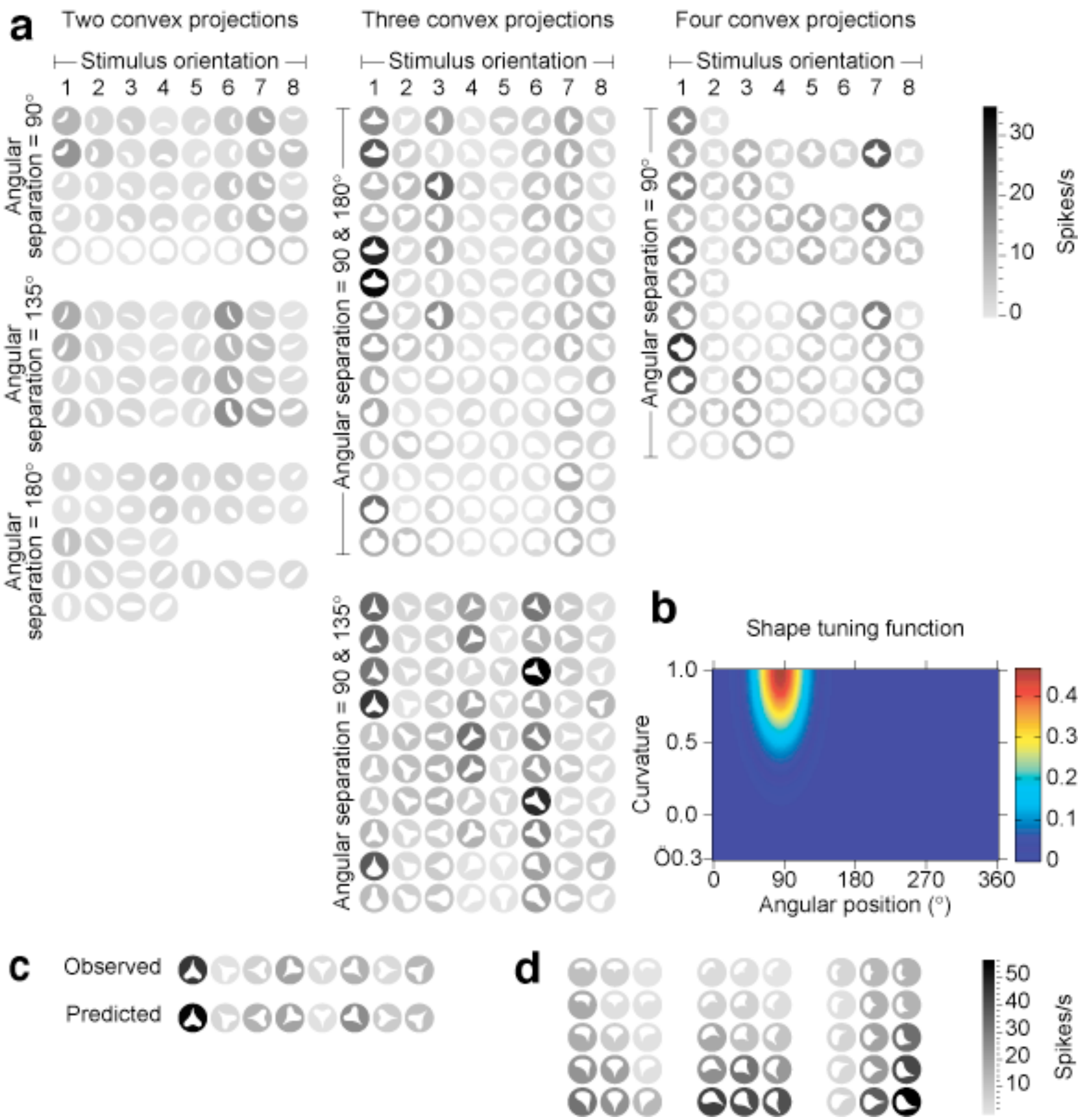
A structural (parts-based) shape-coding scheme based on contour fragments. *A*, The example shape, a bold numeral 2, can be decomposed into contour fragments (*a-g*) with different curvatures, orientations, and positions. *B*, The curvature and orientation of each contour fragment is plotted on a 2-D domain. *C*, The positions of the contour fragments (relative to the object center) are plotted on a 2-D domain. Together, plots *B* and *C* represent a 4-D domain for describing contour fragments.

What shape features drive V4 response?

Adapted from C.E. Connor

<u>Make a basis for shapes</u>:
each shape = set of curved elements
each element = (ang position, curvature)

<u>Hypothesis</u>:
V4 neurons are tuned in this basis



Pasupathy and Connor (V4)
Brincat and Connor (PIT)

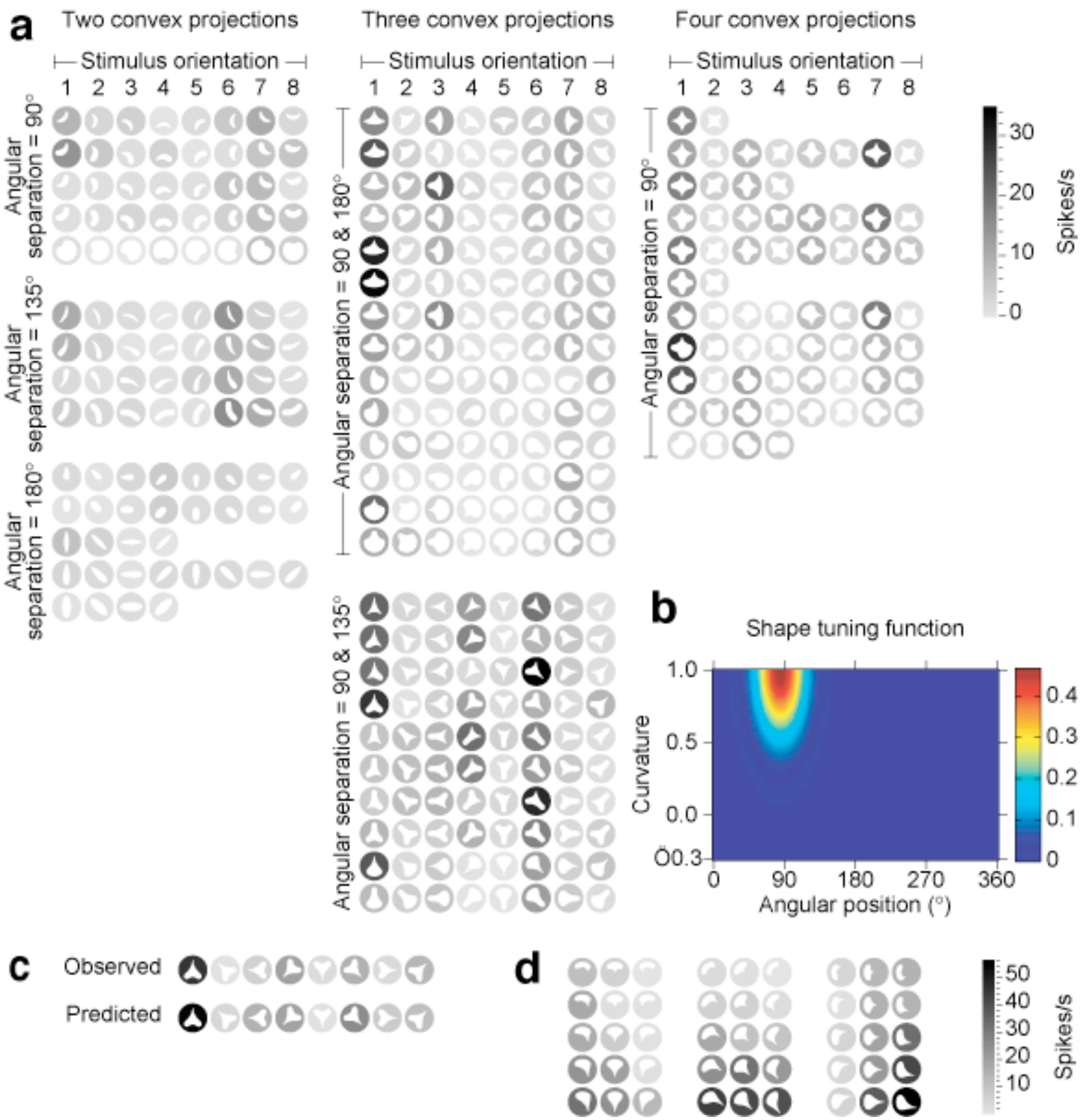
What shape features drive V4 response?

Adapted from C.E. Connor

<u>Make a basis for shapes</u>:
each shape = set of curved elements
each element = (ang position, curvature)

<u>Hypothesis</u>:
V4 neurons are tuned in this basis

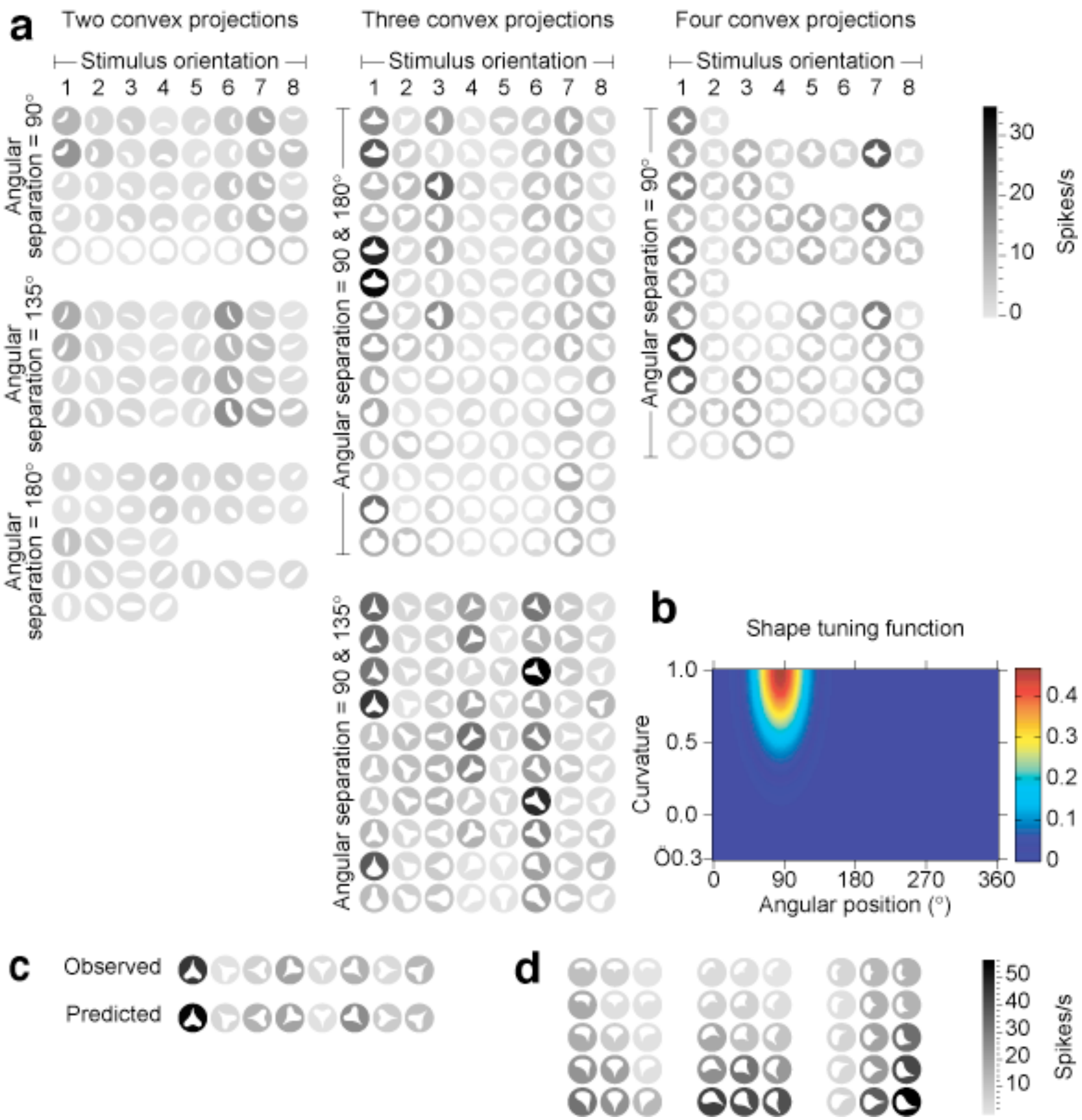
<u>Experimental result</u>:
Hypothesis explains ~50% of the explainable response variance for these types of stimuli



Pasupathy and Connor (V4)
Brincat and Connor (PIT)

What shape features drive V4 response?

Adapted from C.E. Connor



<u>Make a basis for shapes</u>:
each shape = set of curved elements
each element = (ang position, curvature)

<u>Hypothesis</u>:
V4 neurons are tuned in this basis

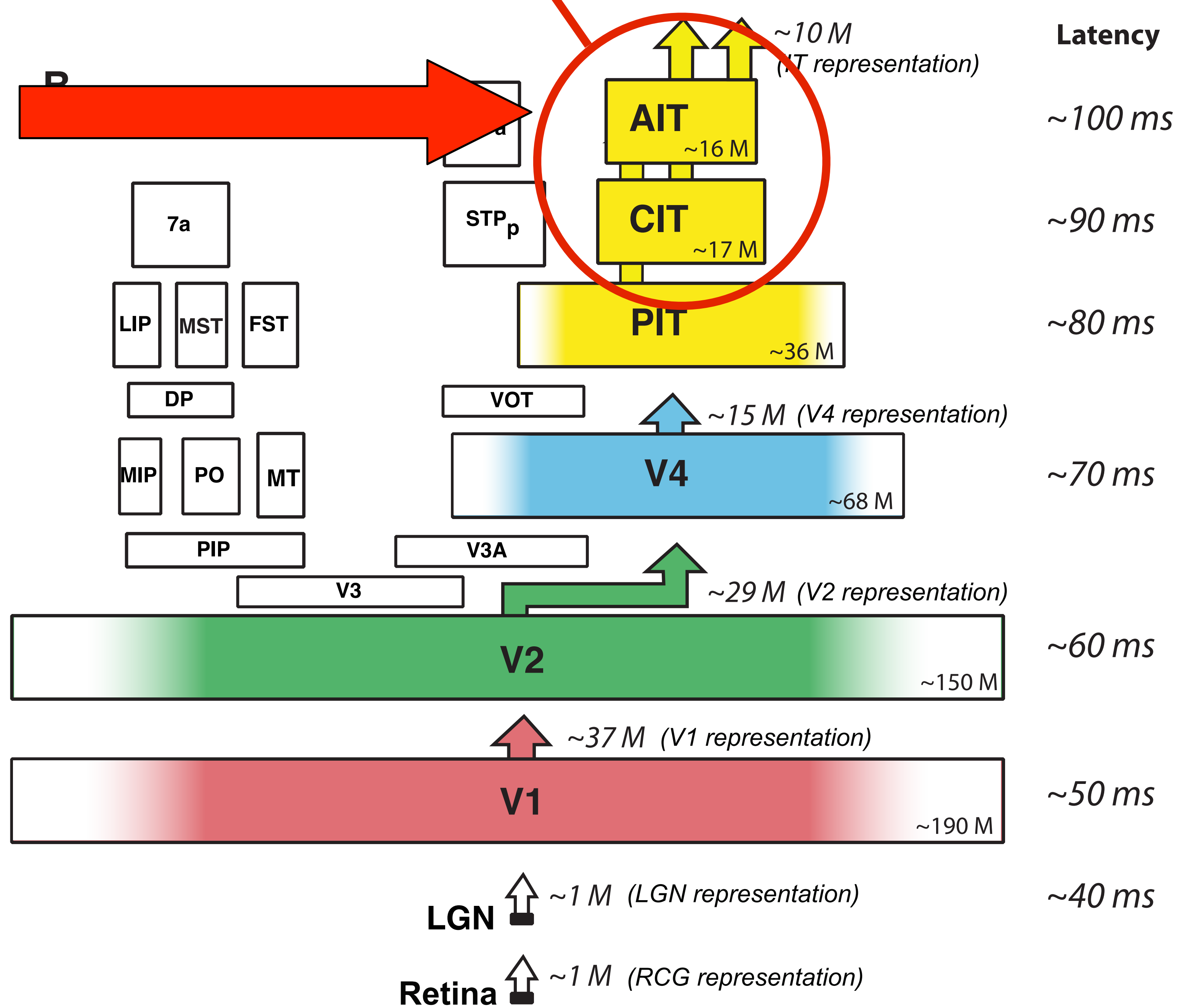
<u>Experimental result</u>:
Hypothesis explains ~50% of the explainable response variance for these types of stimuli

Problem:
No predictions for any other images.
i.e.
is not an "image-computable" model

Pasupathy and Connor (V4)
Brincat and Connor (PIT)

“IT” (Inferior temporal cortex)

You are here.



Adapted from DiCarlo et al. 2012

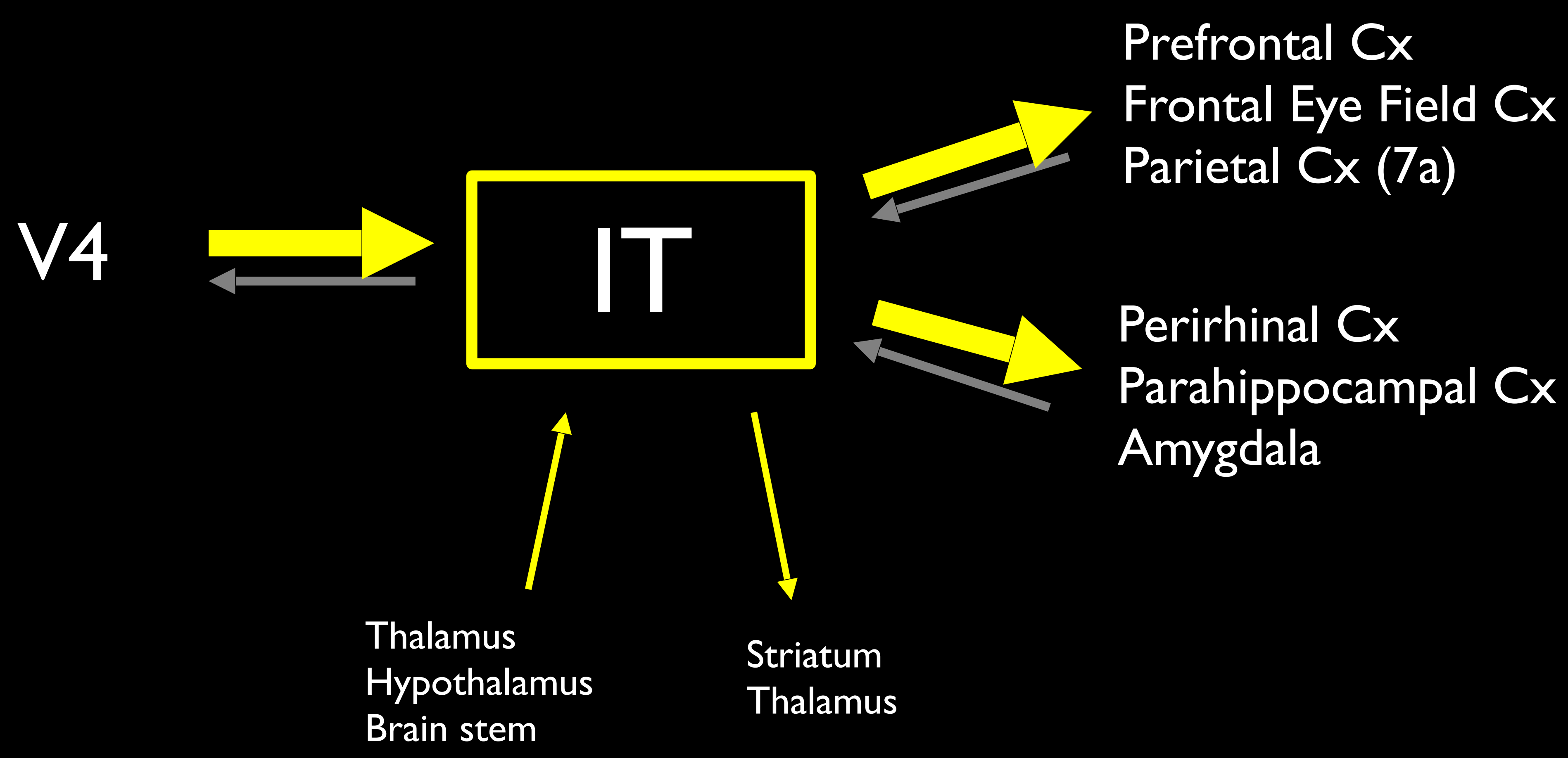
IT statistics (rhesus monkey)

~ 7.7 cm^2

~ 8% of neocortex (~ 15% of visual cortex)

~ 90 million neurons

Subregions: (PIT, CIT, AIT) (TEO, TE)



Stimulus selectivity in inferotemporal cortex

Gross, Rocha-Miranda & Bender 1972

Increasing ability to drive this IT neuron -->

The use of [these] stimuli was begun one day when, having failed to drive a unit with any light stimulus, we waved a hand at the stimulus screen and elicited a very vigorous response from the previously unresponsive neuron...

We then spent the next 12 hr testing various paper cutouts in an attempt to find the trigger feature for this unit. When the entire set of stimuli used were ranked according to the strength of the response that they produced, we could not find a simple physical dimension that correlated with this rank order. However, the rank order of adequate stimuli did correlate with similarity (for us) to the shadow of a monkey hand" (Gross et al., 1972).

Stimulus selectivity in inferotemporal cortex
Gross, Rocha-Miranda & Bender 1972

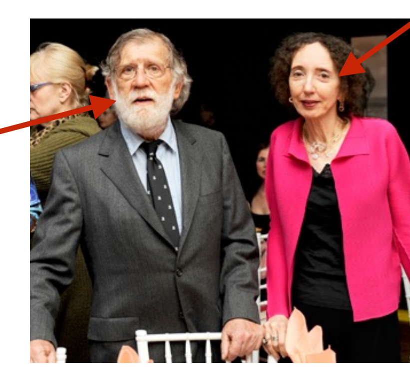
Increasing ability to drive this IT neuron -->

The use of [these] stimuli was begun one day when, having failed to drive a unit with any light stimulus, we waved a hand at the stimulus screen and elicited a very vigorous response from the previously unresponsive neuron...

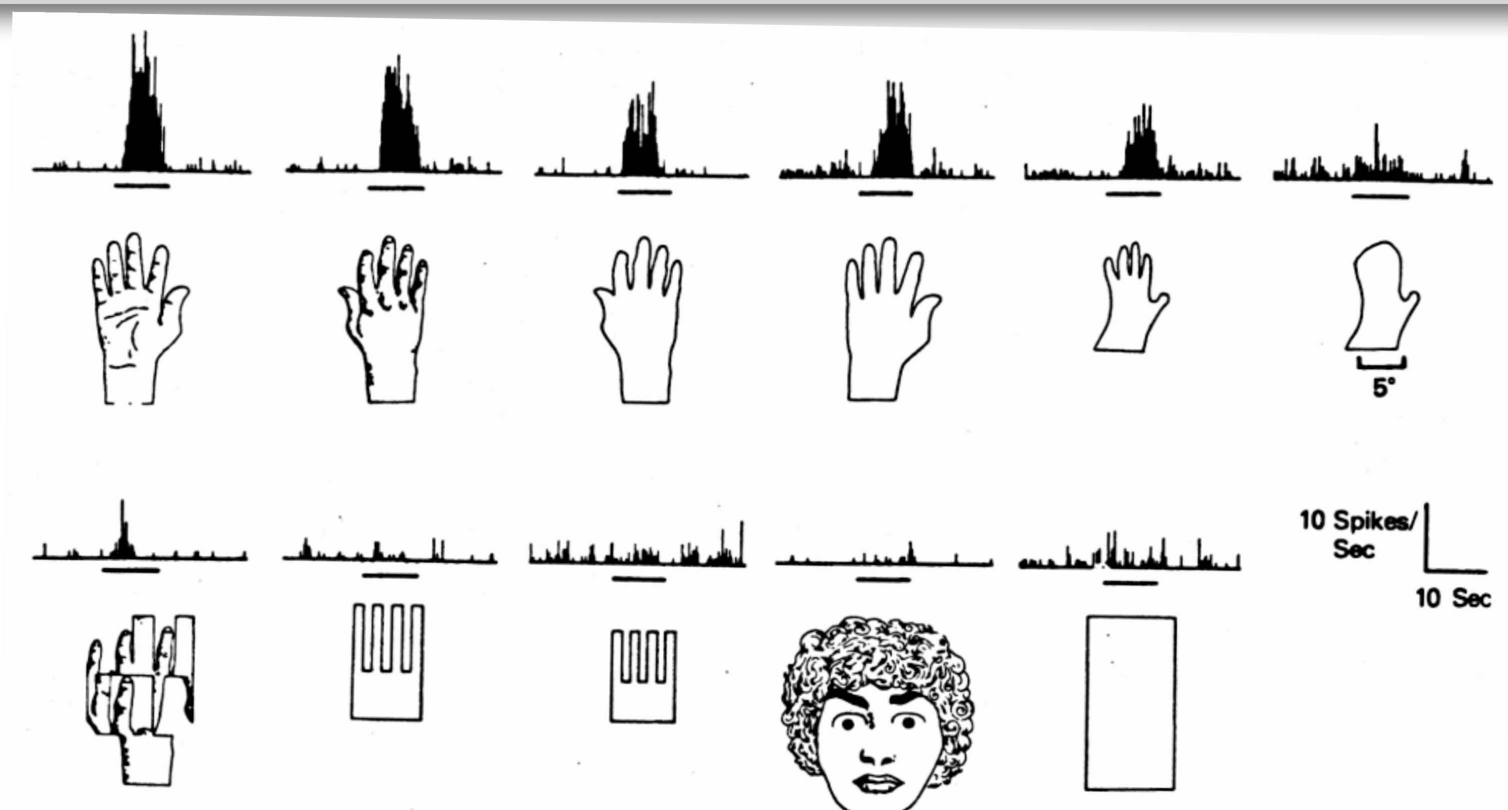
We then spent the next 12 hr testing various paper cutouts in an attempt to find the trigger feature for this unit. When the entire set of stimuli used were ranked according to the strength of the response that they produced, we could not find a simple physical dimension that correlated with this rank order. However, the rank order of adequate stimuli did correlate with similarity (for us) to the shadow of a monkey hand" (Gross et al., 1972).

Joyce Carol Oates!

Charlie Gross

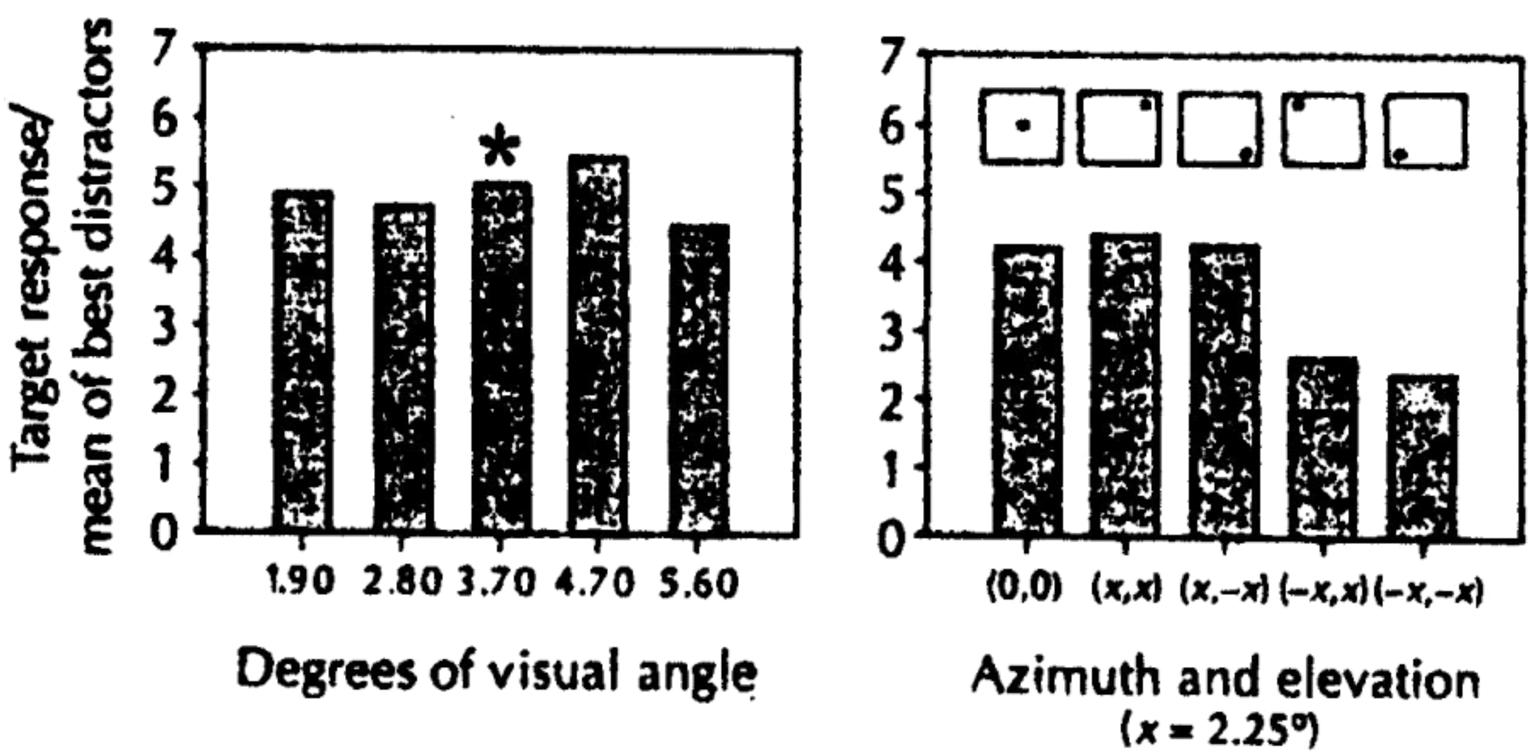


What stimulus feature are IT neurons actually "tuned" to?



Desimone et al. (1984)

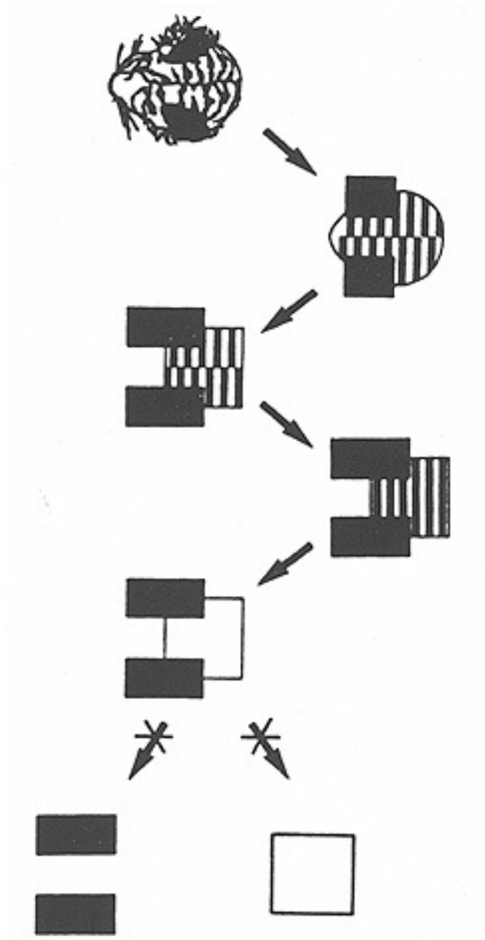
IT neurons can be tuned to specific combinations of features (high "selectivity")

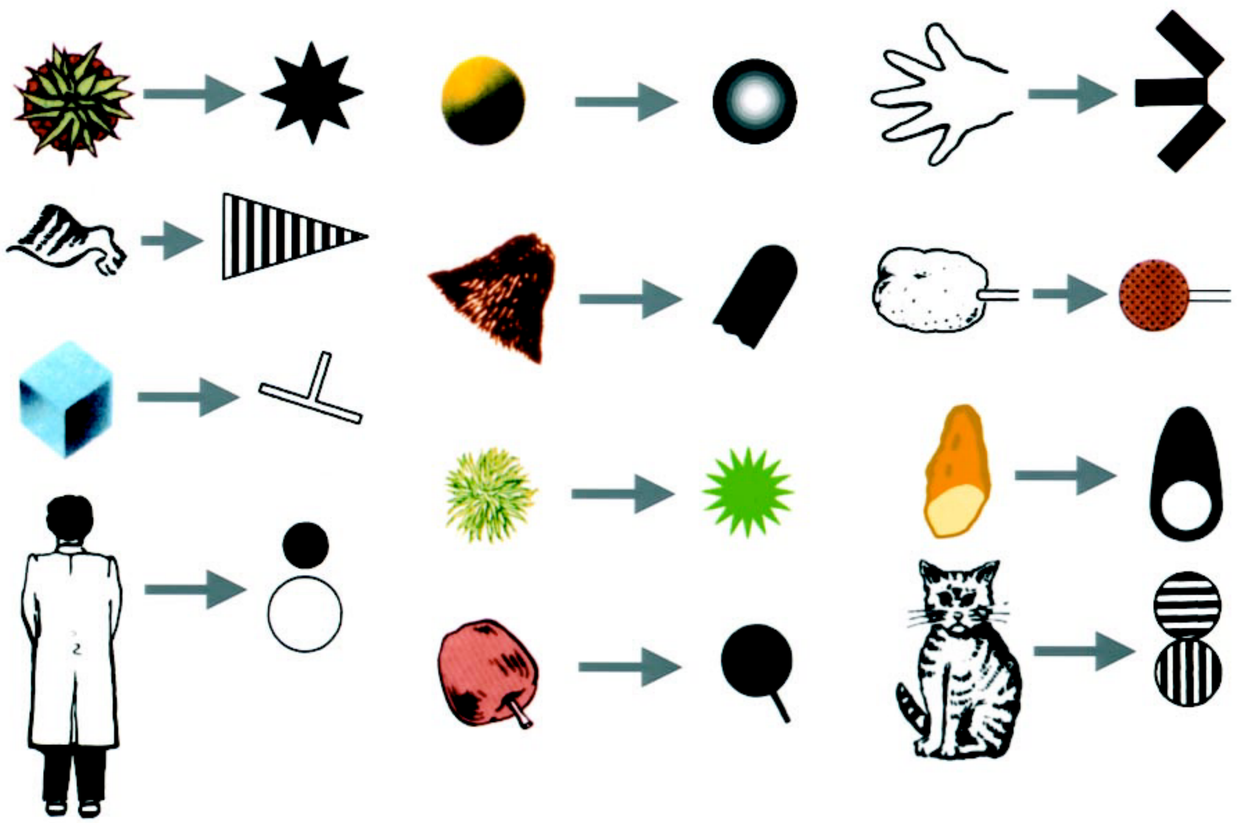


Logothetis et al. (1995)

That selectivity is tolerant to changes in position and size

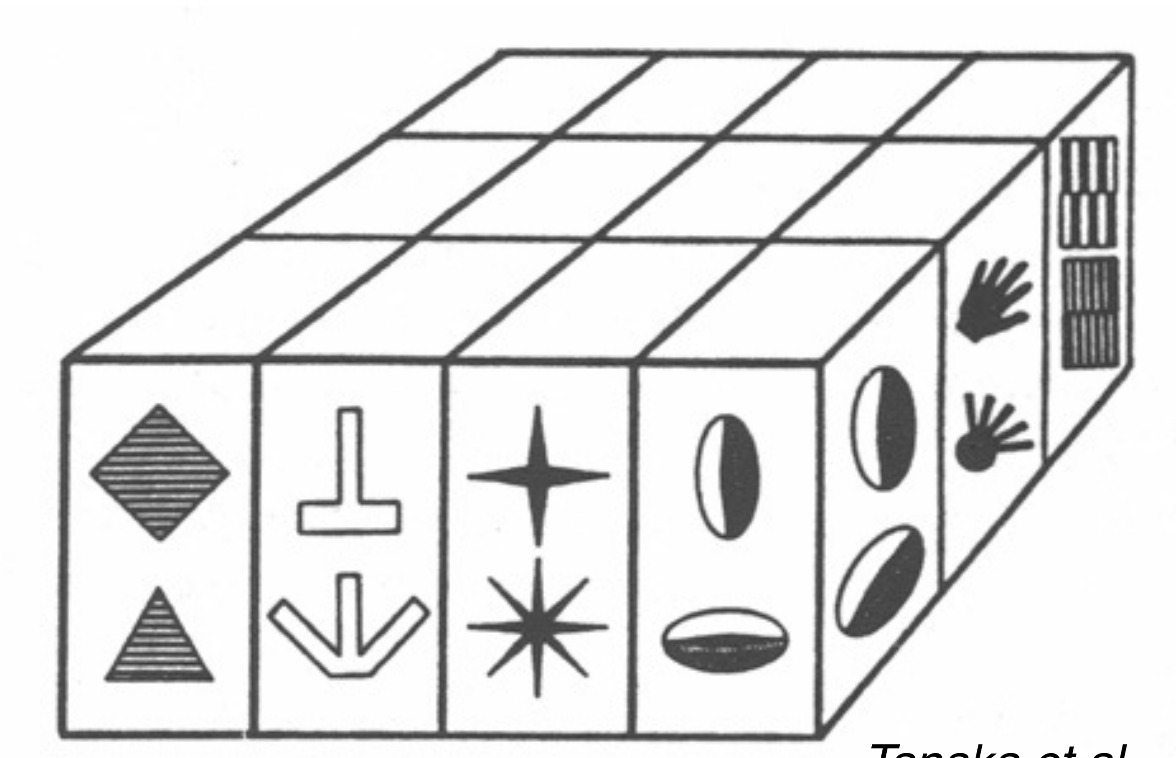
What stimulus feature are IT neurons actually "tuned" to?



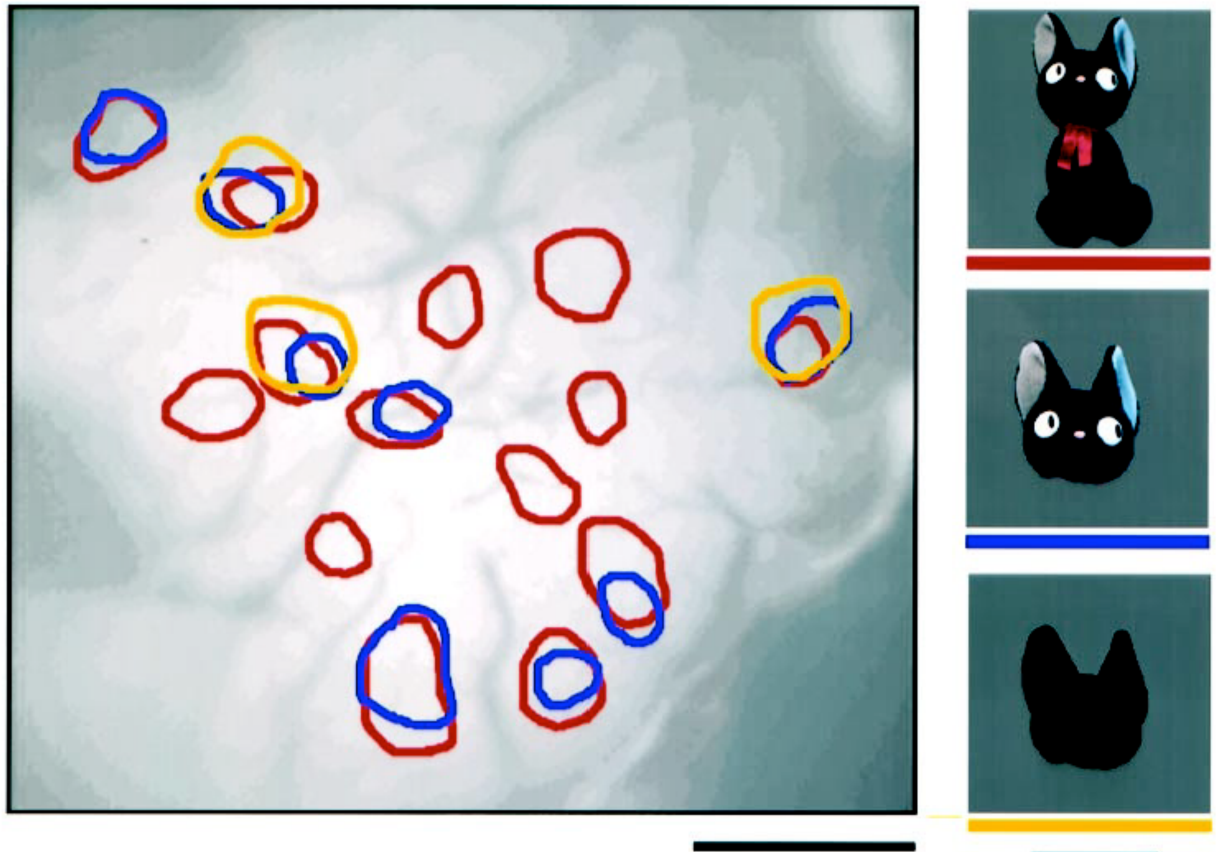


Tanaka et al.

IT has spatial organization at 500 um - 1 mm scale



Tanaka et al.



1 mm

Tsunoda et al.

Larger scale (2-6 mm) organization for some image contrasts

Winrich Freiwald

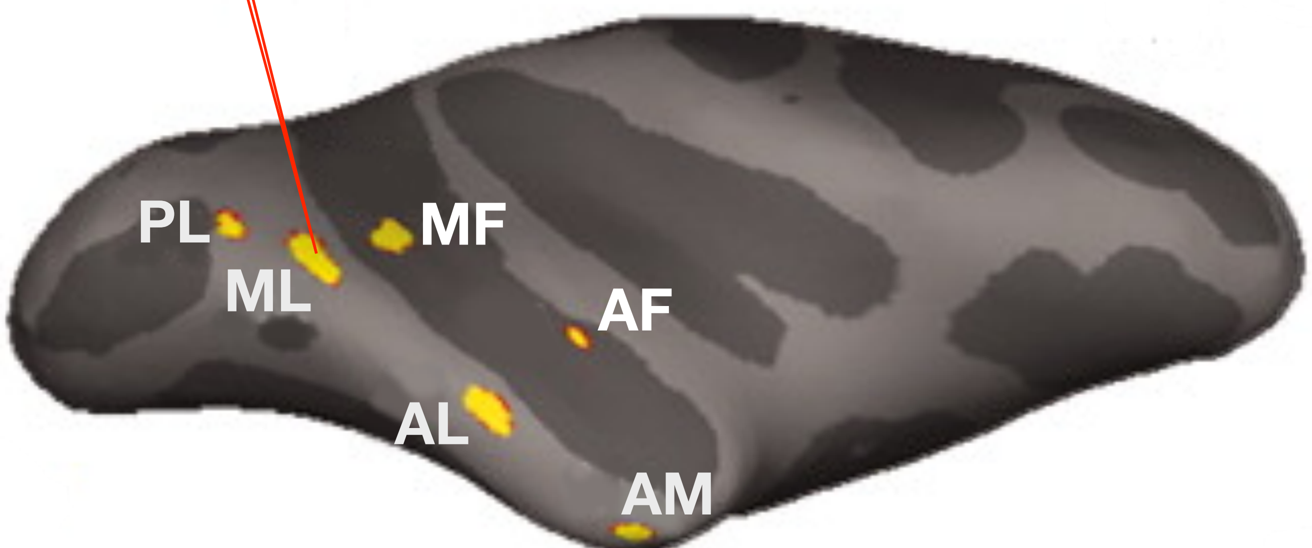
Doris Tsao

Nancy Kanwisher

Face Patches in IT

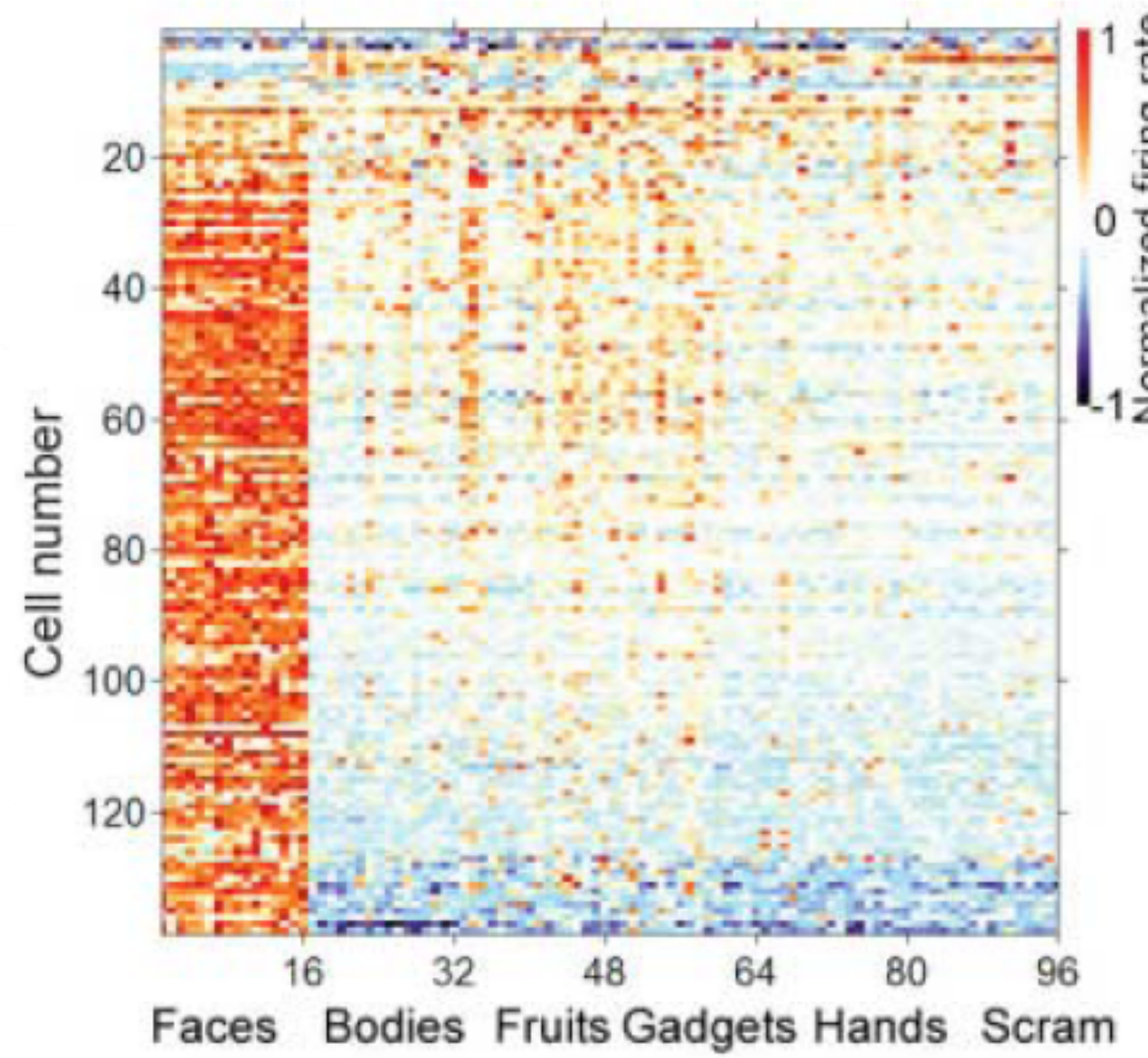
fMRI

Faces vs Objects



-

ML

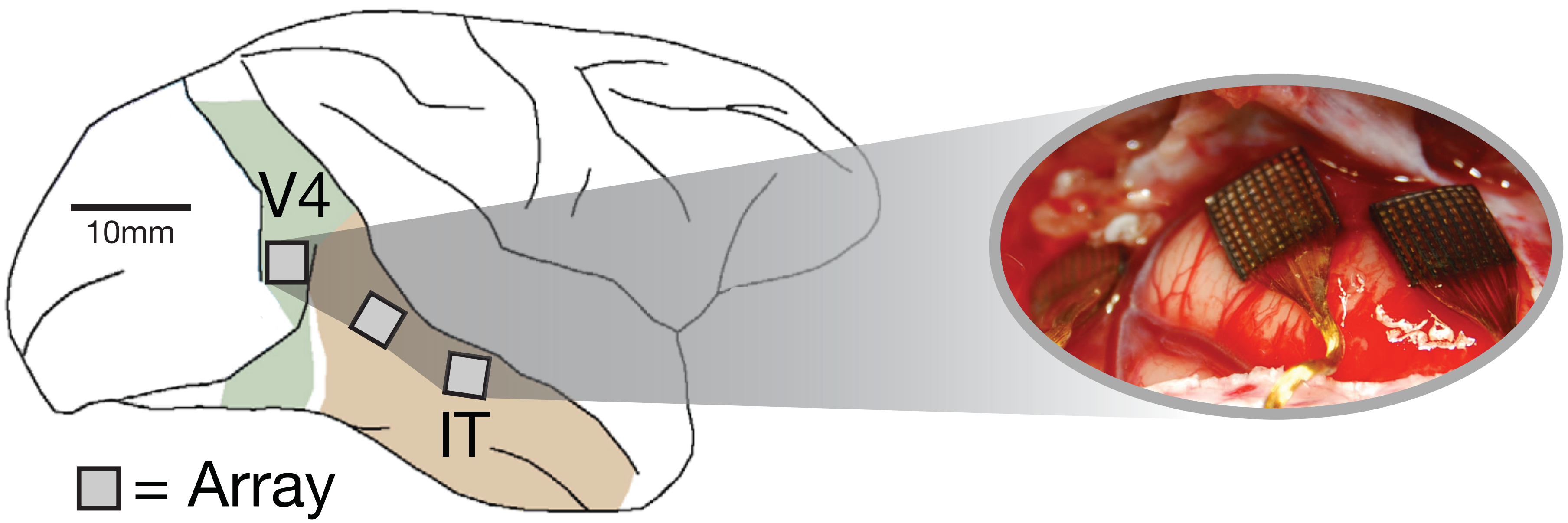


Tsao, Freiwald, and Livingstone used fMRI to reveal a set of face selective regions in macaque IT (aka "face patches")

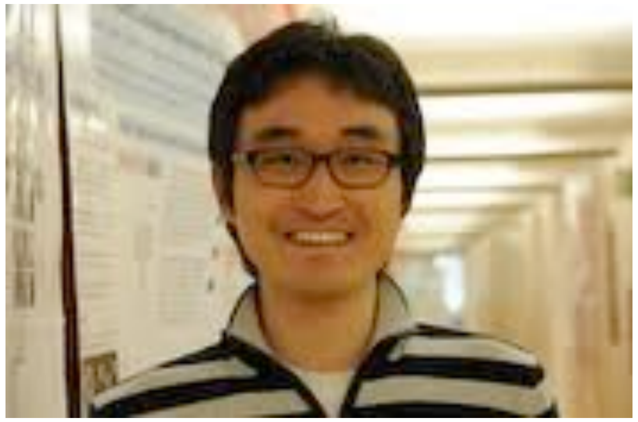
Most of the single neurons in these regions showed a preference for frontal faces

Tsao et al., *Science* 2006

Multi-array electrophysiology in macaque V4 and IT.



About 300 total sites



Ha Hong

Jim DiCarlo

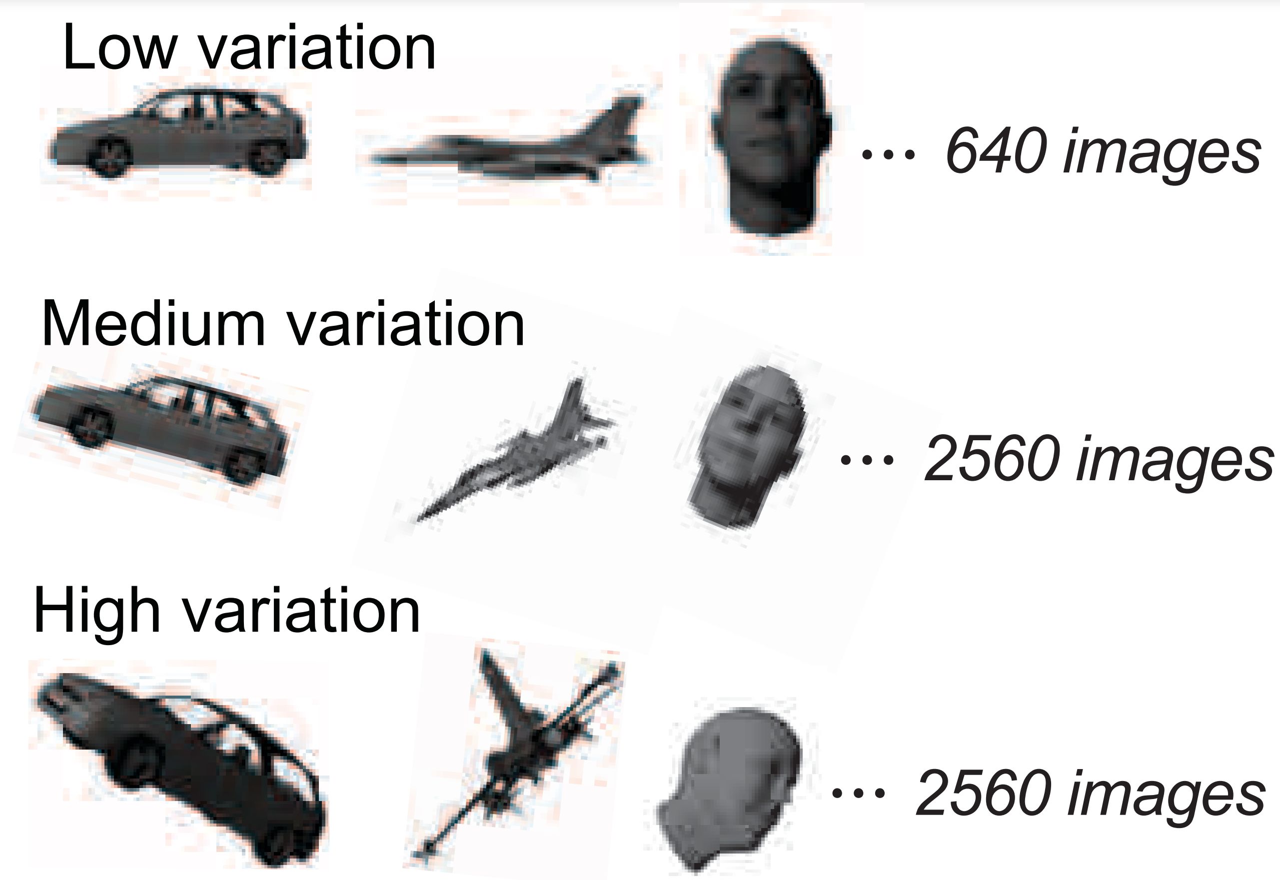
Multi-array Electrophysiology Experiment

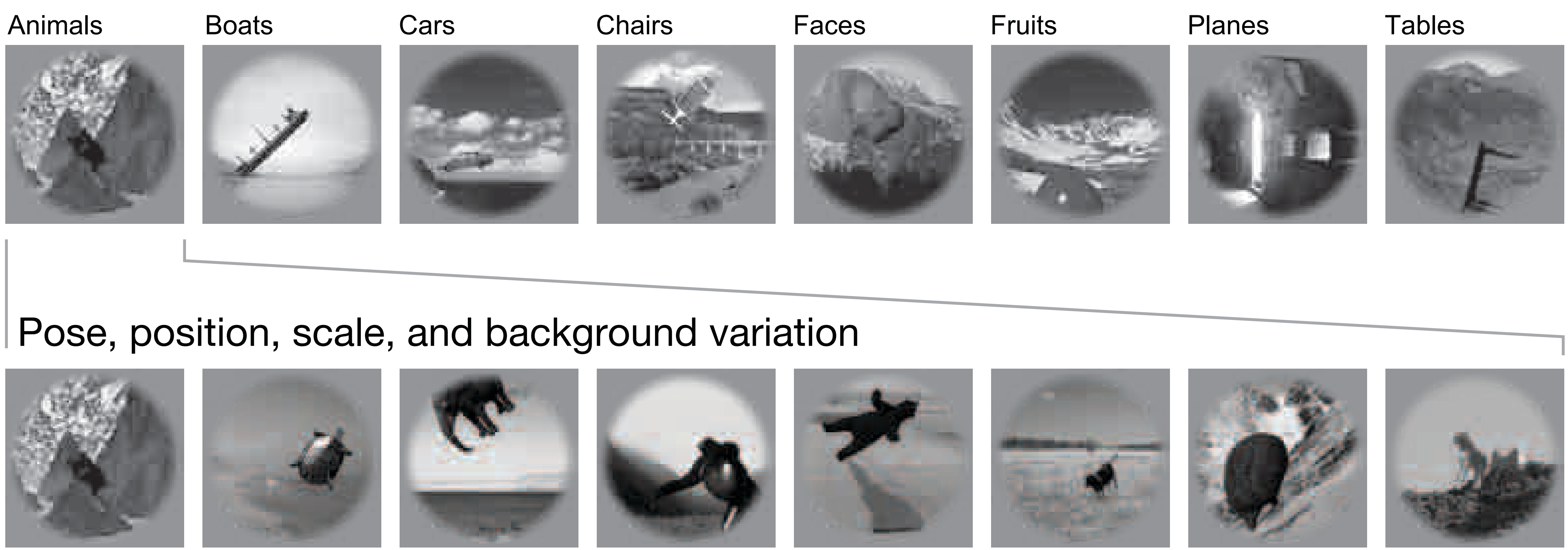
5760 images

64 objects

8 categories

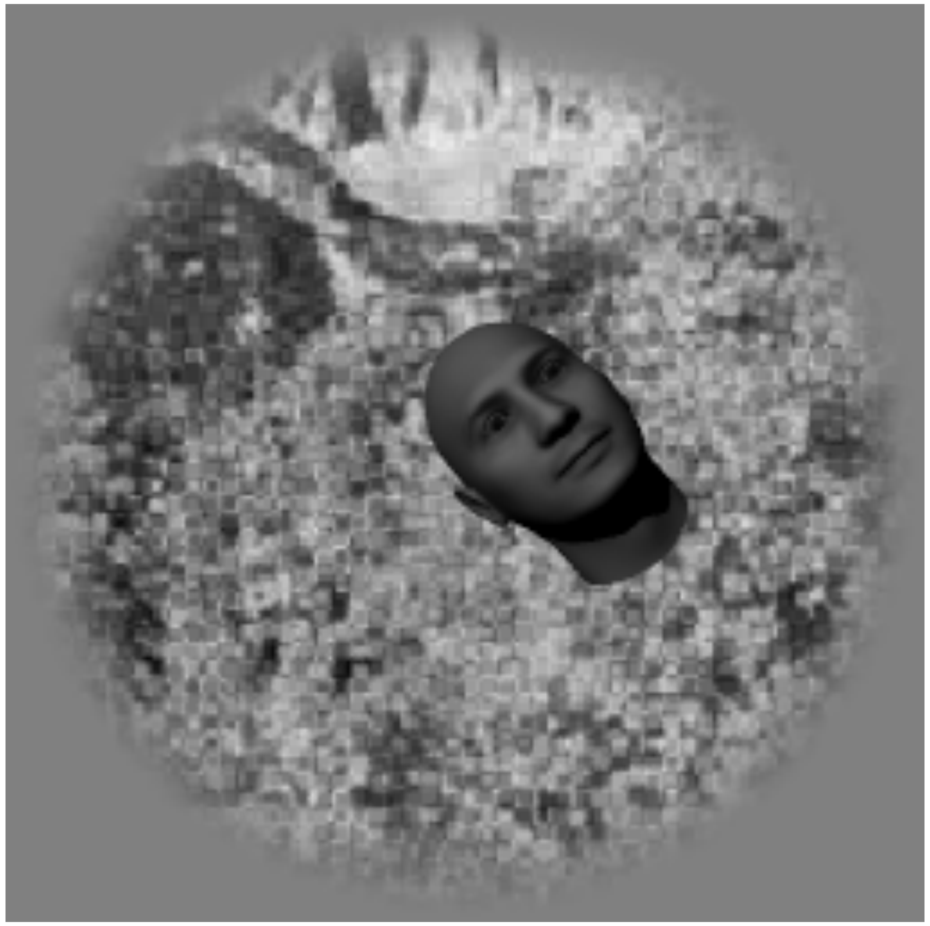
uncorrelated photo backgrounds





Multi-array Electrophysiology Experiment

Multi-array Electrophysiology Experiment



complex, uncorrelated backgrounds **prevent low-level cheating**

part of what we mean by "complex task"

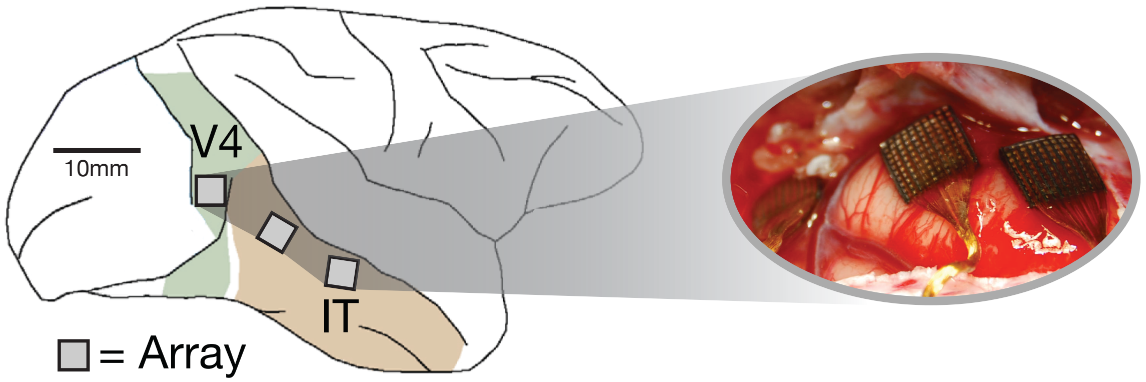
Multi-array Electrophysiology Experiment

Ellie. *C. Shay & K. Kar (Winter 2019)*

complex, uncorrelated backgrounds **prevent low-level cheating**

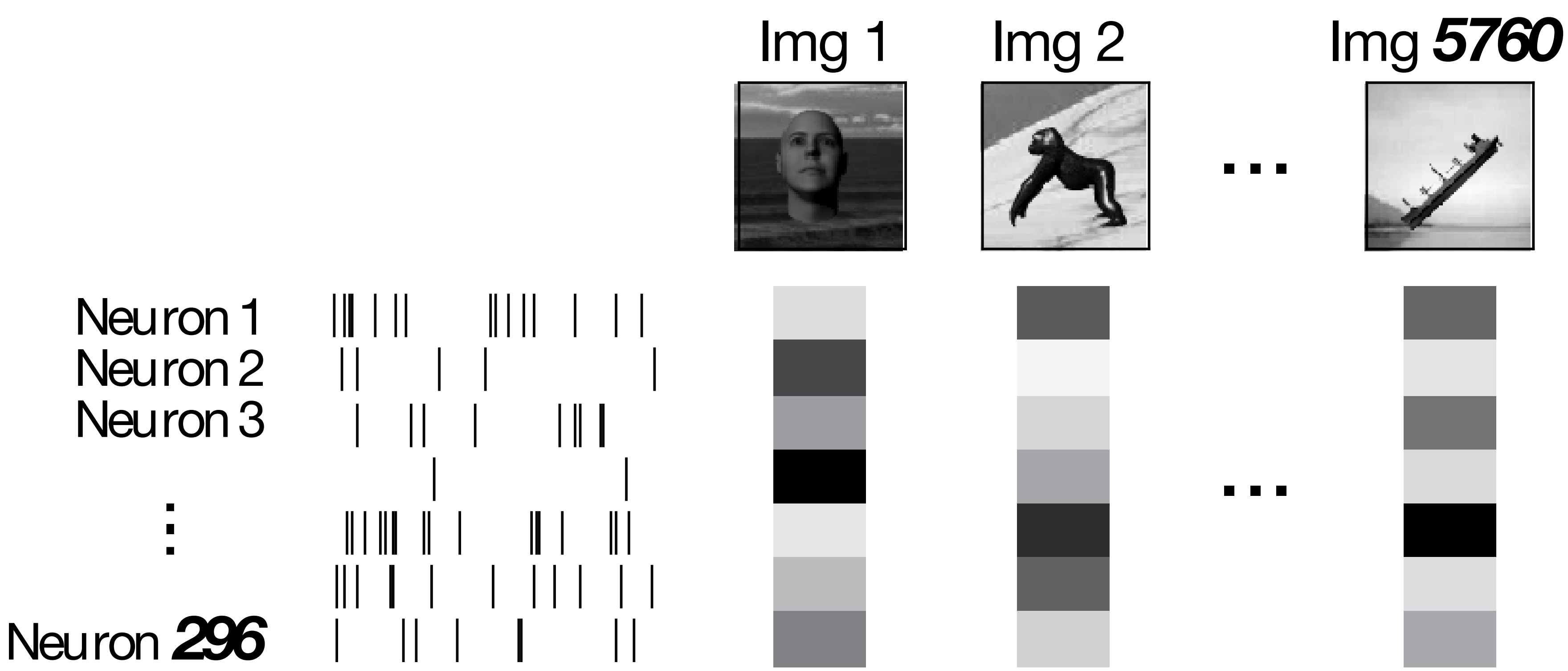
part of what we mean by "complex task"

Multi-array Electrophysiology Experiment



About 300 total sites

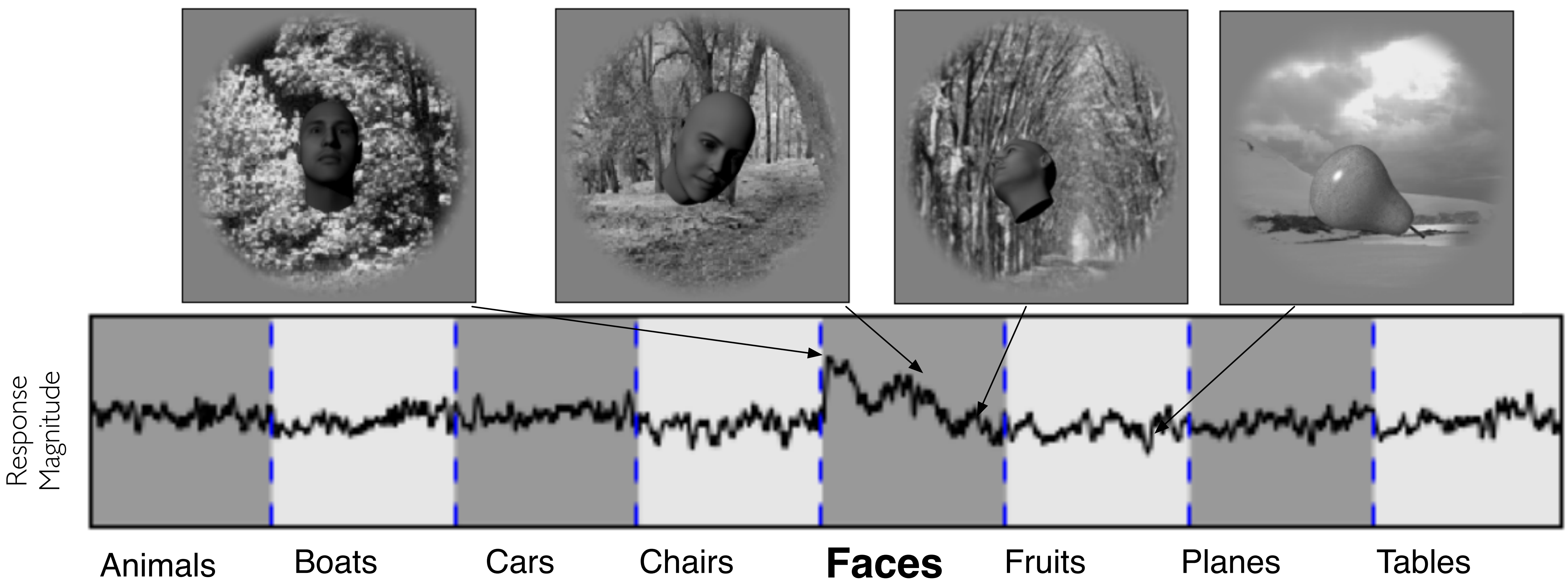
Output = Binned spike counts 70ms-170ms post stimulus presentation averaged over 25-50 reps of each image.



Multi-array Electrophysiology Experiment

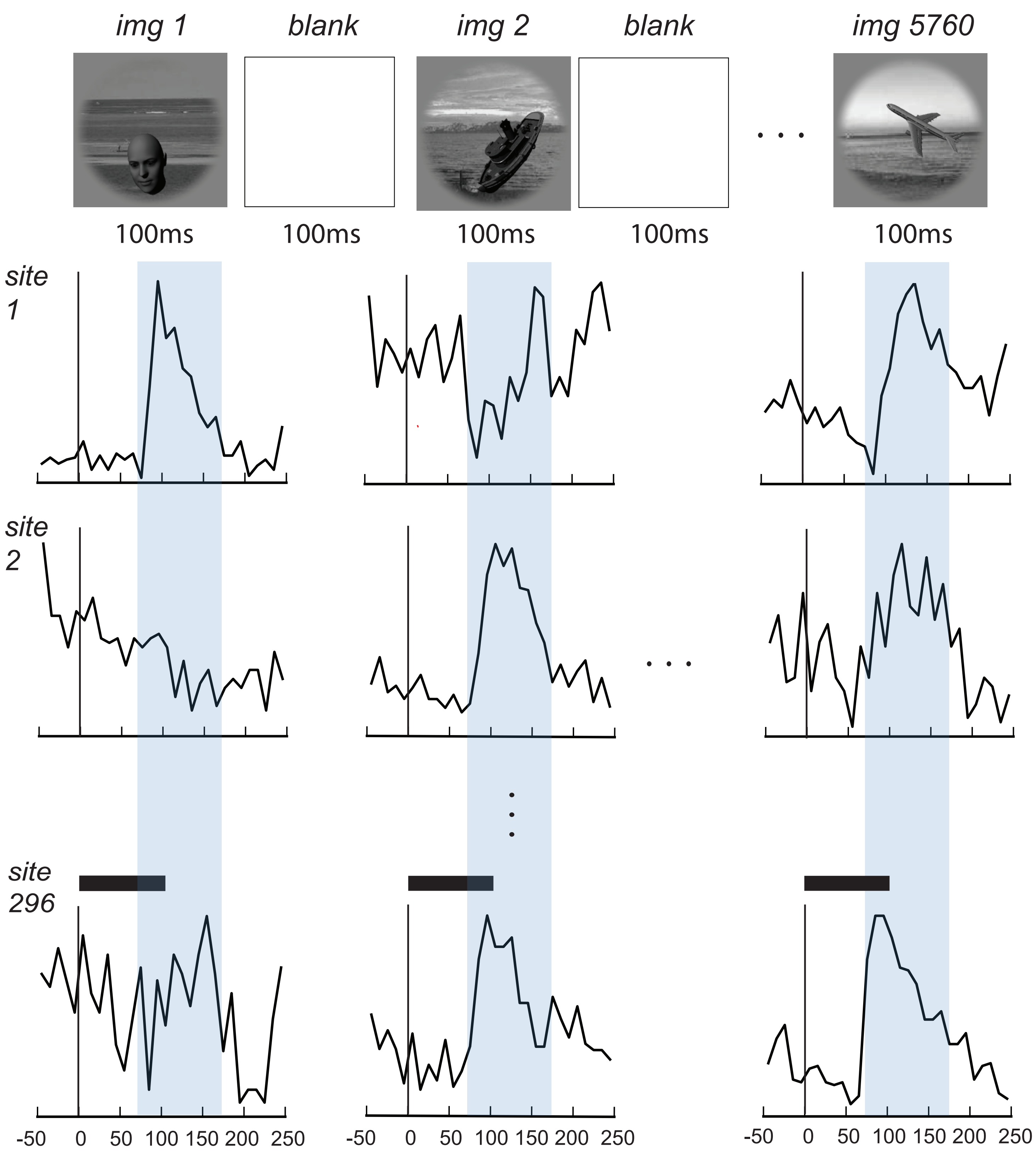
Responses to 1600 test images of two example units

IT unit 53

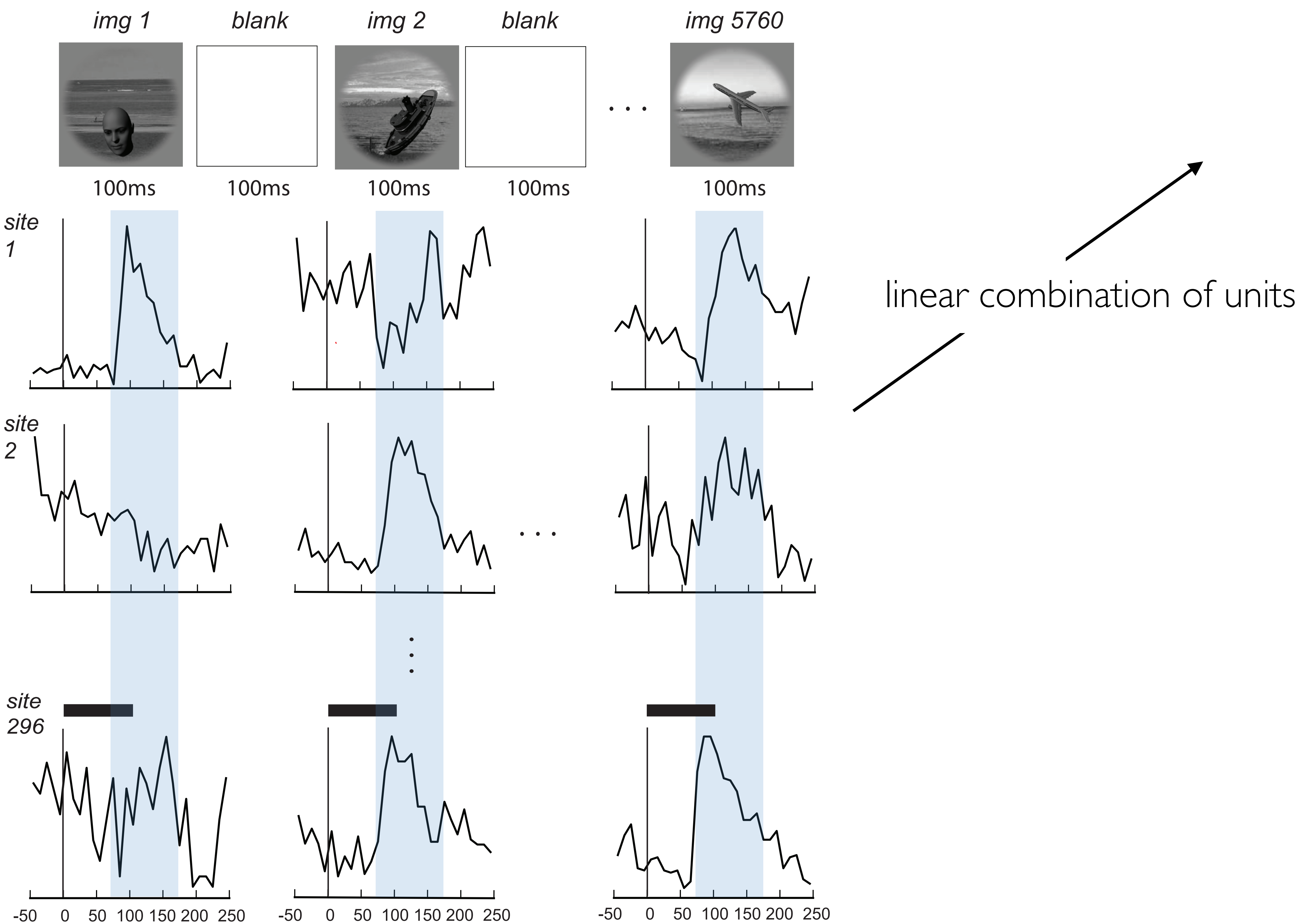


Images sorted first by **category**, then **variation level**.

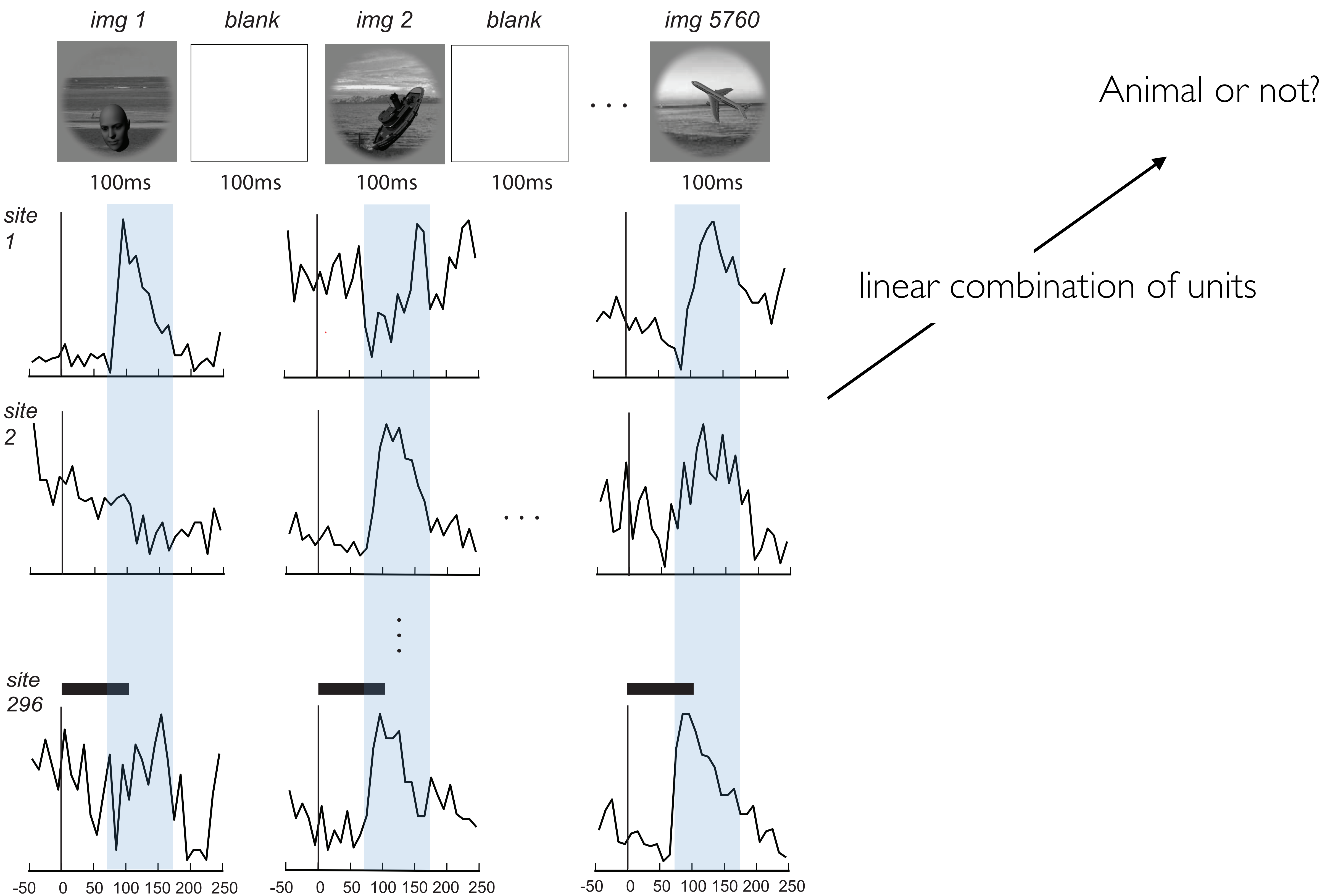
Neural-Behavior Decoding



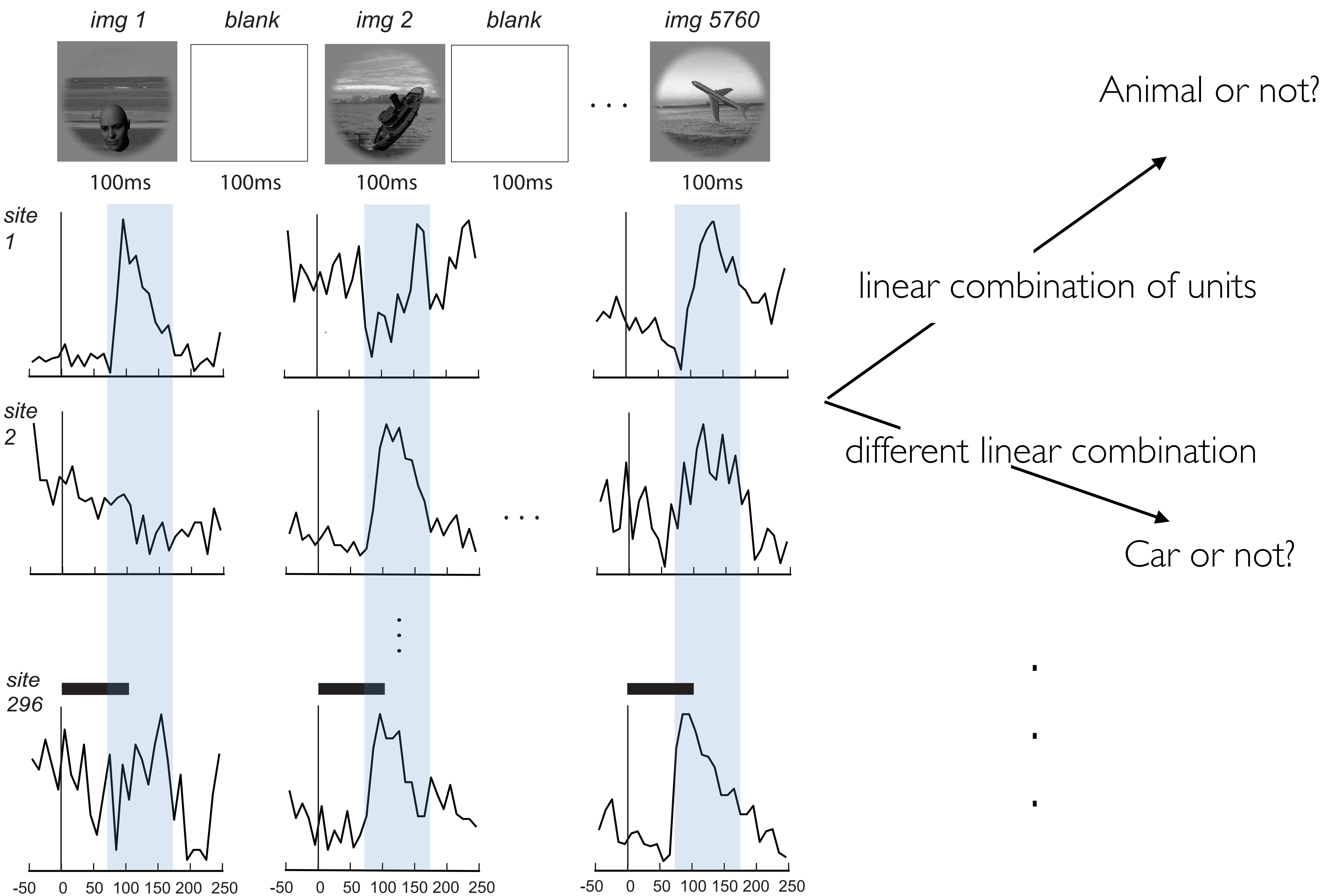
Neural-Behavior Decoding



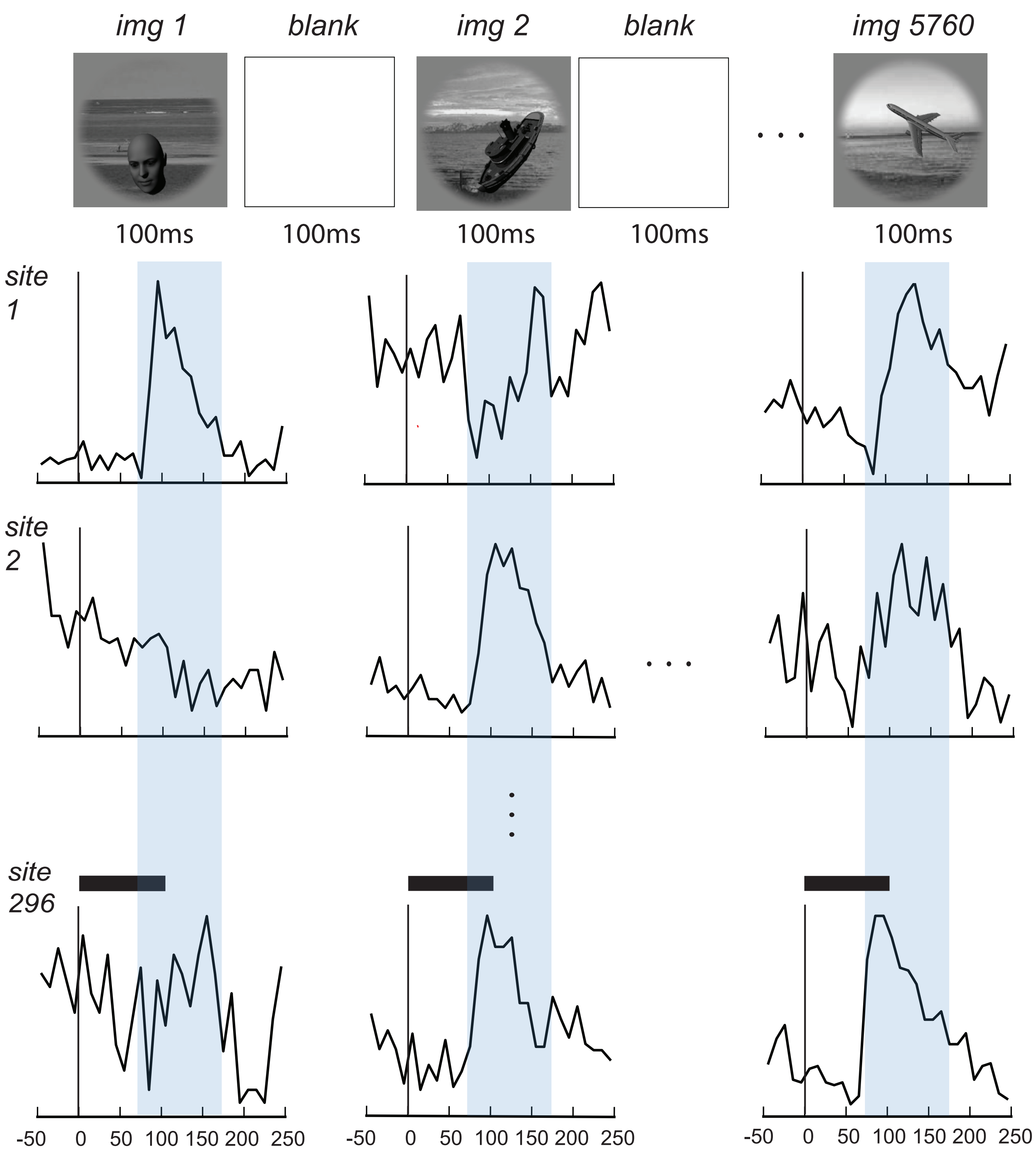
Neural-Behavior Decoding



Neural-Behavior Decoding



Neural-Behavior Decoding



linear combination of units

different linear combination

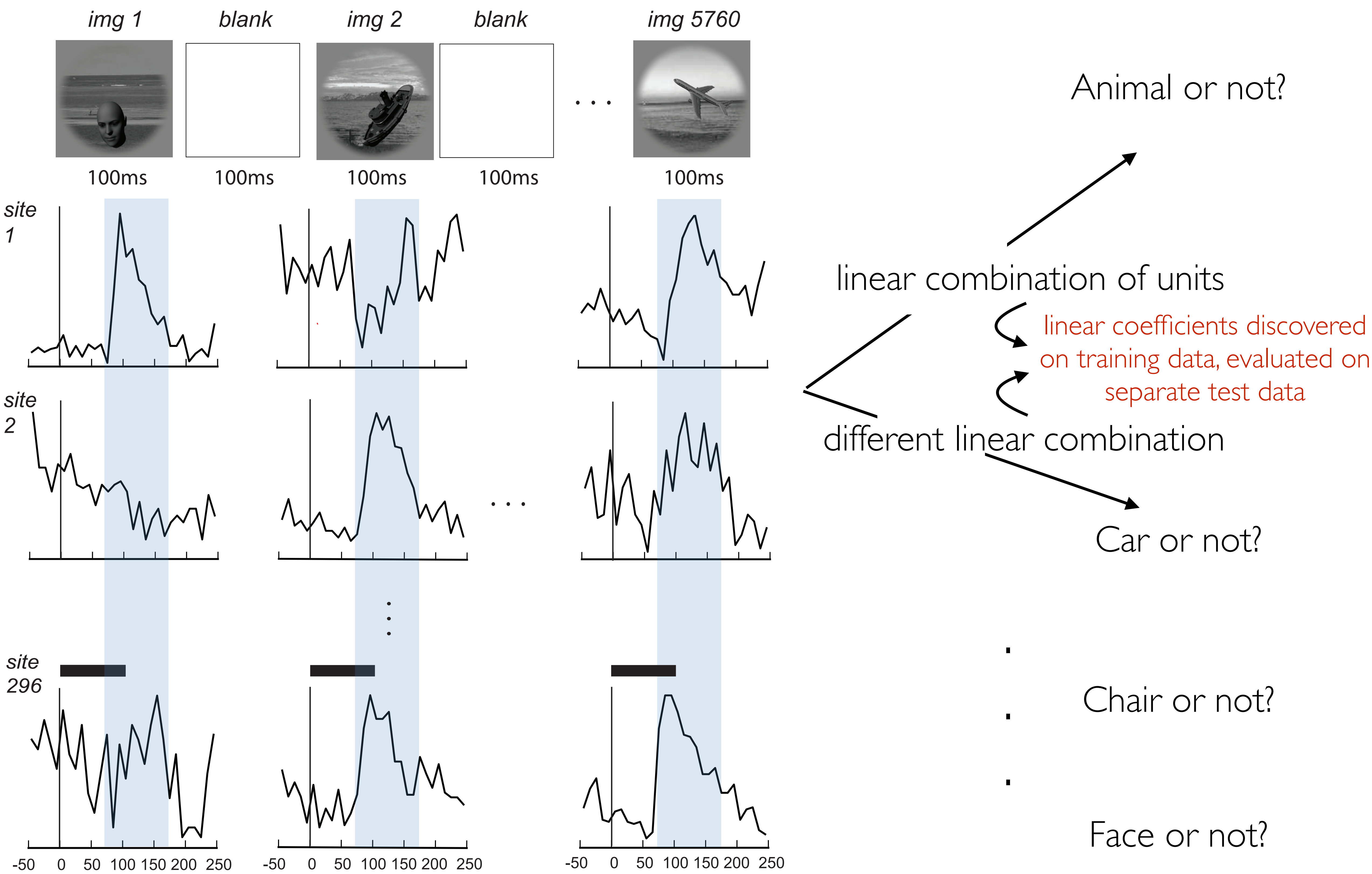
Animal or not?

Car or not?

Chair or not?

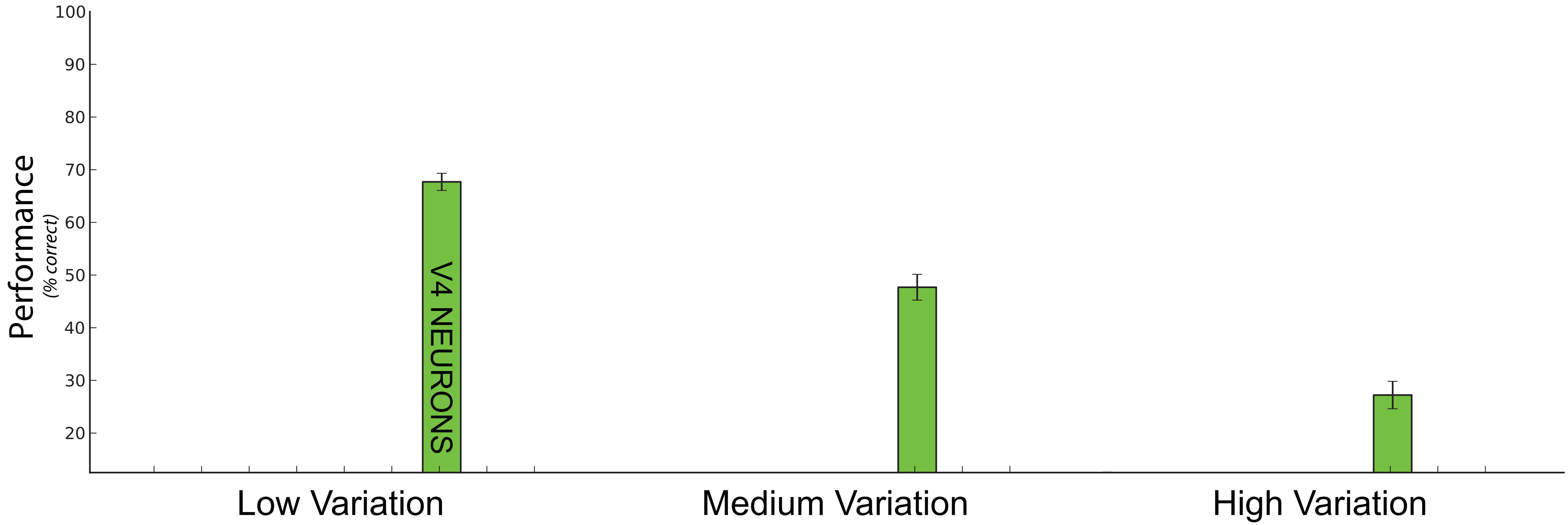
Face or not?

Neural-Behavior Decoding



Decoding Behaviorally Output from Neural Populations

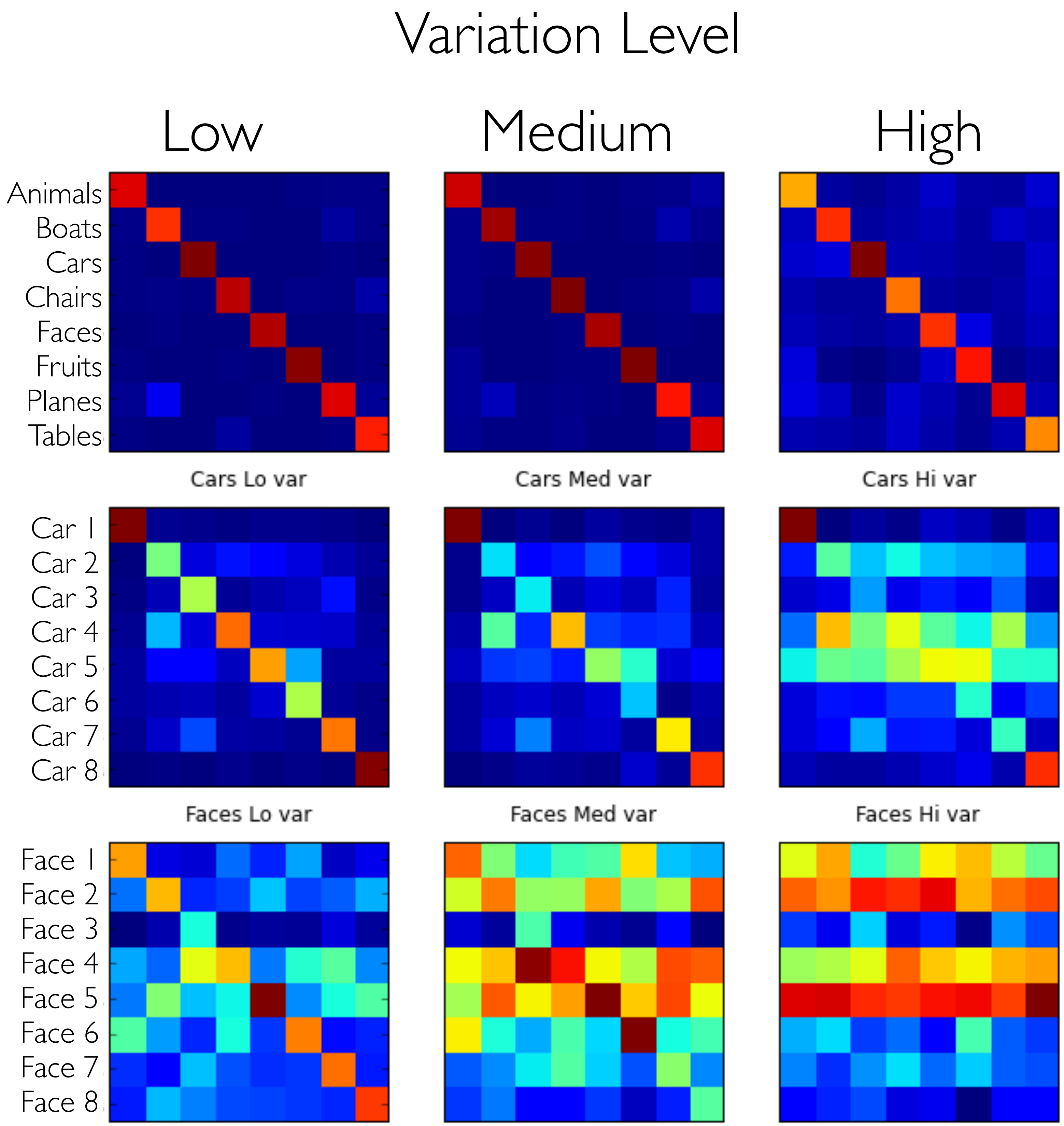
V4 loses out at higher variation:



Yamins and Hong* et. al.* ***PNAS*** *(2014)*

Range of Human Behavior

at
ceiling …

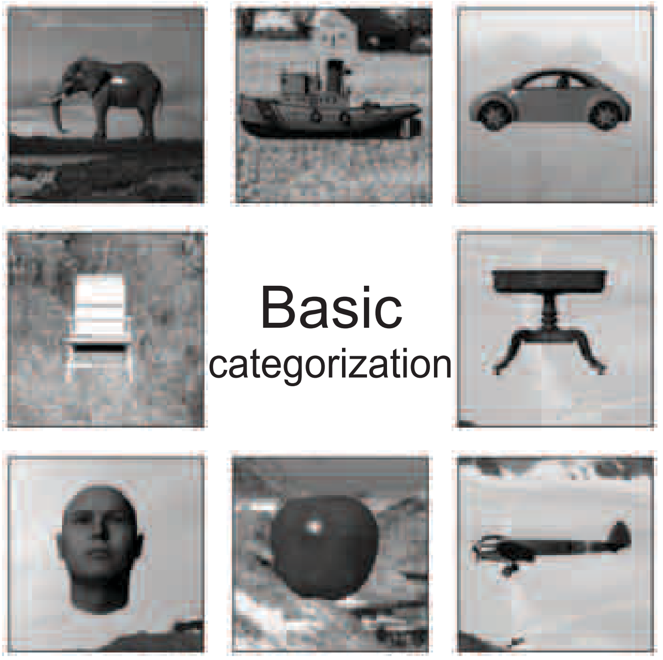


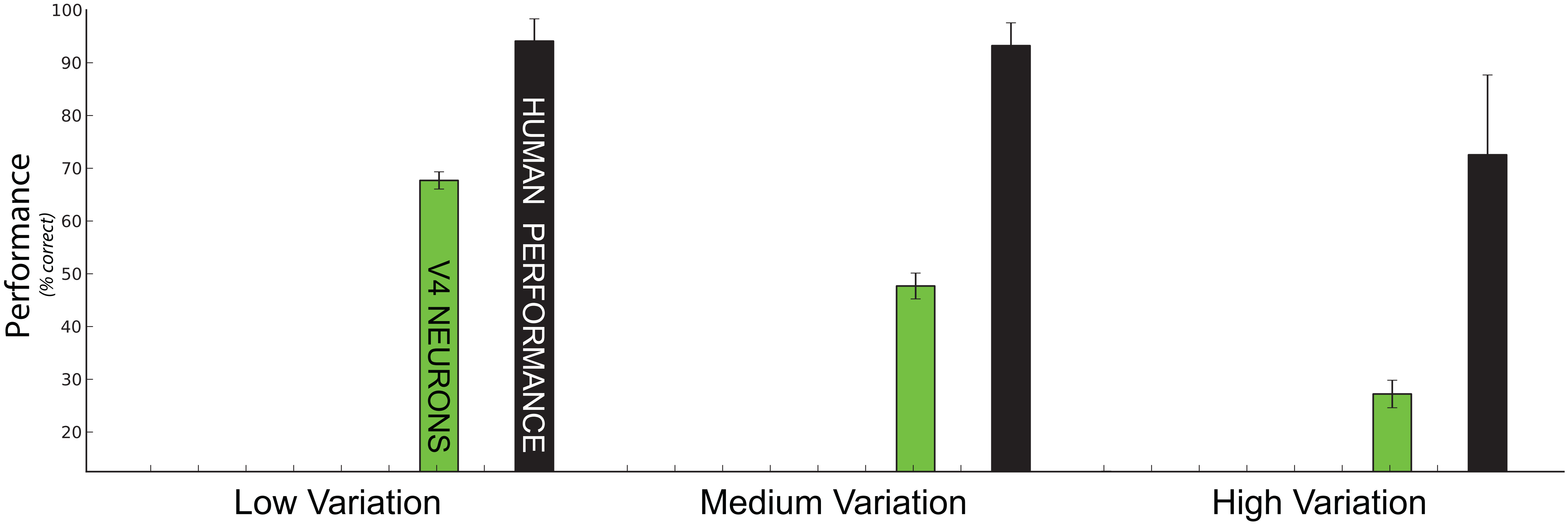
… at
chance

Decoding Behaviorally Output from Neural Populations

V4 loses out at higher variation:

… but humans are much less affected.





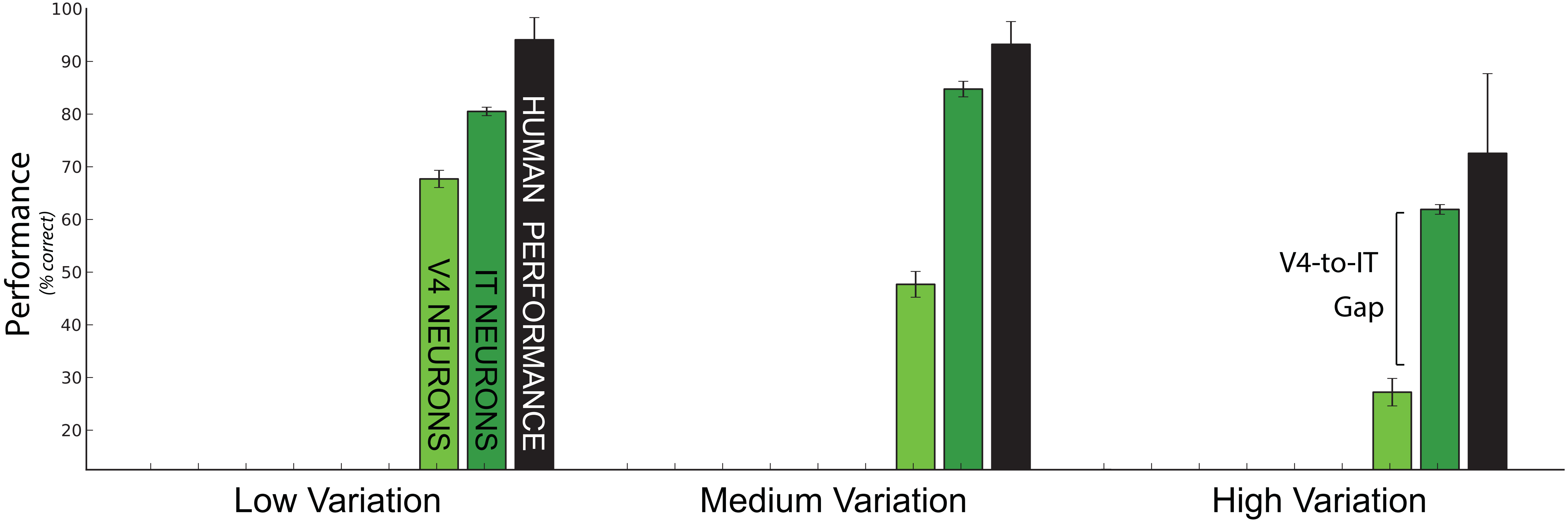
Yamins and Hong* et. al.* ***PNAS*** *(2014)*

IT Neurons Track Human Performance

V4 loses out at higher variation:

… but humans are much less affected.

… as is the IT neural population.



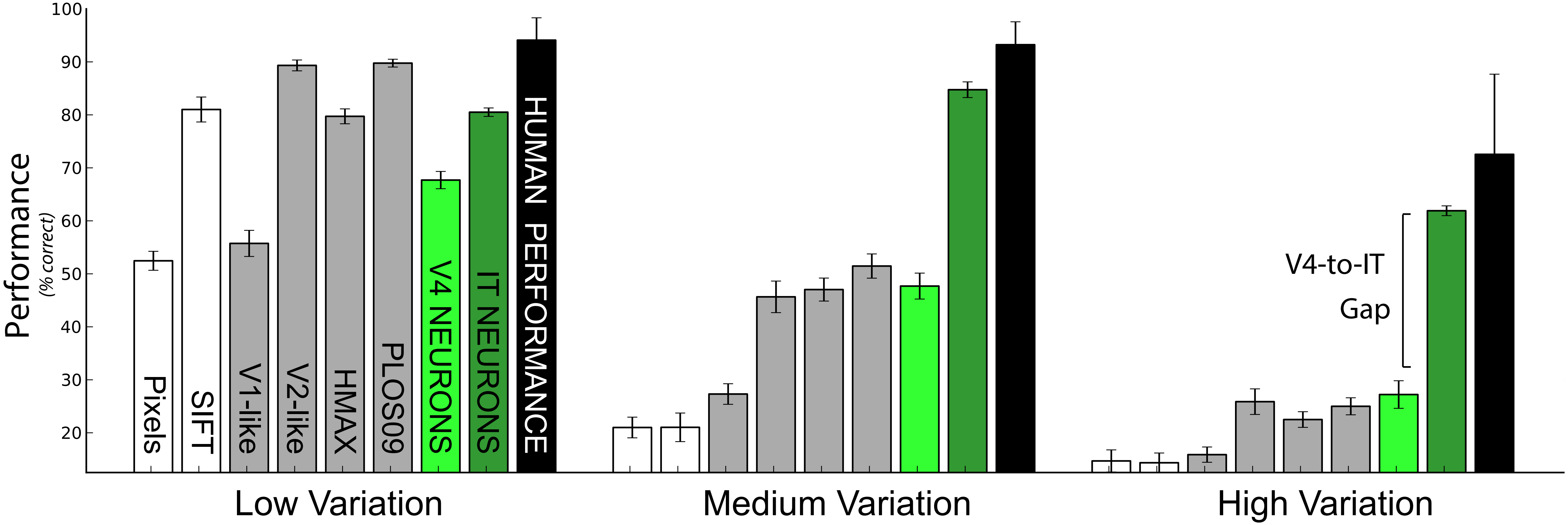
Yamins and Hong* et. al.* ***PNAS*** *(2014)*

IT Neurons Track Human Performance

V4 loses out at higher variation:

… but humans are much less affected.

… as is the IT neural population.

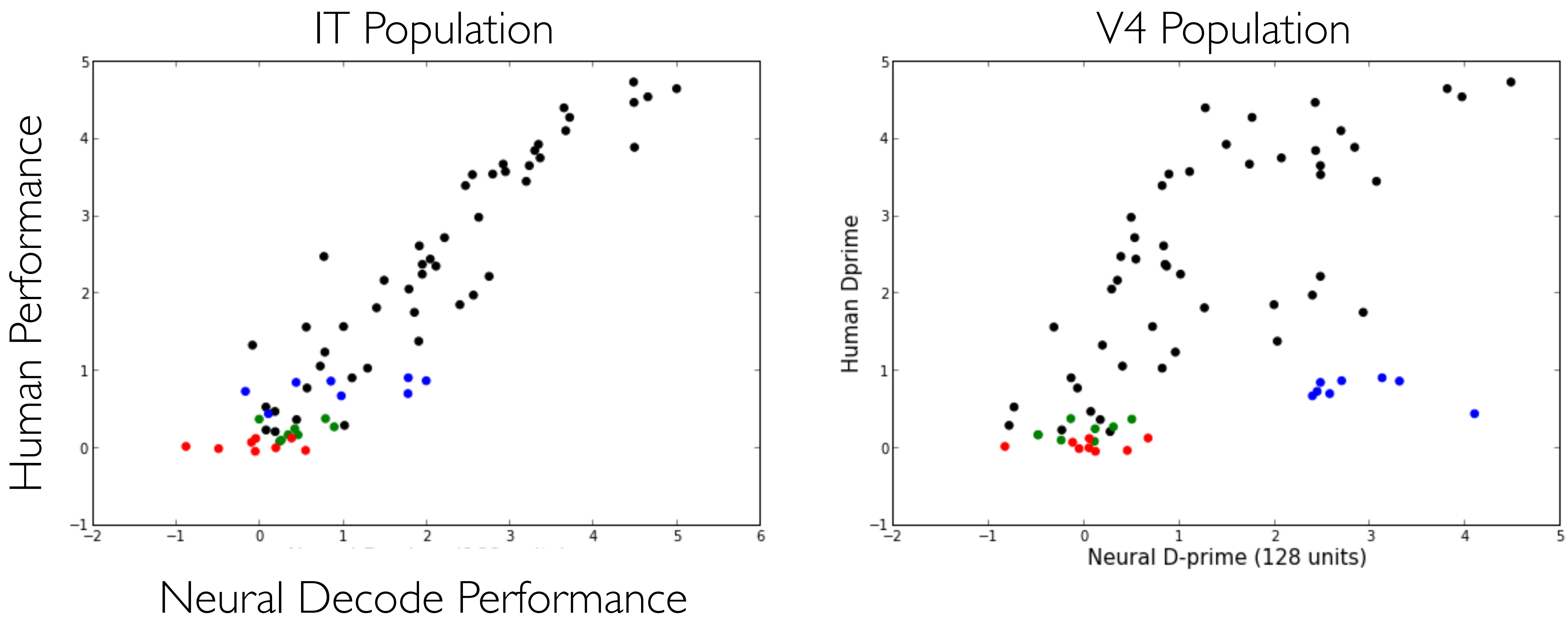


Yamins and Hong* et. al.* ***PNAS*** *(2014)*

At <u>high variation levels</u>, IT much better than V4 and existing models.

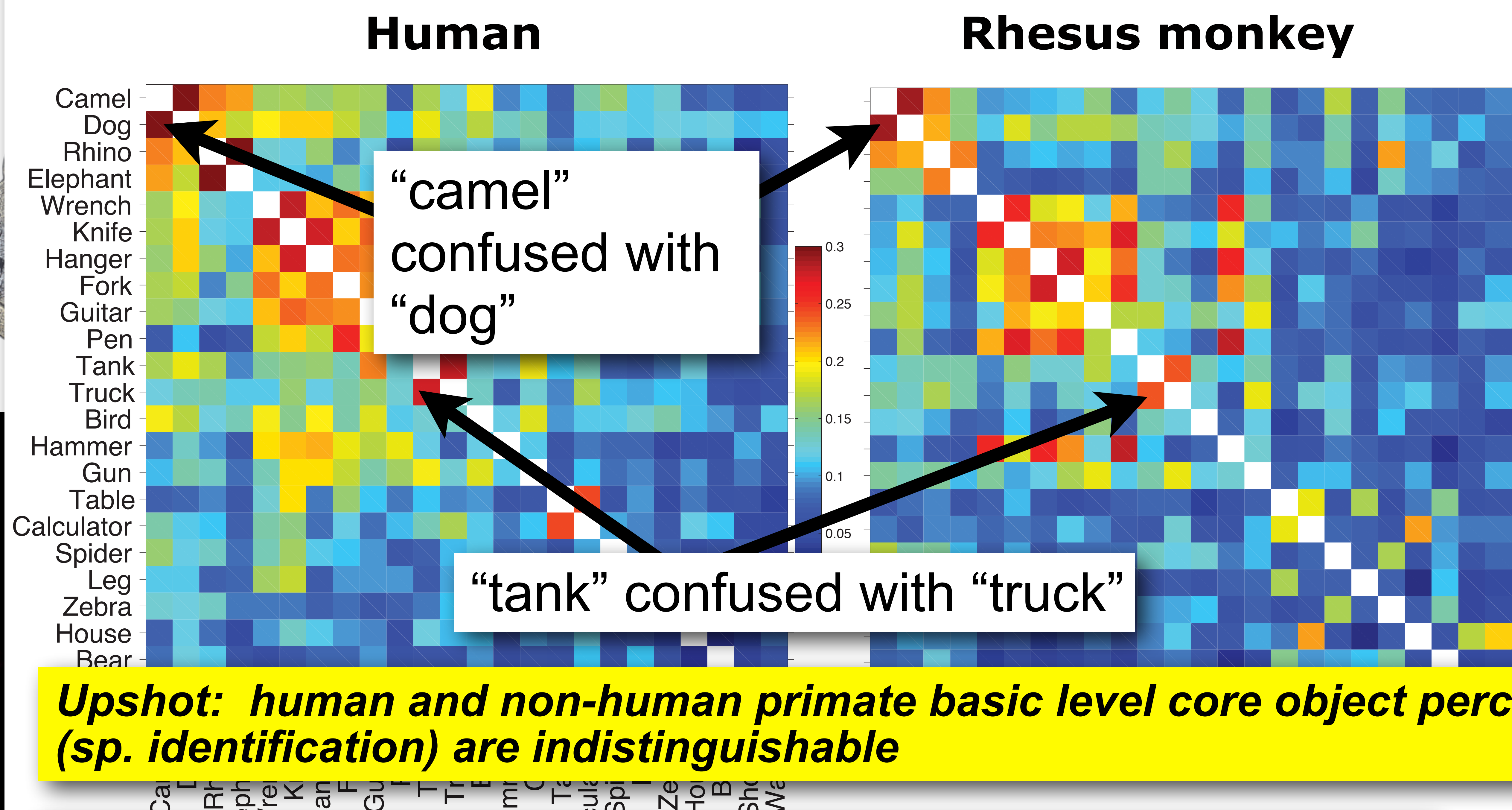
IT Neurons Track Human Performance

IT matches human error patterns as well as raw performance.



- Low-Variation Face subordinate tasks.

Human / Monkey similarities



Upshot: human and non-human primate basic level core object percept (sp. identification) are indistinguishable

Does not depend on reporting effector (touch vs. eye movement)

Comparison of Object Recognition Behavior in Human and Monkey

R. Rajalingham, K Schmidt, J.J. DiCarlo, ***Vision Sciences Society*** *(2014)*

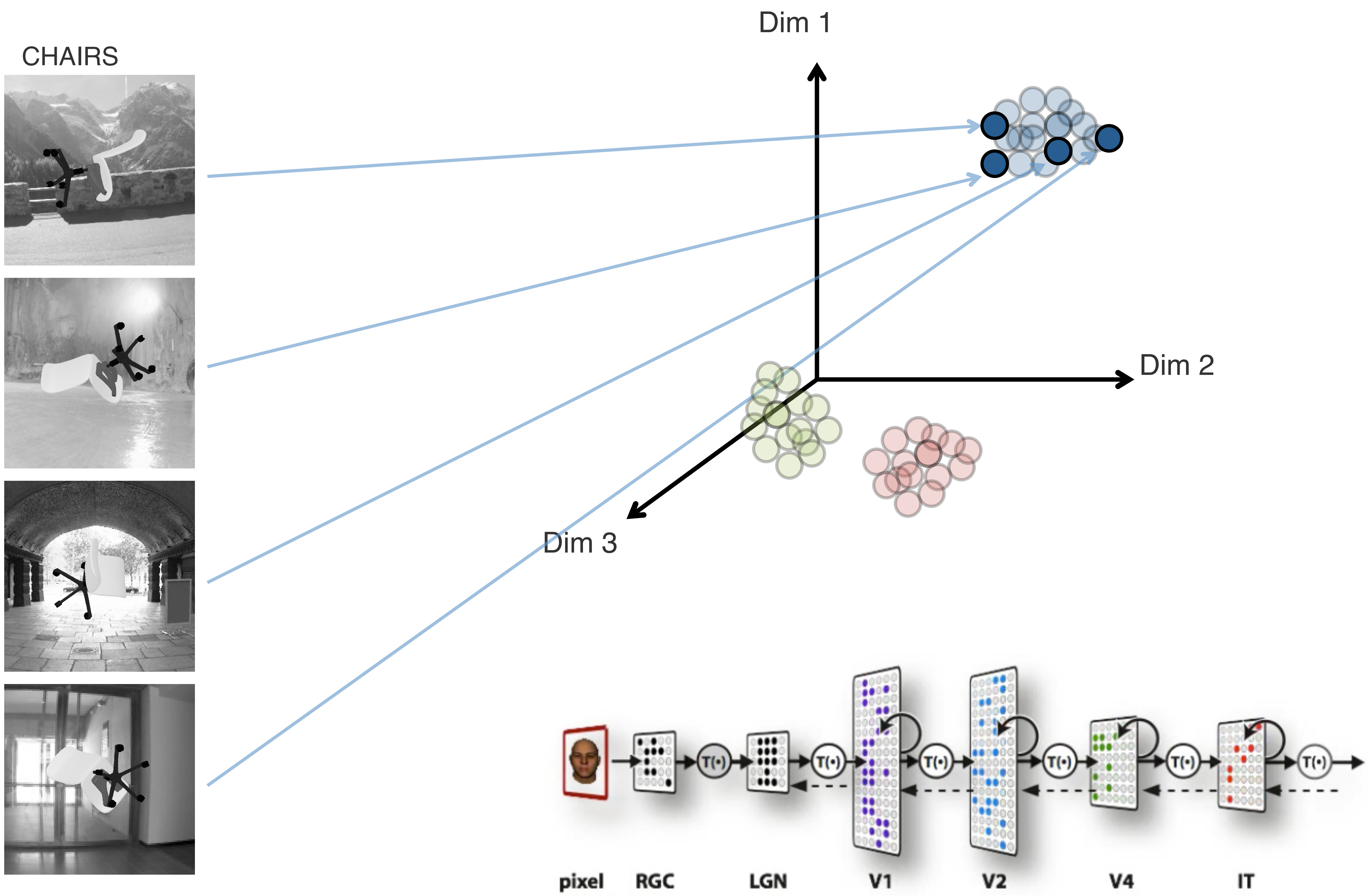
R. Rajalingham, K Schmidt, J.J. DiCarlo, ***J. Neuroscience*** *(2015)*

Adapted from Motter and Mountcastle 1981

Feature Space as Encoding

Pixel space: $R^{\sim 1000000}$

Feature space: $R^{4000(?)}$



Feature Space as Encoding

Behavior = Feature space + Simple decision rule

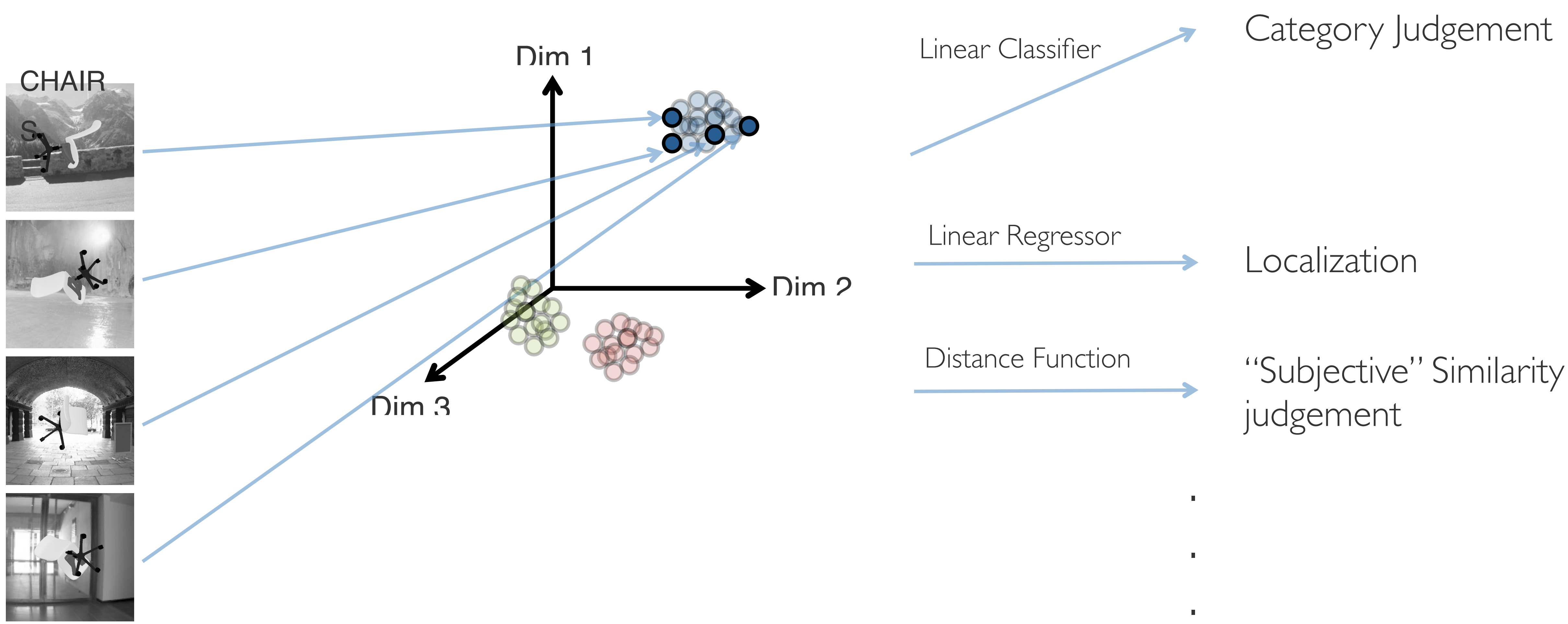
= encoding + decoding

Pixel space: $R^{1000000}$

Output

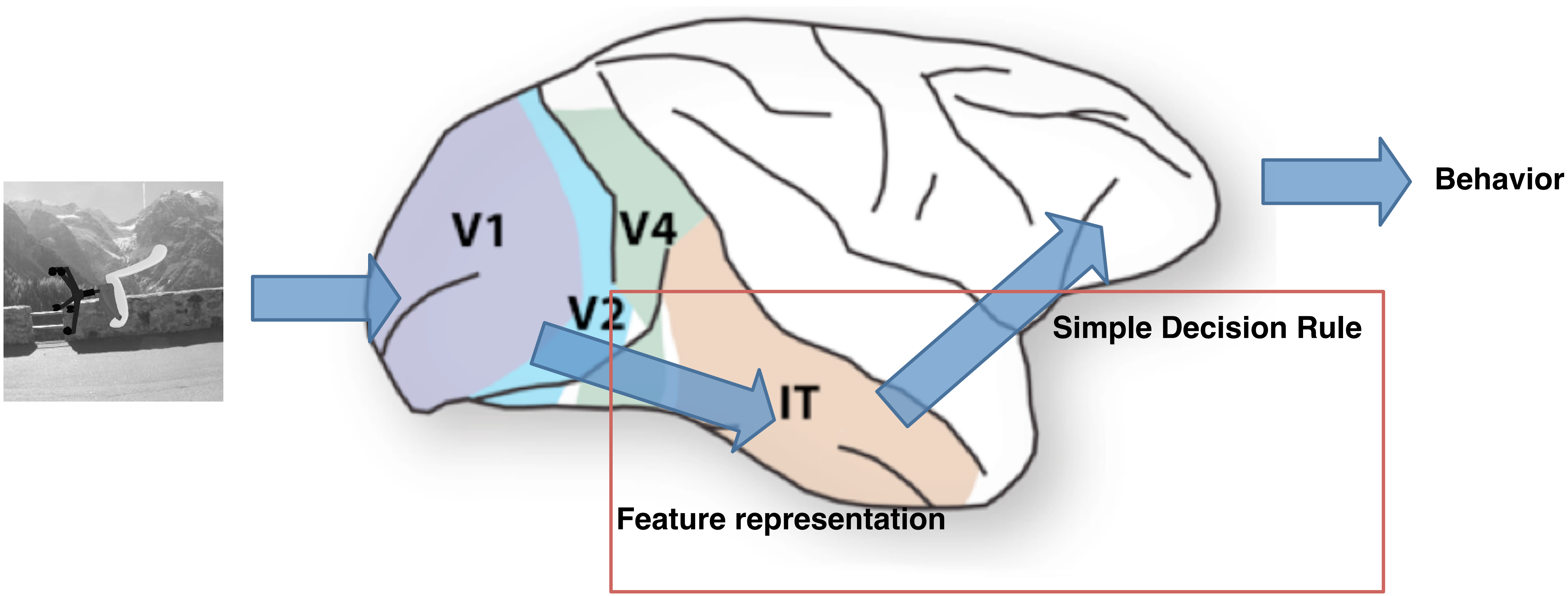
Feature space: $R^{4000(?)}$

Behavioral



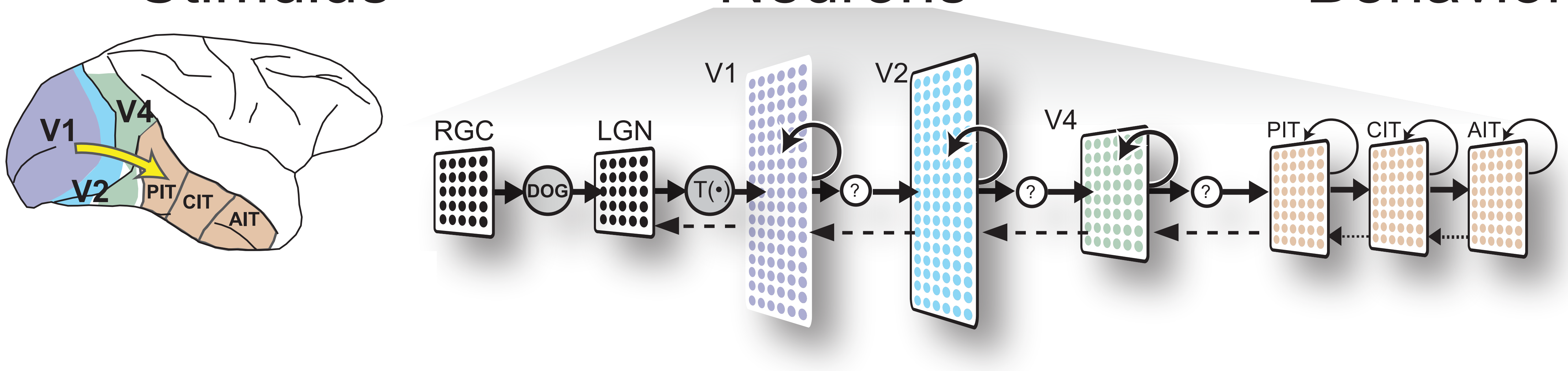
Encoding & Decoding

Behavior = Feature space + Simple decision rule
= encoding + decoding

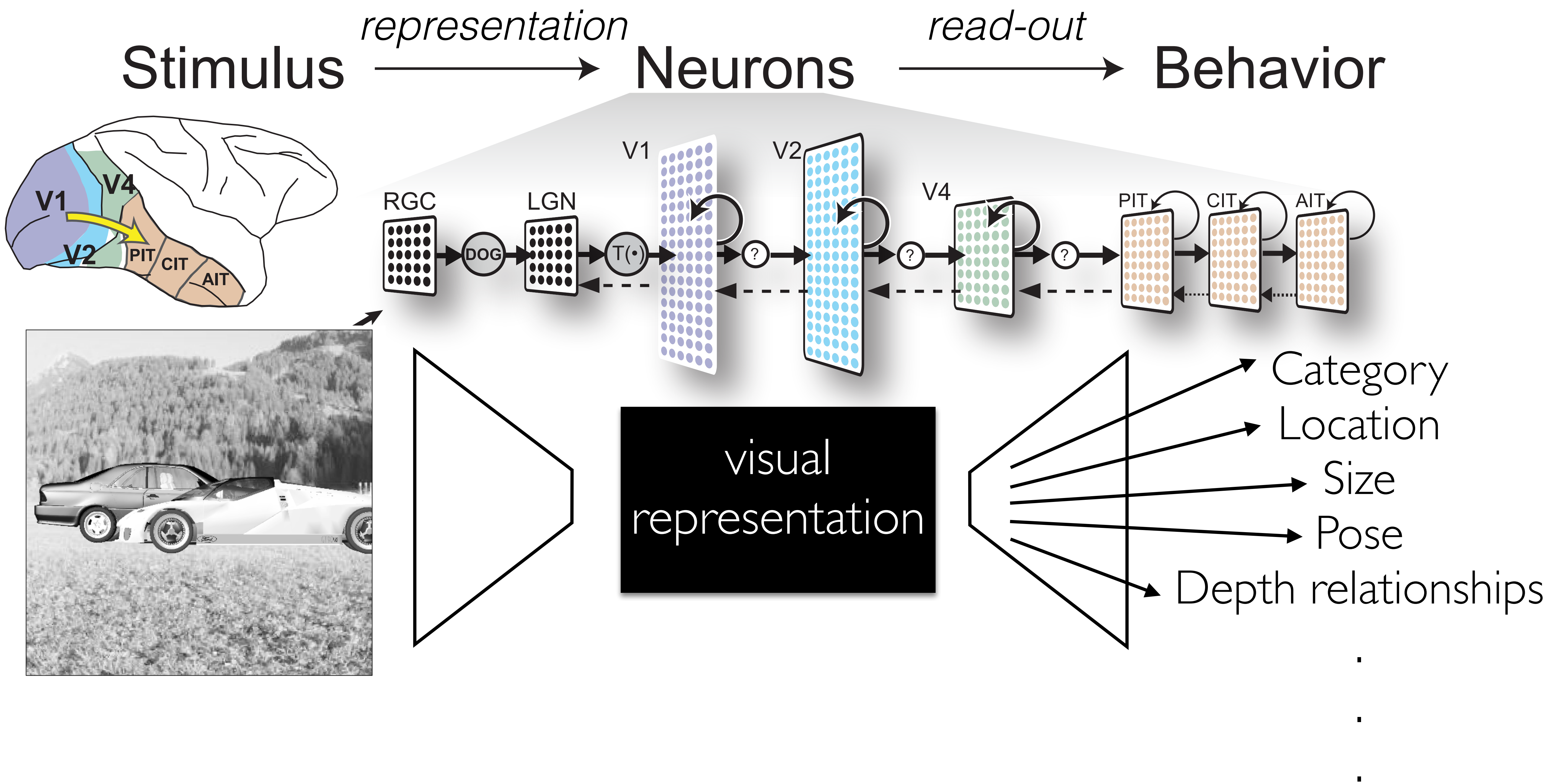


Encoding & Decoding

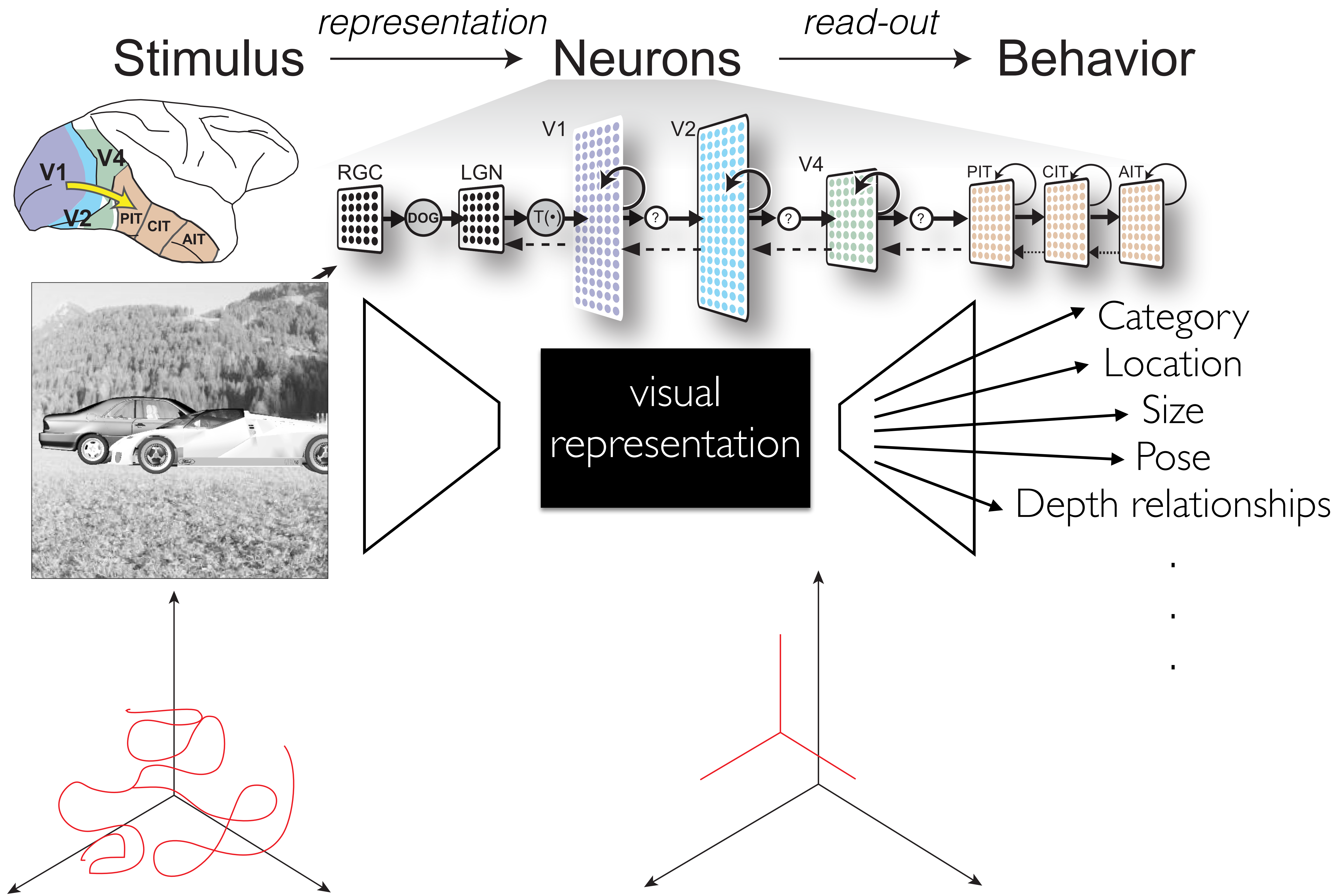
Stimulus $\xrightarrow{\textit{representation}}$ Neurons $\xrightarrow{\textit{read-out}}$ Behavior



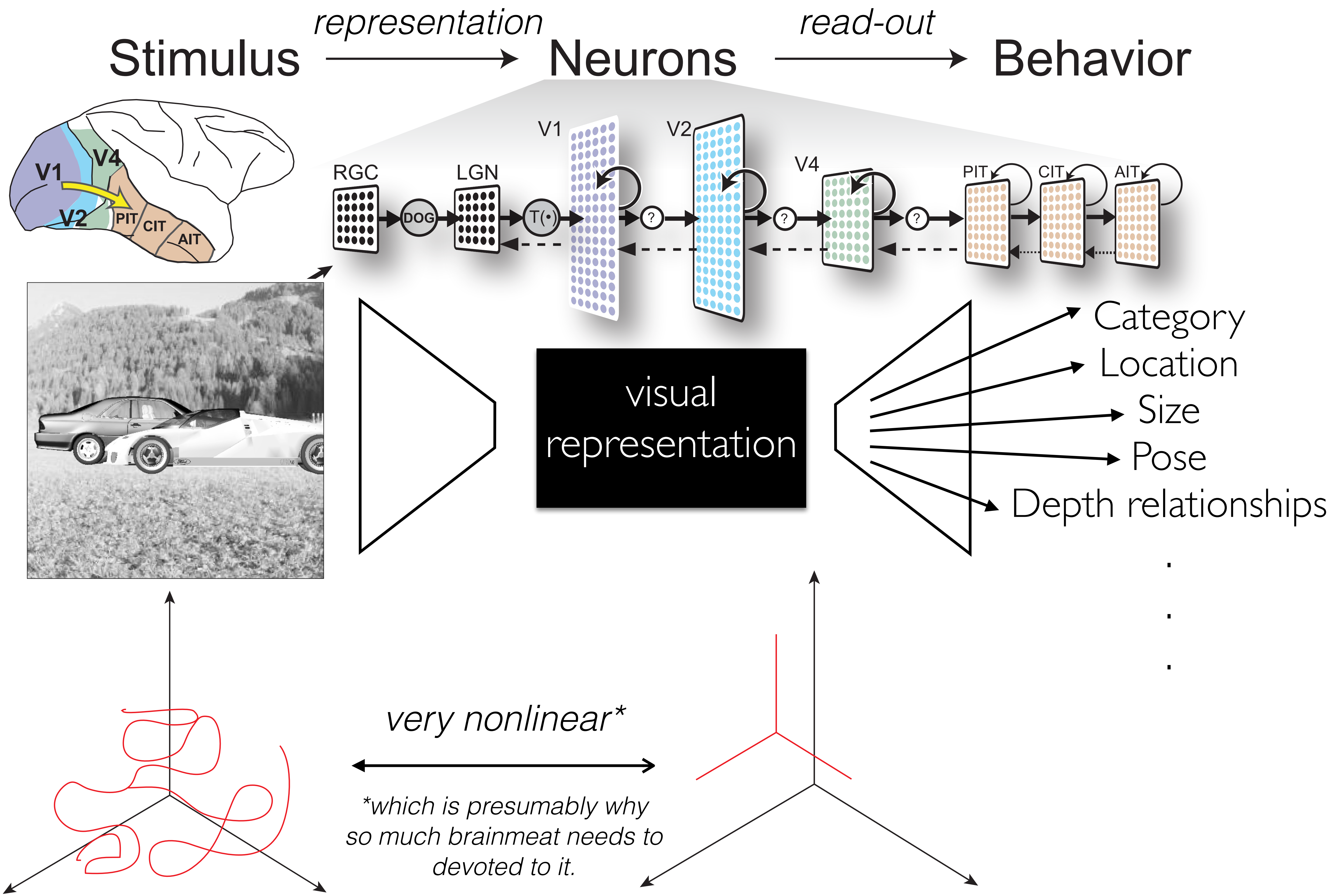
Encoding & Decoding



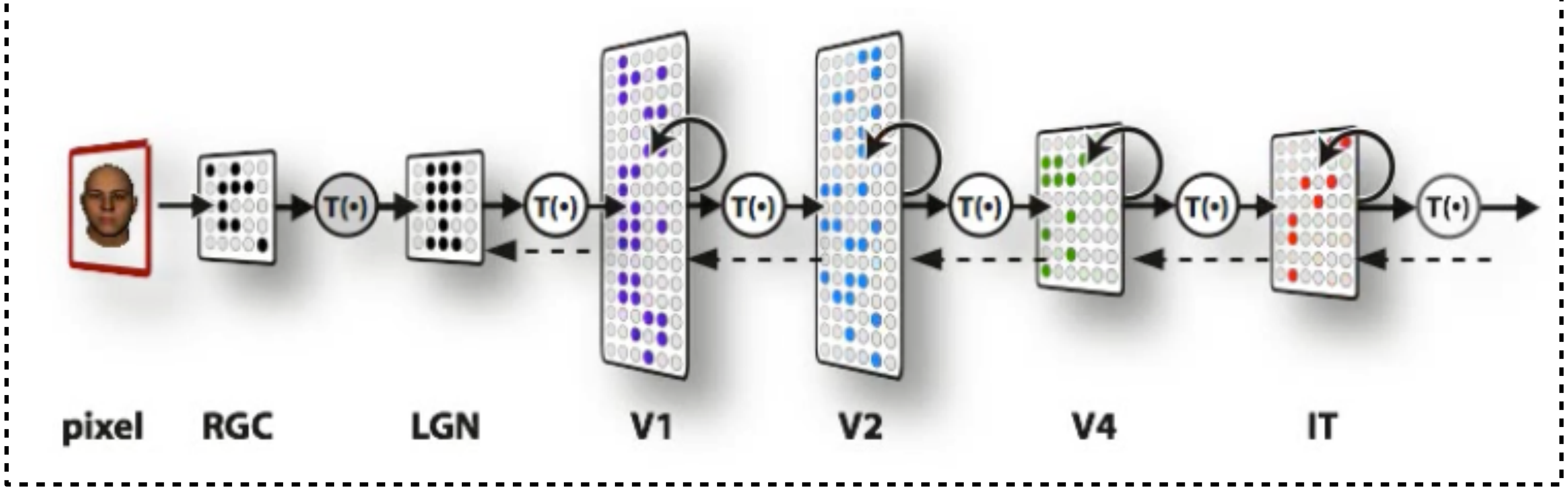
Encoding & Decoding



Encoding & Decoding



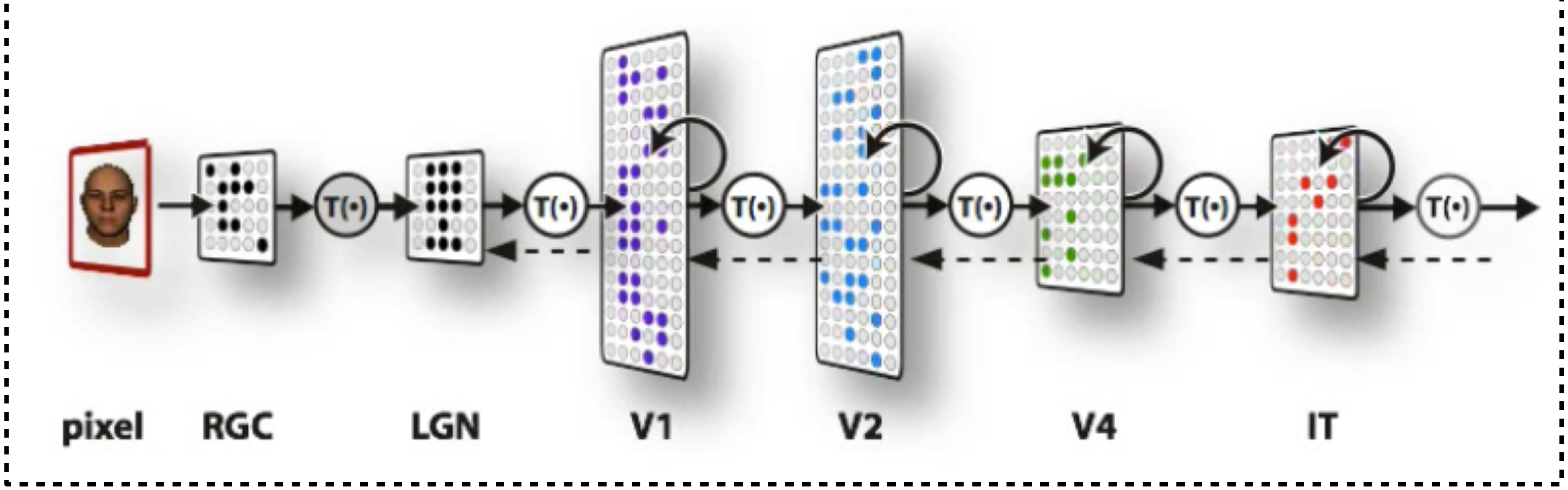
GOAL: Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.



Key questions:

(a) how many layers?

GOAL: Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.



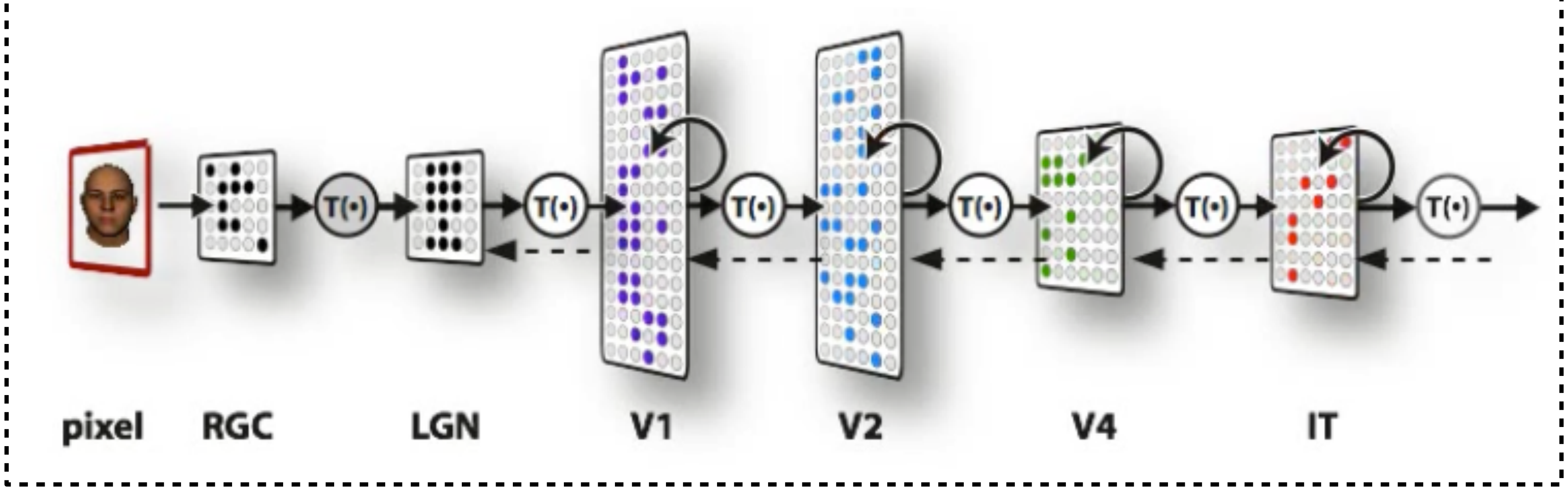
Key questions:

(a) how many layers?

(b) what's in each layer, specifically?

Ventral Stream

<u>GOAL: Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.</u>



Key questions:

(a) how many layers?

(b) what's in each layer, specifically?

(c) what behavioral goals and biophysical facts constrain it to be as it is?

<u>How are we supposed to use all this hard-won (Retina-IT) neuroscience knowledge to make an actual model?</u>