Class 19

A model summarizes what we know about how visual cortex “works”

Tomaso Poggio

Plan for class 21-22- 23

- Class 19: A class of models of the ventral stream of visual cortex

- Class 20: A “magic” theory of the ventral stream: Part I

- Class 21: A “magic” theory of the ventral stream: Part I I and III

Intro and connections with other classes

1. Problem of visual recognition, visual cortex
2. Historical background
3. Neurons and areas in the visual system
4. Feedforward hierarchical models
5. Beyond hierarchical models

Connection with the topic of learning theory

Notices of the American Mathematical Society (AMS), Vol. 50, No. 5, 537-544, 2003.

The Mathematics of Learning: Dealing with Data

Tomaso Poggio and Steve Smale

How then do the learning machines described in the theory compare with brains?

- One of the most obvious differences is the ability of people and animals to learn from very few examples. The algorithms we have described can learn an object recognition task from a few thousand labeled images but a child, or even a monkey, can learn the same task from just a few examples. Thus an important area for future theoretical and experimental work is learning from partially labeled examples

- A comparison with real brains offers another, related, challenge to learning theory. The "learning algorithms" we have described in this paper correspond to one-layer architectures. Are hierarchical architectures with more layers justifiable in terms of learning theory? It seems that the learning theory of the type we have outlined does not offer any general argument in favor of hierarchical learning machines for regression or classification.

- Why hierarchies? There may be reasons of *efficiency* – computational speed and use of computational resources. For instance, the lowest levels of the hierarchy may represent a dictionary of features that can be shared across multiple classification tasks.

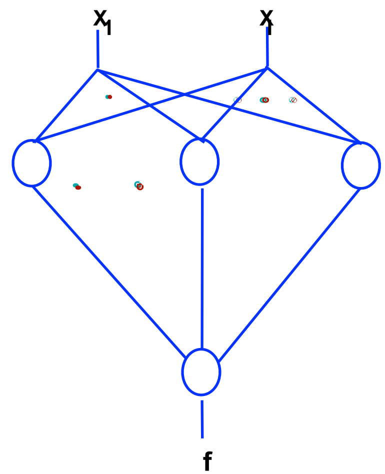
- There may also be the more fundamental issue of *sample complexity*. Learning theory shows that the difficulty of a learning task depends on the size of the required hypothesis space. This complexity determines in turn how many training examples are needed to achieve a given level of generalization error. Thus our ability of learning from just a few examples, and its limitations, may be related to the hierarchical architecture of cortex.

Classical learning theory and Kernel Machines (Regularization in RKHS)

$$\min_{f \in H} \left[\frac{1}{n} \sum_{i=1}^{n} V(f(x_i) - y_i) + \lambda \ \|f\|_K^2 \right]$$

implies

$$f(\mathbf{x}) = \sum_i^n \alpha_i K(\mathbf{x}, \mathbf{x}_i)$$



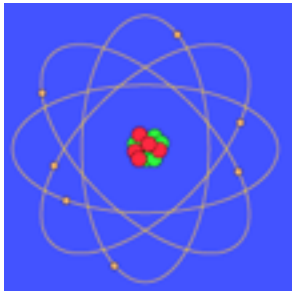
Remark:

Kernel machines correspond to *shallow* networks

WARNING:

using a class of models to summarize/interpret experimental results

- Models are cartoons of reality, eg Bohr's model of the hydrogen atom



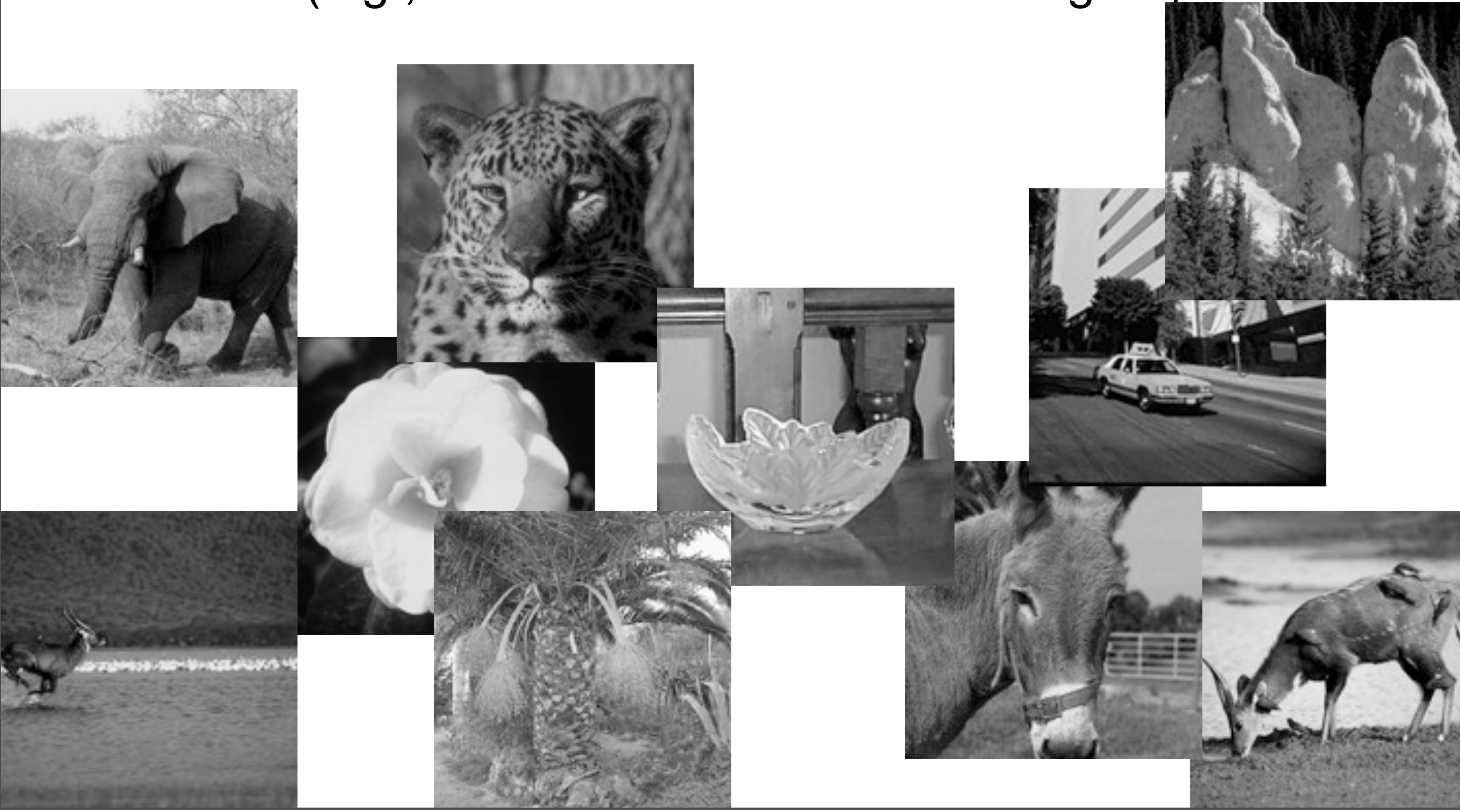
- All models are "wrong"

- Some models can be useful summaries of data and some can be a good starting point for a real *theory*

1. Problem of visual recognition, visual cortex
2. Historical background
3. Neurons and areas in the visual system
4. Feedforward hierarchical models
5. Beyond hierarchical models

Learning and Recognition in Visual Cortex: what is where

Unconstrained visual recognition is a difficult problem
(e.g., "is there an animal in the image?")



Vision: what is where

The MIT Press

YOUR PROFILE | TO ORDER | CONTACT US

The MIT Press is the only university press in the United States whose list is based in science and technology. This does not

Vision

A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr
Foreword by Shimon Ullman
Afterword by Tomaso Poggio

David Marr's posthumously published *Vision* (1982) influenced a generation of brain and cognitive scientists, inspiring many to enter the field. In *Vision*, Marr describes a general framework for understanding visual perception and touches on broader questions about how the brain and its functions can be studied and understood. Researchers from a range of brain and cognitive sciences have long valued Marr's creativity, intellectual power, and ability to integrate insights and data from neuroscience, psychology, and computation. This MIT Press edition makes Marr's influential work available to a new generation of students and scientists.

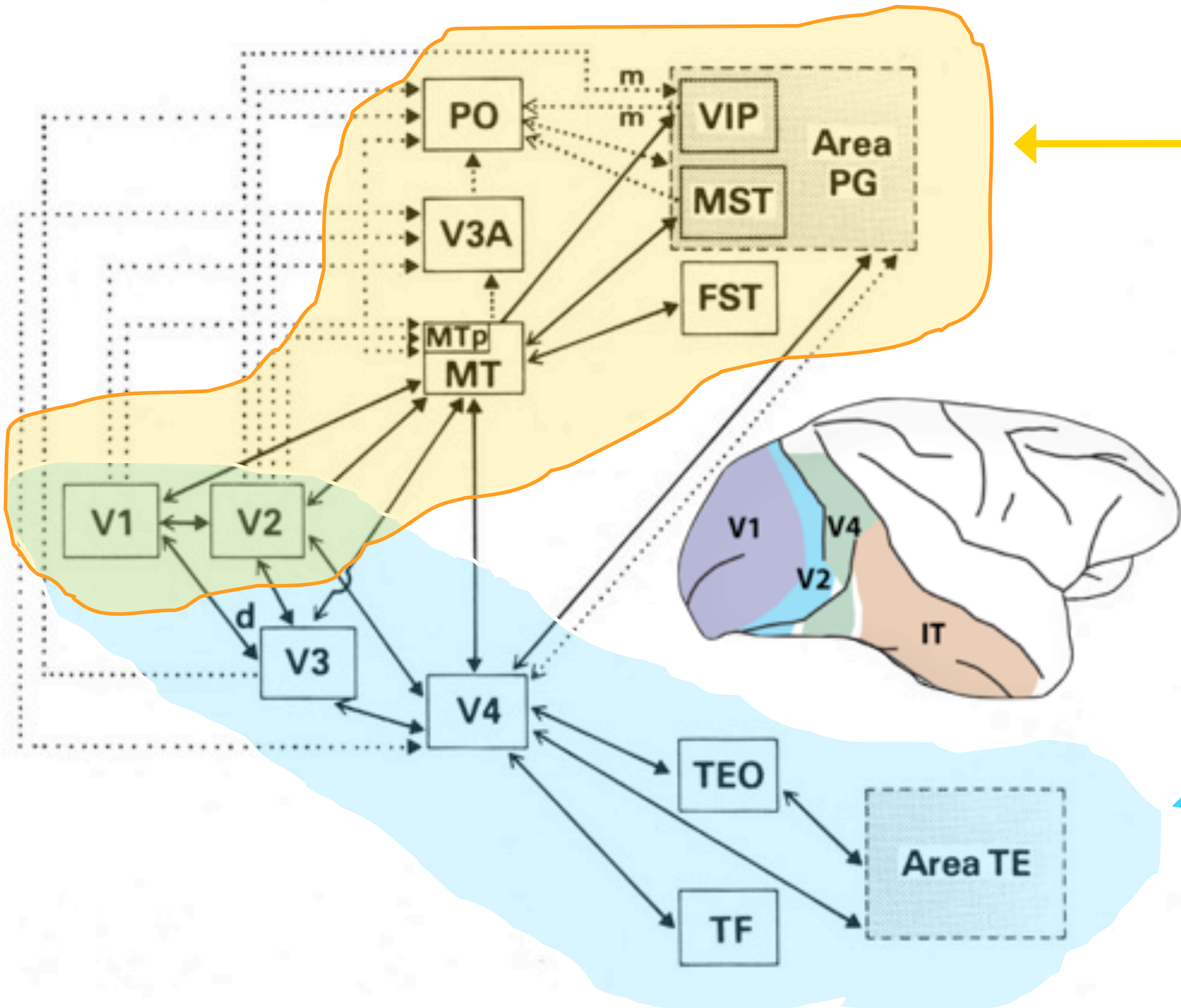
In Marr's framework, the process of vision constructs a set of representations, starting from a description of the input image and culminating with a description of three-dimensional objects in the surrounding environment. A central theme, and one that has had far-reaching influence in both neuroscience and cognitive science, is the notion of different levels of analysis—in Marr's framework, the computational level, the algorithmic level, and the hardware implementation level.

Now, thirty years later, the main problems that occupied Marr remain fundamental open problems in the study of perception. *Vision* provides inspiration for the continui

Vision: what is where

~ 1979 , with David Marr and Francis Crick, Borego Desert

Vision: what is where

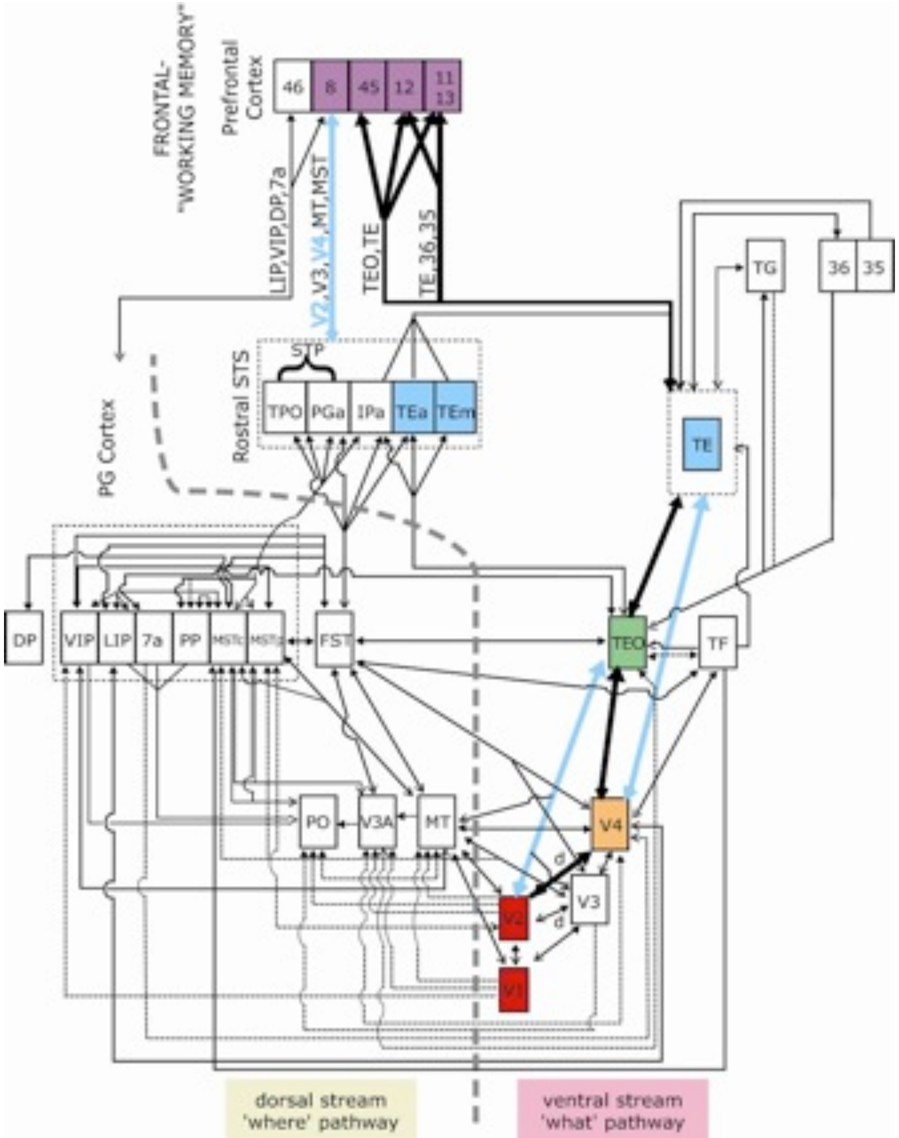


dorsal stream: "where"

ventral stream: "what"

Desimone & Ungerleider 1989

The ventral stream...

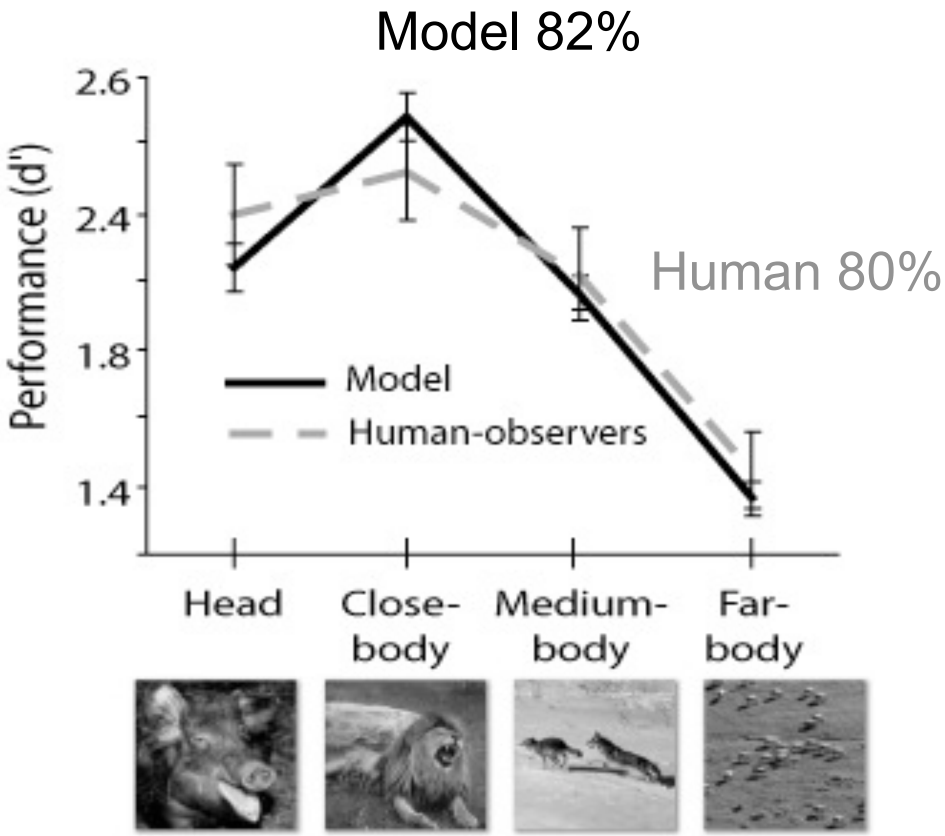


Feedforward connections only?

...``solves" the problem

(if the mask forces feedforward processing)...

- d'~ standardized error rate
- the higher the d', the better the performance



Serre Oliva & Poggio 2007

1. Problem of visual recognition, visual cortex
2. Historical background
3. Neurons and areas in the visual system
4. Feedforward hierarchical models
5. Beyond hierarchical models

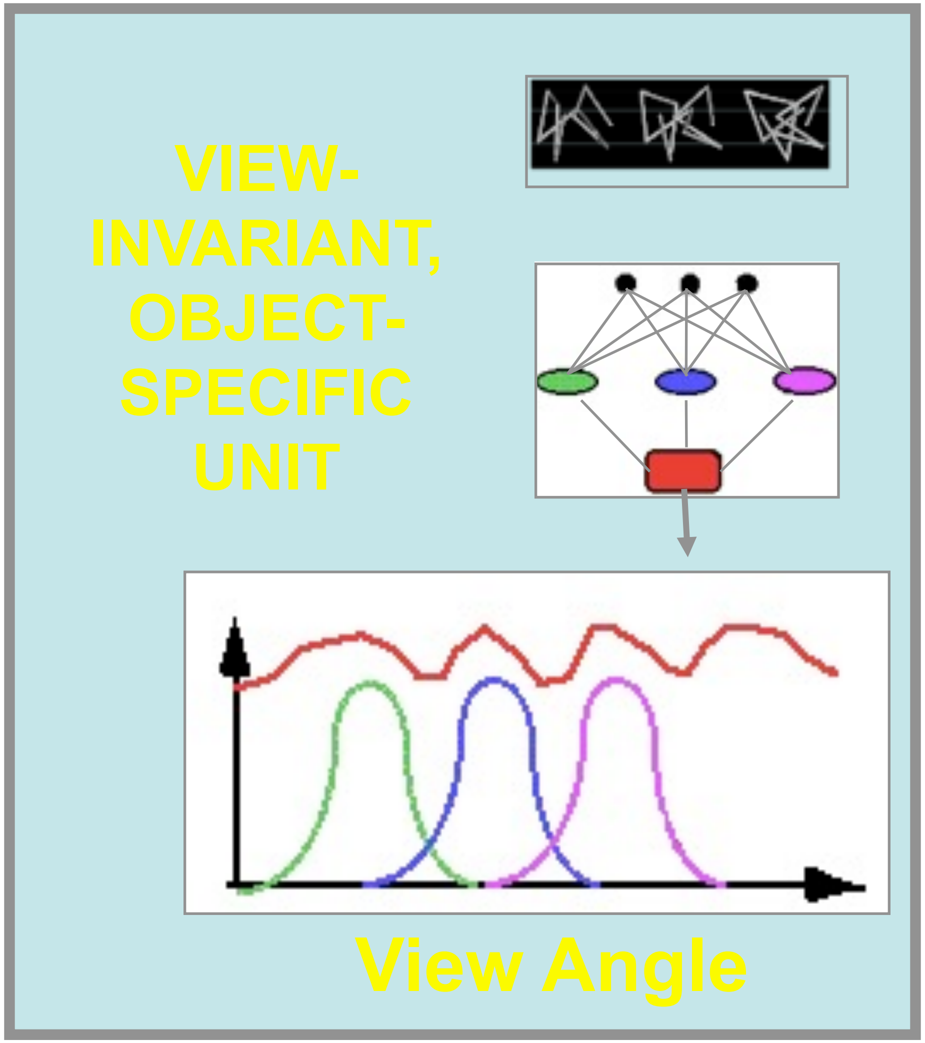
Some personal history:

First step in developing a model:
learning to recognize 3D objects in IT cortex

Poggio & Edelman 1990

An idea for a module for view-invariant identification

Architecture that accounts for invariances to 3D effects (>1 view needed to learn!)



Regularization Network (GRBF) with Gaussian kernels

Prediction: neurons become view-tuned through learning

Poggio & Edelman 1990

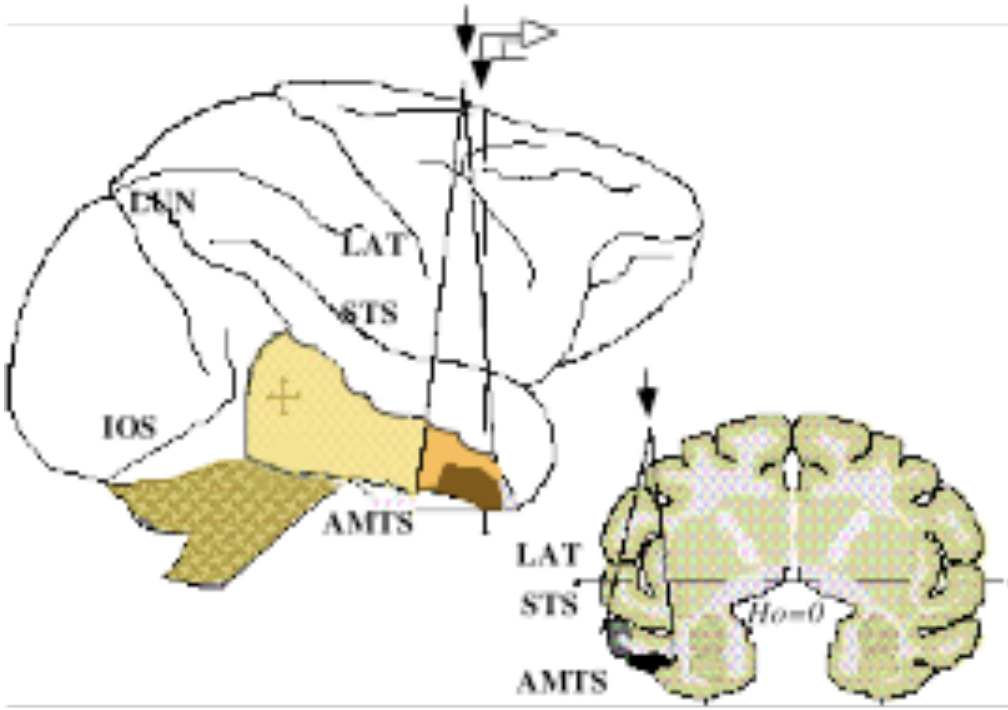
Learning to Recognize 3D Objects in IT Cortex

After human psychophysics (Buelthoff, Edelman, Tarr, Sinha, *to be added next year*…), which supports models based on view-tuned units...

… physiology!

Logothetis Pauls & Poggio 1995

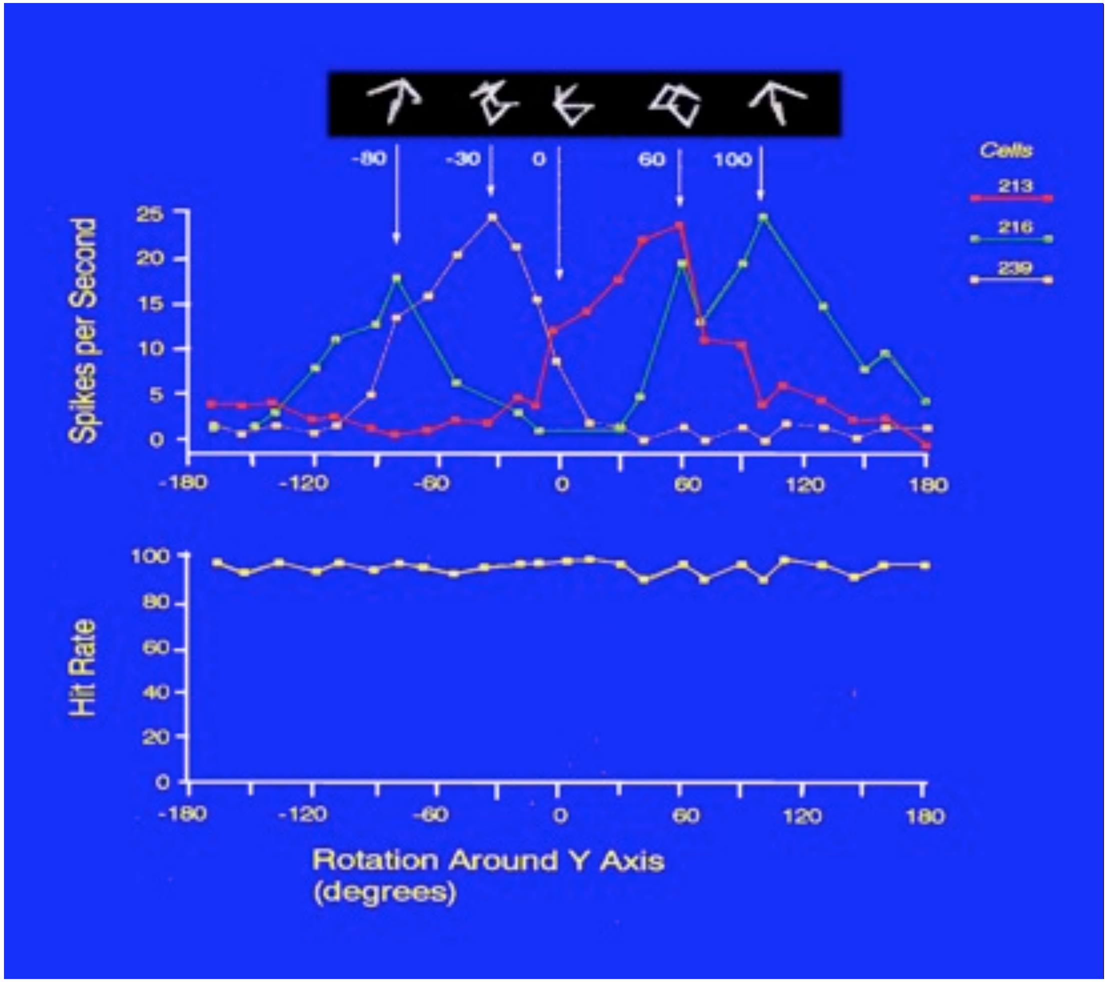
Recording Sites in Anterior IT



…neurons tuned to faces are intermingled nearby….

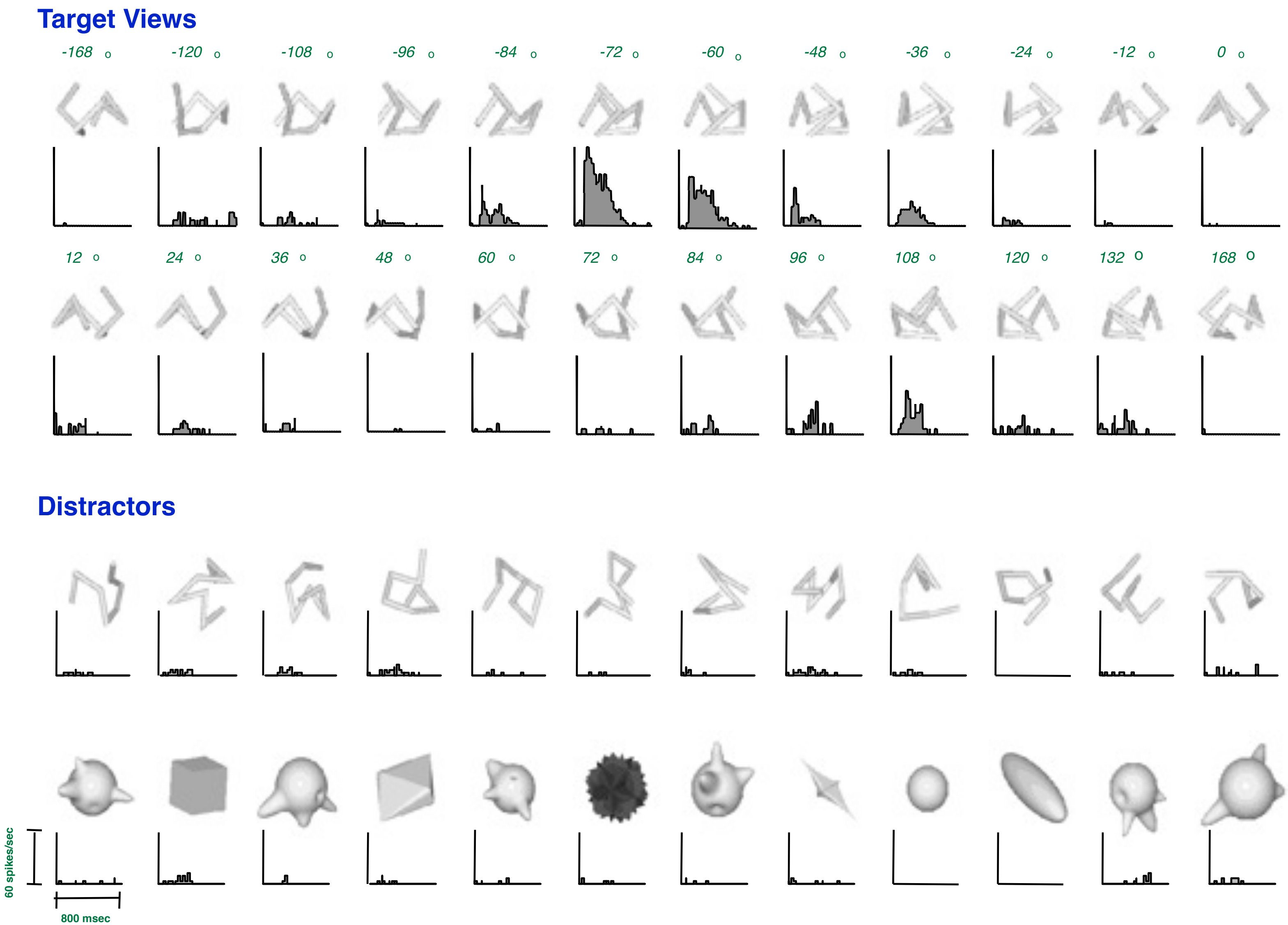
Logothetis, Pauls & Poggio 1995

Neurons tuned to object views, as predicted by model!



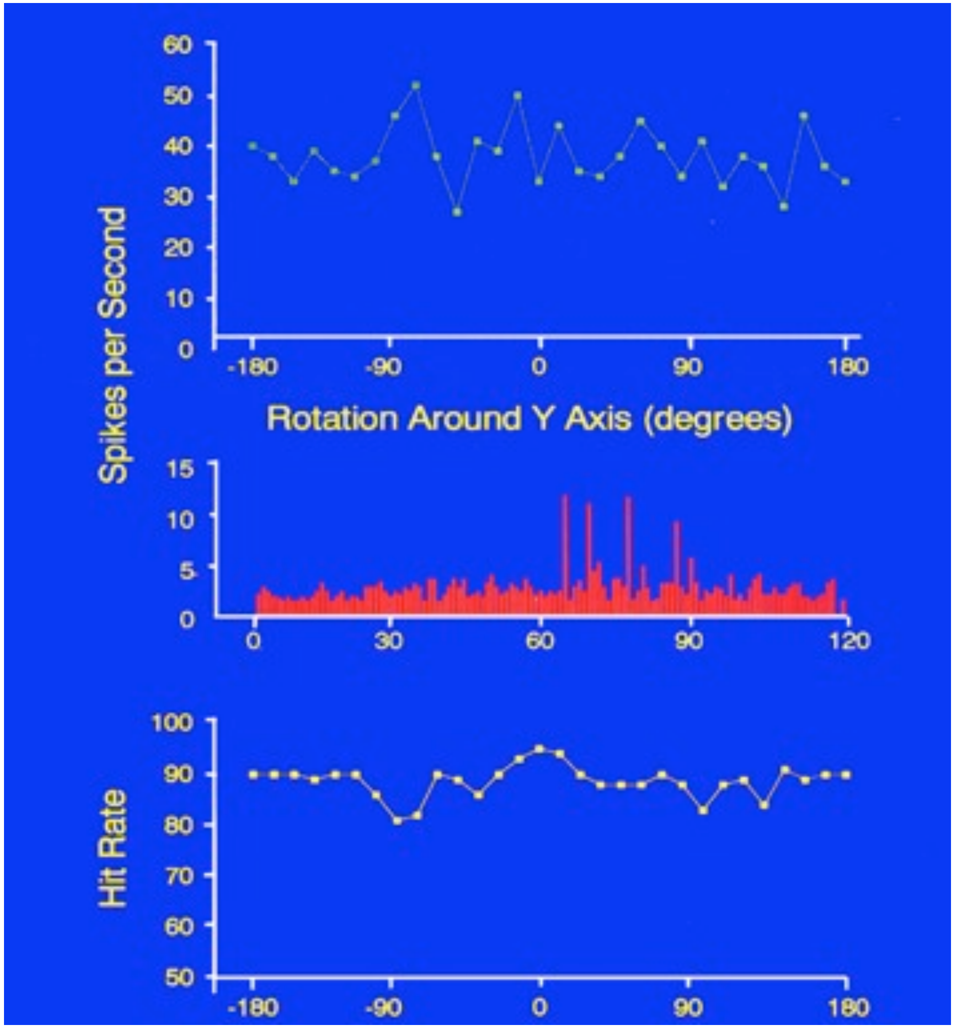
Logothetis Pauls & Poggio 1995

A “View-Tuned” IT Cell



Logothetis Pauls & Poggio 1995

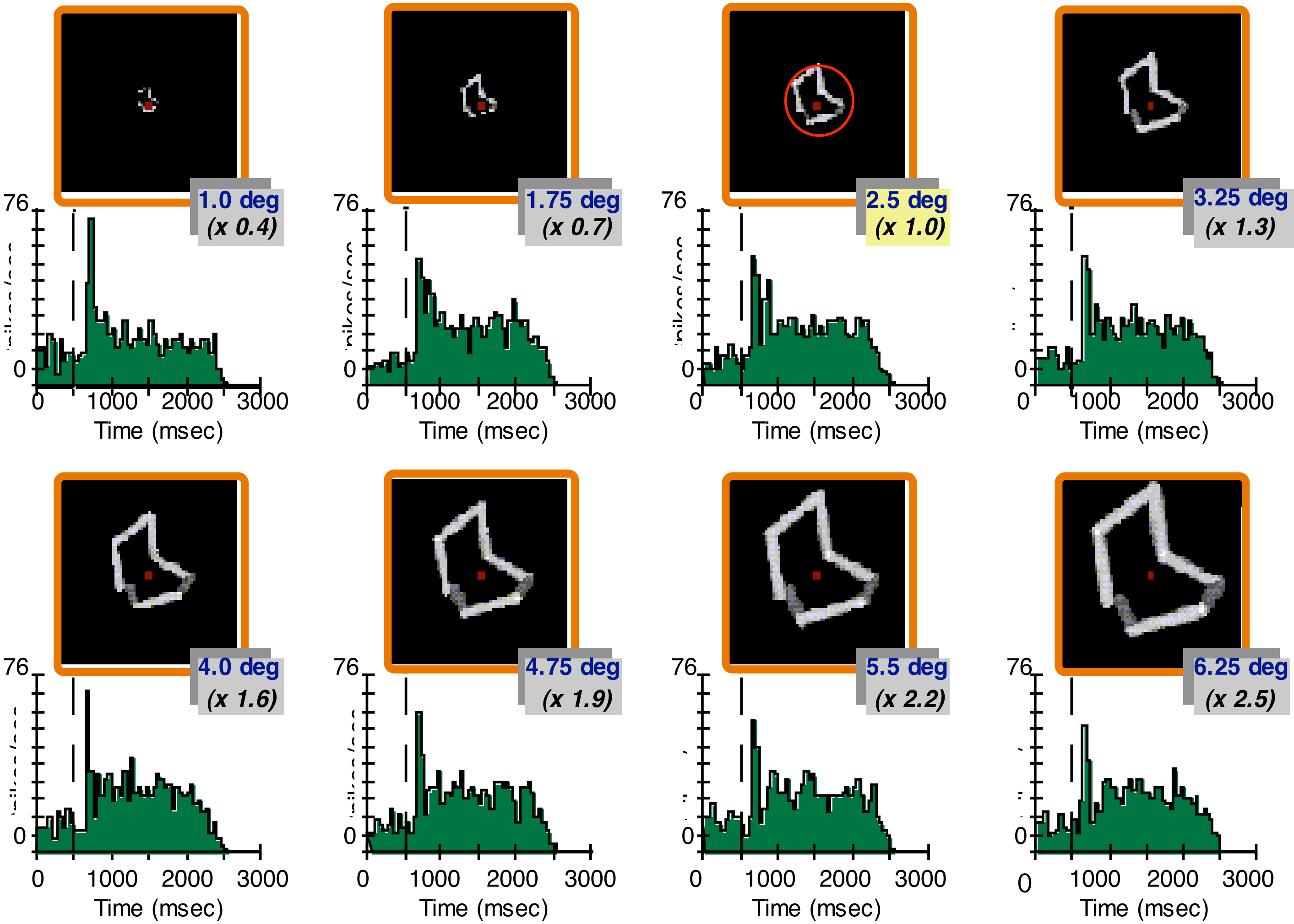
But also view-invariant object-specific neurons (5 of them over 1000 recordings)



Logothetis Pauls & Poggio 1995

View-tuned cells:

scale invariance (one training view only) motivates present model



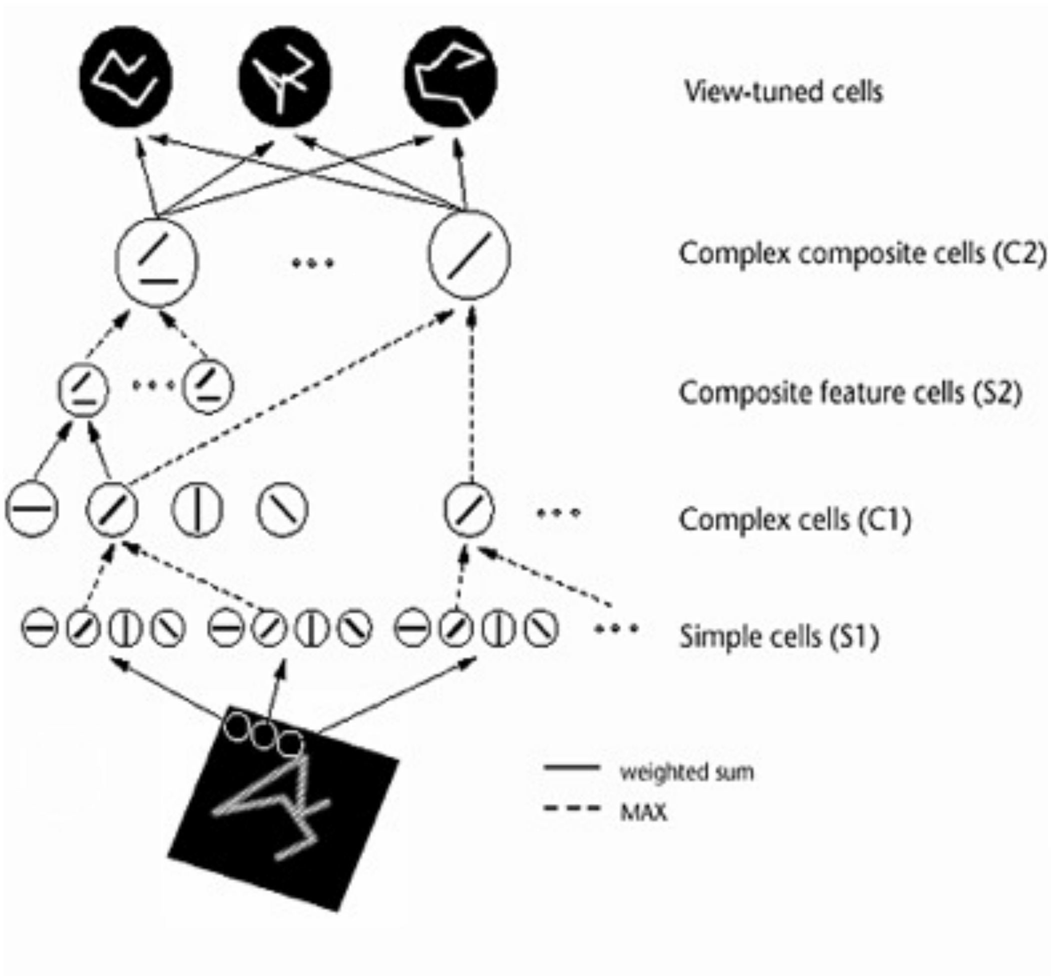
Logothetis Pauls & Poggio 1995

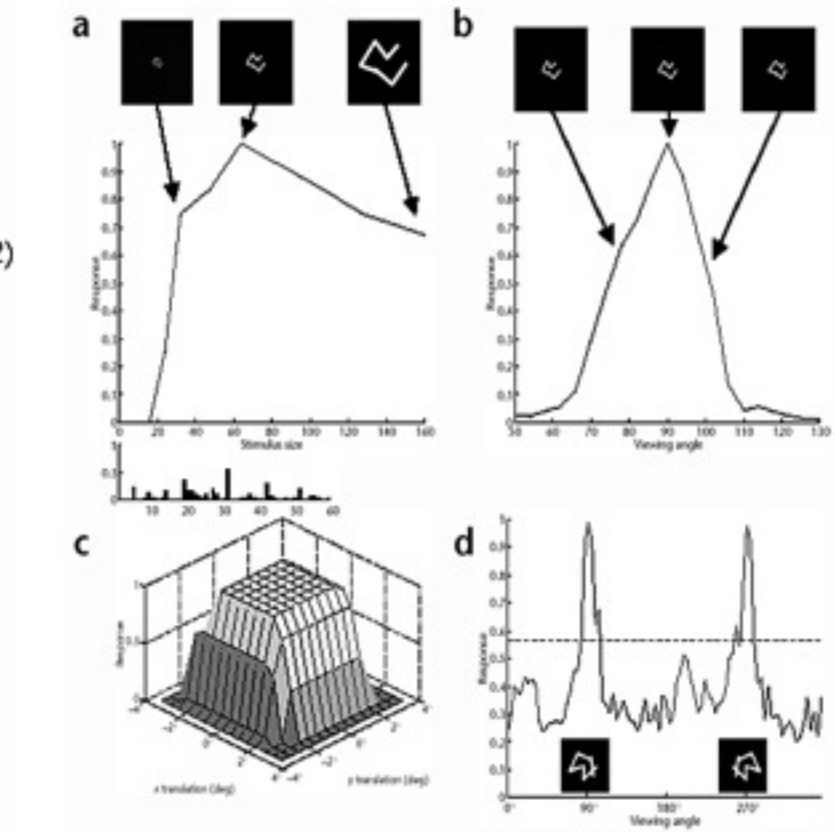
Hierarchy

- Gaussian centers (Gaussian Kernels) tuned to complex multidimensional features as composition of lower dimensional Gaussian

- What about tolerance to position and scale?
- Answer: hierarchy of invariance and tuning operations

Answer: the "HMAX" model





Riesenhuber & Poggio 1999, 2000

From HMAX to the present model

How the new version of the model evolved from the original one

1. **The two key operations:** Operations for selectivity and invariance, originally computed in a simplified and idealized form (i.e., a multivariate Gaussian and an exact max, see Section 2) have been replaced by more plausible operations, normalized dot-product and softmax

2. **S1 and C1 layers:** In [Serre and Riesenhuber, 2004] we found that the S1 and C1 units in the original model were too broadly tuned to orientation and spatial frequency and revised these units accordingly. In particular at the S1 level, we replaced Gaussian derivatives with Gabor filters to better fit parafoveal simple cells' tuning properties. We also modified both S1 and C1 receptive field sizes.

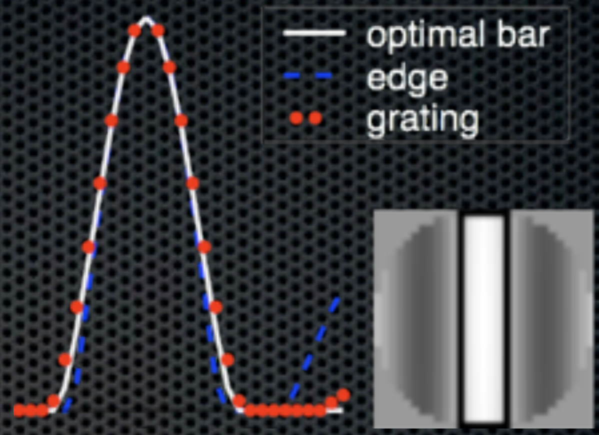
3. **S2 layers:** They are now learned from natural images. S2 units are more complex than the old ones (simple 2 °— 2 combinations of orientations). The introduction of learning, we believe, has b een the key factor for the model to achieve a high-level of performance on natural images, see [Serre et al., 2002].

4. **C2 layers:** Their receptive field sizes, as well as range of invariances to scale and position have been decreased so that C2 units now better fit V4 data.

5. **S3 and C3 layers:** They were recently added and constitute the top-most layers of the model along with the S2b and C2b units (see Section 2 and above). The tuning of the S3 units is also learned from natural images.

6. **S2b and C2b layers:** We added those two layers to account for the bypass route (that projects directly from V1/V2 to PIT, thus bypassing V4 [see Nakamura et al., 1993]).

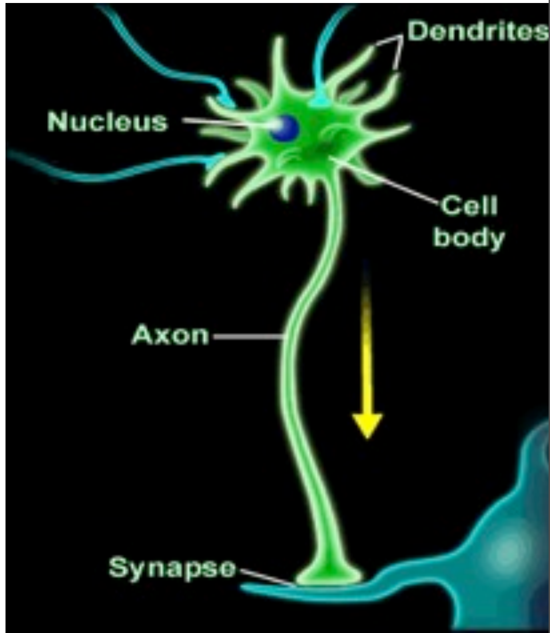
	Model	Cortex	References
	Receptive field sizes		
simple cells	$0.2° - 1.1°$	$\approx 0.1° - 1.0°$	[Schiller et al., 1976e; Hubel and Wiesel, 1965]
complex cells	$0.4° - 1.6°$	$\approx 0.2° - 2.0°$	
	Peak frequencies (cycles / deg)		
simple cells	range: 1.6 – 9.8 mean/med: 3.7/2.8	bulk $\approx$ 1.0 – 4.0 mean: $\approx$ 2.2 range: $\approx$ 0.5 – 8.0	[DeValois et al., 1982a])
complex cells	range: 1.8 – 7.8 mean/med: 3.9/3.2	bulk $\approx$ 2.0 – 5.6 mean: 3.2 range $\approx$ 0.5 – 8.0	
	Frequency bandwidth at 50% amplitude (cycles / deg)		
simple cells	range: 1.1 – 1.8 med: $\approx$ 1.45	bulk $\approx$ 1.0 – 1.5 med: $\approx$ 1.45 range $\approx$ 0.4 – 2.6	[DeValois et al., 1982a]
complex cells	range: 1.5 – 2.0 med: 1.6	bulk $\approx$ 1.0 – 2.0 med: 1.6 range $\approx$ 0.4 – 2.6	
	Frequency bandwidth at 71% amplitude (index)		
simple cells	range: 44 – 58 med: 55	bulk $\approx$ 40 – 70	[Schiller et al., 1976d]
complex cells	range 40 – 50 med. 48	bulk $\approx$ 40 – 60	
	Orientation bandwidth at 50% amplitude (octaves)		
simple cells	range: $38° - 49°$ med: $44°$	—	[DeValois et al., 1982b]
complex cells	range: $27° - 33°$ med: $43°$	bulk $\approx 20° - 90°$ med: $44°$	
	Orientation bandwidth at 71% amplitude (octaves)		
simple cells	range: $27° - 33°$ med: $30°$	bulk $\approx 20° - 70°$	[Schiller et al., 1976c]
complex cells	range: $27° - 33°$ med: $31°$	bulk $\approx 20° - 90°$	



Serre & Riesenhuber 2004

1. Problem of visual recognition, visual cortex
2. Historical background
3. Neurons and areas in the visual system
4. Feedforward hierarchical models
5. Beyond hierarchical models

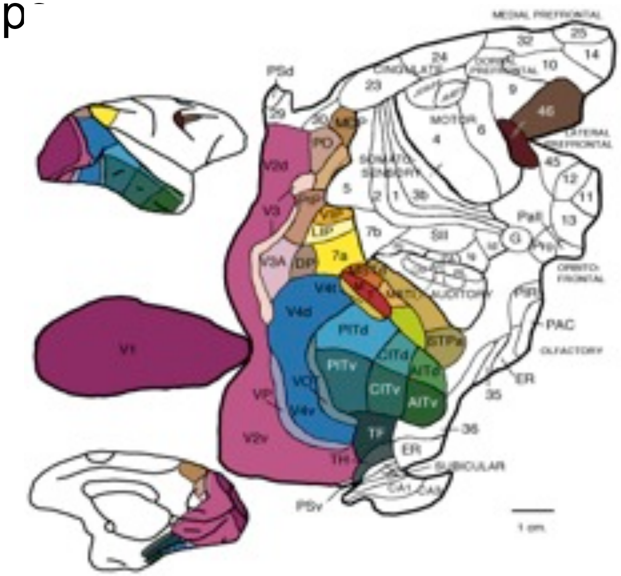
Vision: what is where



- Human Brain
  - 10^{10}-10^{11} neurons (~1 million flies)
  - 10^{14}- 10^{15} synapses

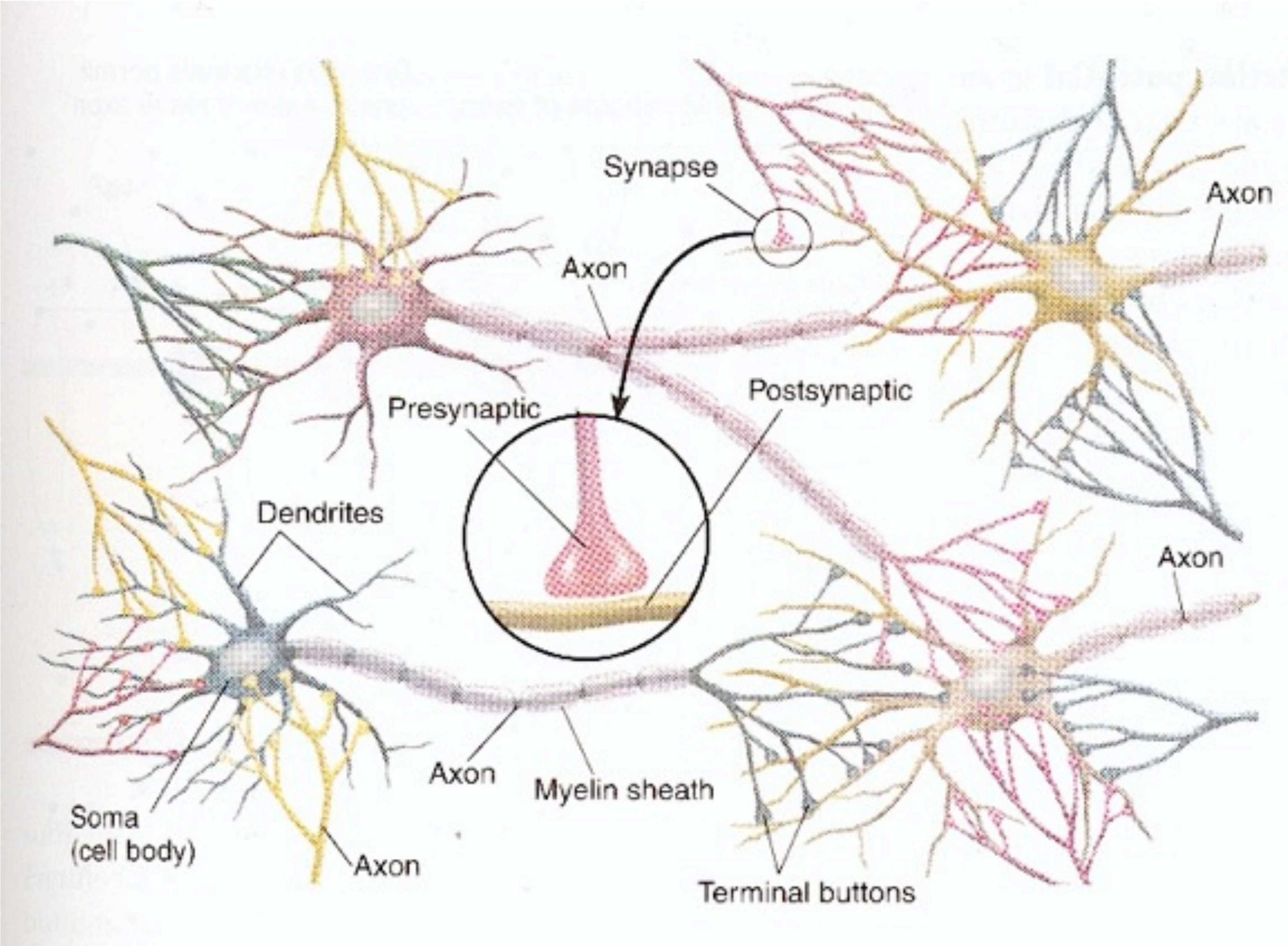
- Neuron
  - Fundamental space dimensions:
    - fine dendrites : 0.1 μ diameter; lipid bilayer membrane : 5 nm thick; specific proteins : pump channels, receptors, enzymes
  - Fundamental time length : 1 msec

- Ventral stream in rhesus monkey
  - ~10^9 neurons in the ventral stream ($350\ 10^6$ in each emisphere)
  - ~$15\ 10^6$ neurons in AIT (Anterior InferoTemporal) cortex



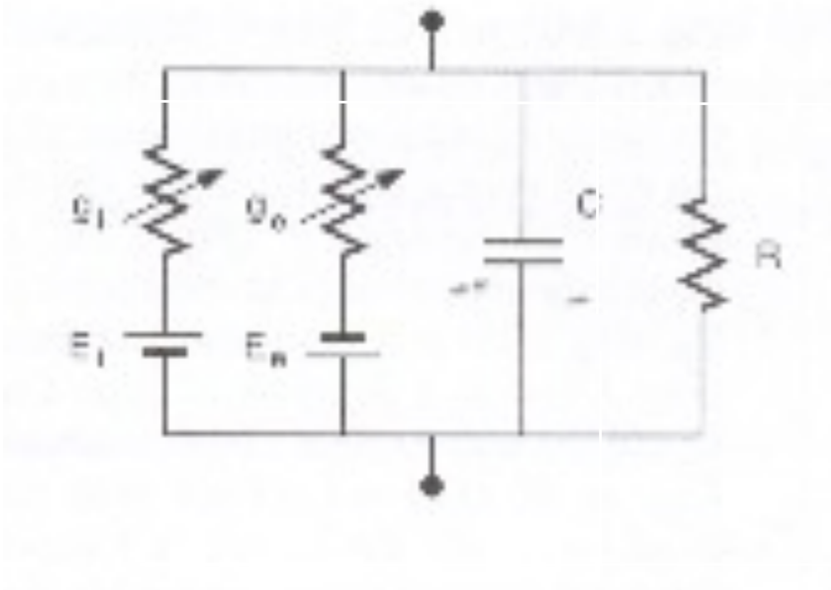
Van Essen & Anderson, 1990

Neural Circuits



Source: Modified from Jody Culham's web slides

Membrane with excitatory and inhibitory synapses

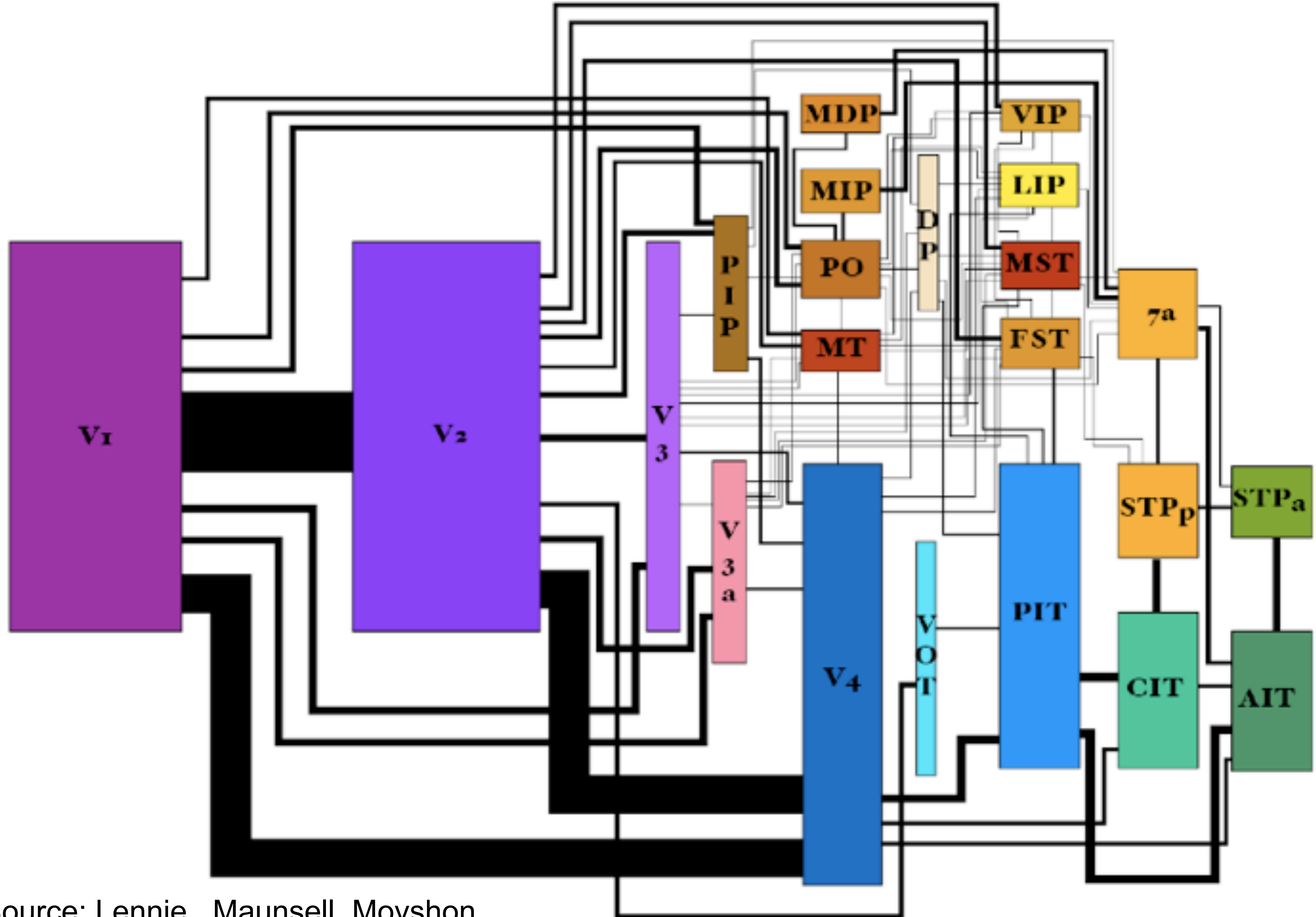


$$C\frac{dV}{dt}+g_i(V-E_i)+g_e(V-E_e)+g_0(V-V_{rest})=0$$

and with $\frac{dV}{dt}\approx 0$, $E_i \approx 0$, $V_{rest}\approx 0$, $\tilde{g}_e=\frac{g_e}{g_0}$ and $\tilde{g}_i=\frac{g_i}{g_0}$ we obtain

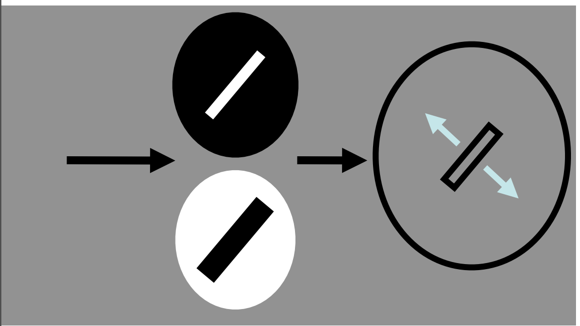
$$V \approx E_e\frac{\tilde{g}_e}{1+\tilde{g}_e+\tilde{g}_i}$$

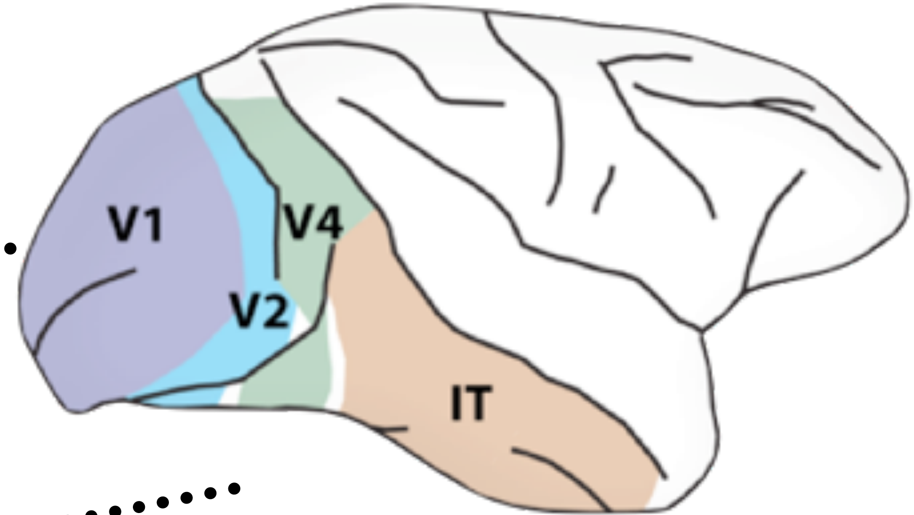
Vision: what is where

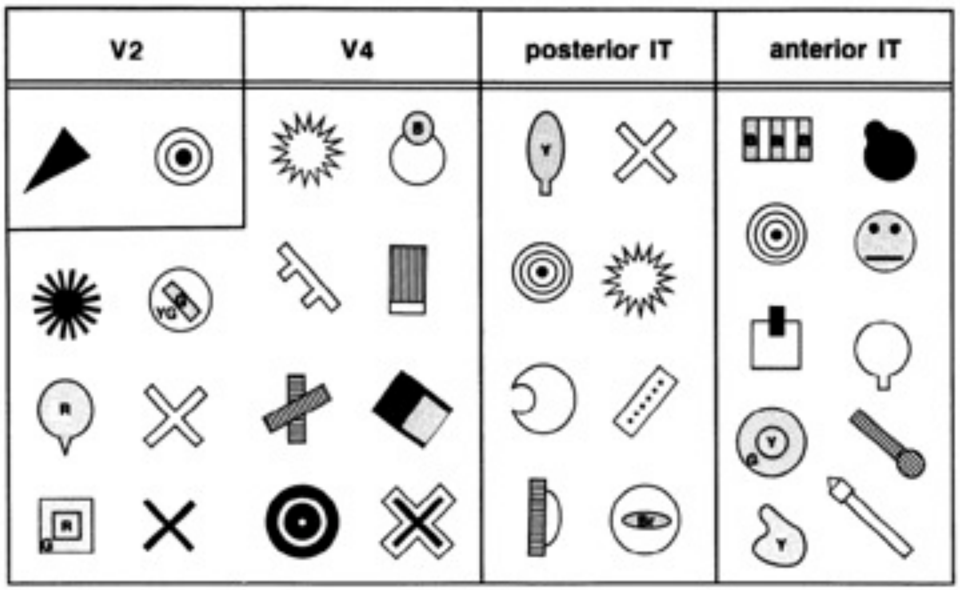


Source: Lennie, Maunsell, Movshon

Vision: what is







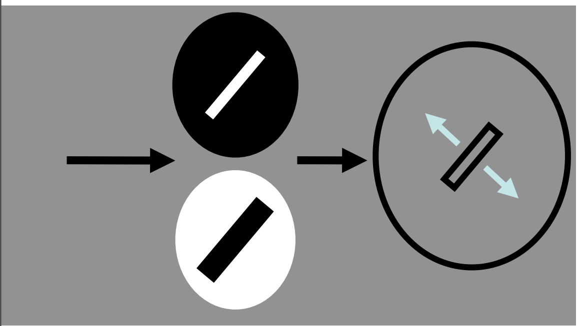
The ventral stream hierarchy: V1, V2, V4, IT

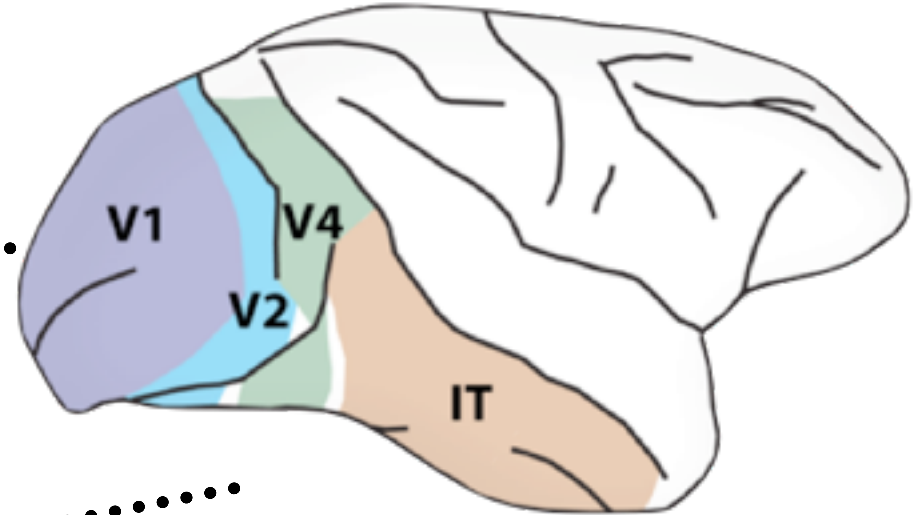
A gradual increase in the receptive field size, in the complexity of the preferred stimulus, in tolerance to position and scale changes

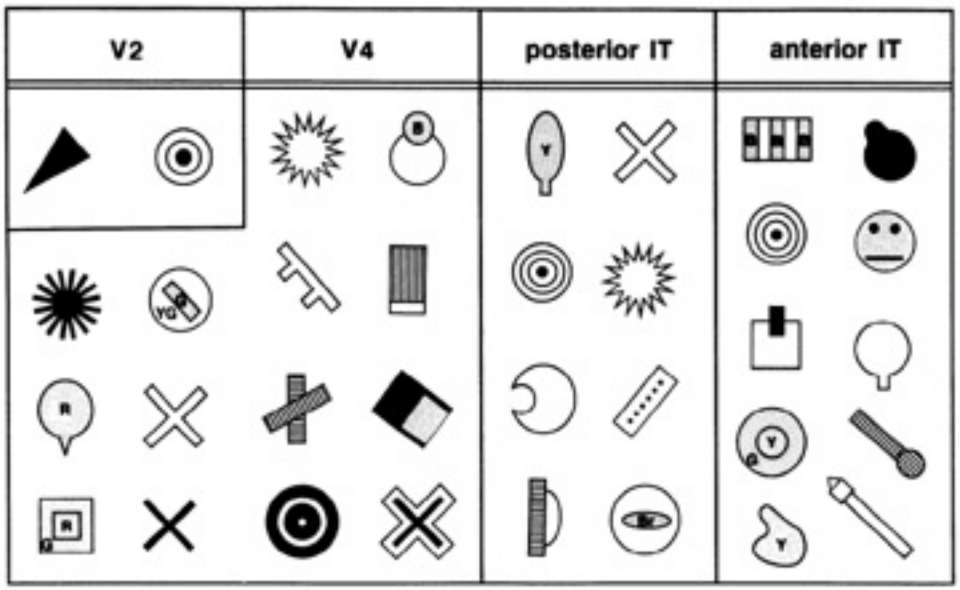
Kobatake & Tanaka, 1994

1. Problem of visual recognition, visual cortex
2. Historical background
3. Neurons and areas in the visual system
4. Feedforward hierarchical models
5. Beyond hierarchical models

The Ventral Stream



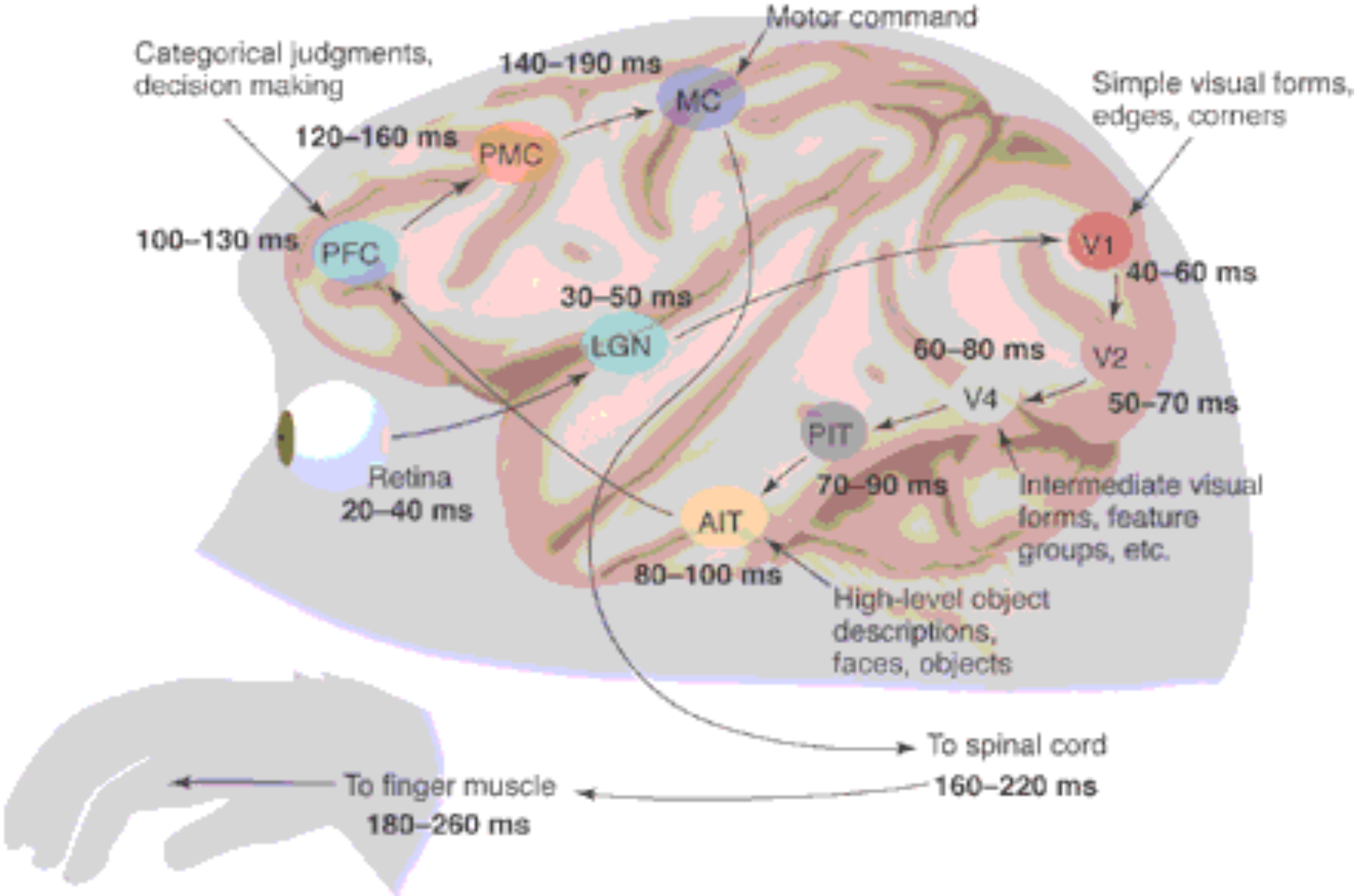




The ventral stream hierarchy: V1, V2, V4, IT

A gradual increase in the receptive field size, in the complexity of the preferred stimulus, in tolerance to position and scale changes

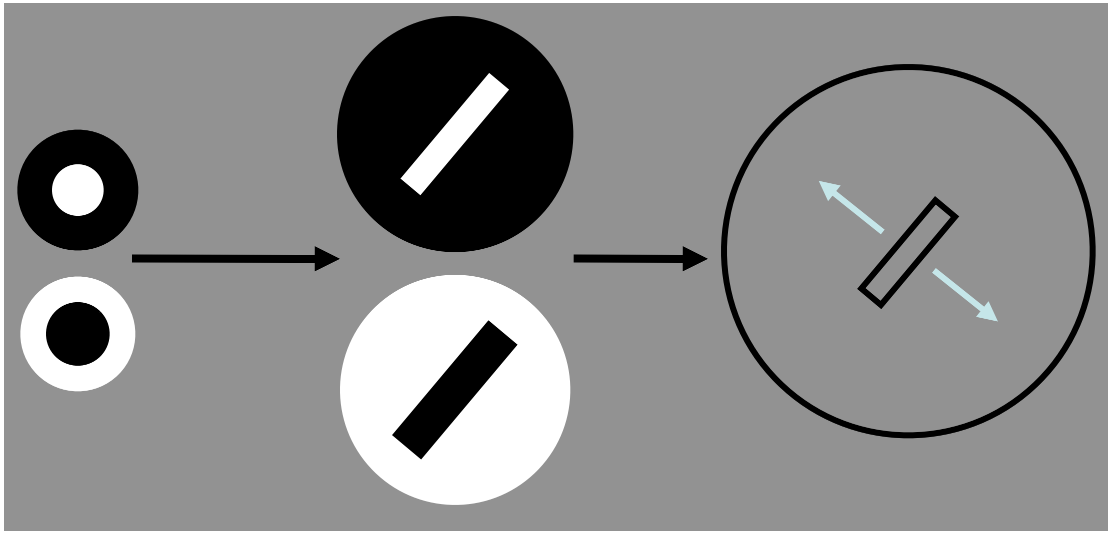
Kobatake & Tanaka, 1994



(Thorpe and Fabre-Thorpe, 2001)

V1: hierarchy of simple and complex cells

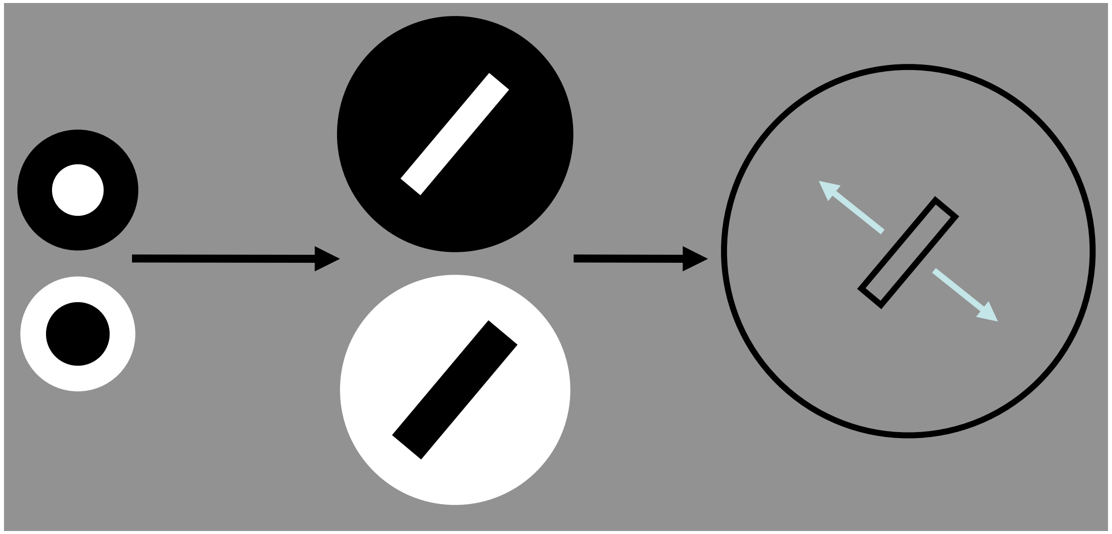
LGN-type cells | Simple cells | Complex cells

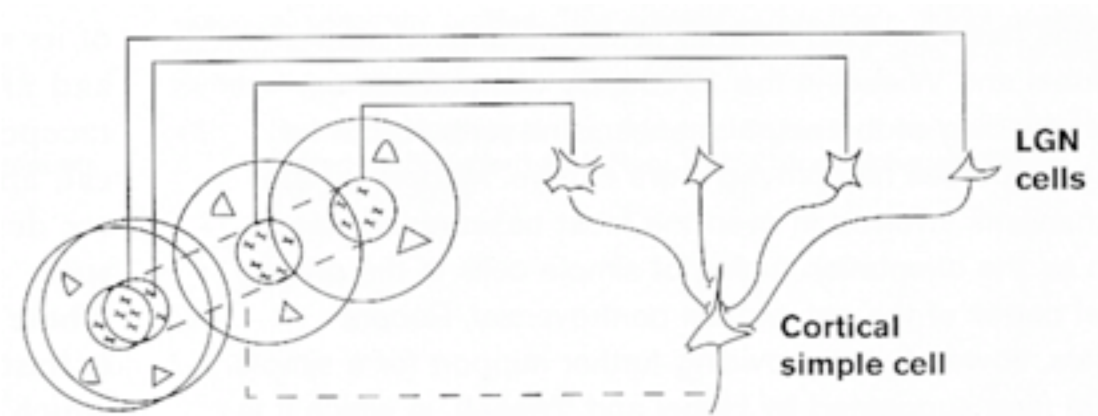


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells — Simple cells — Complex cells

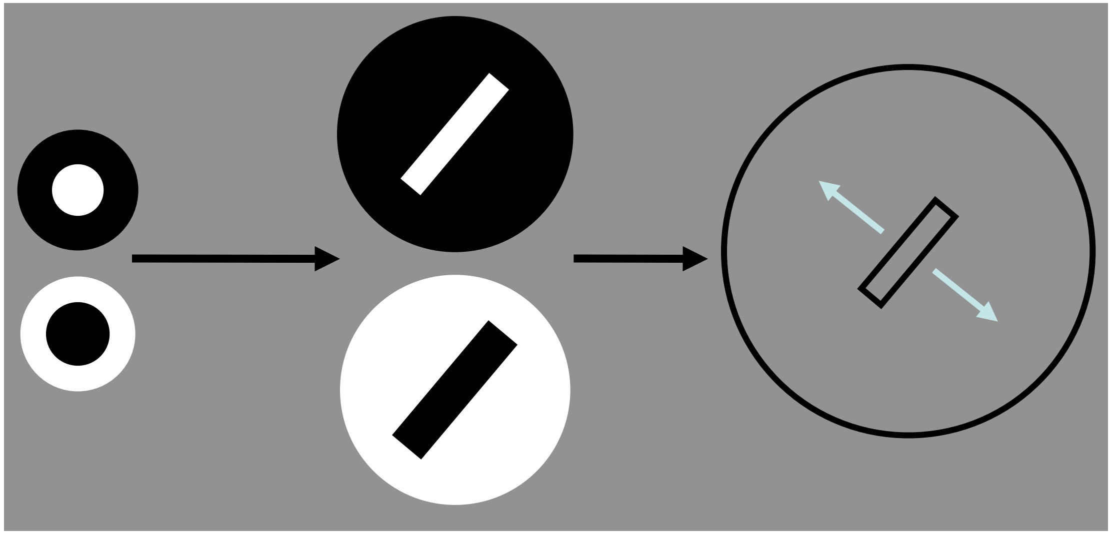


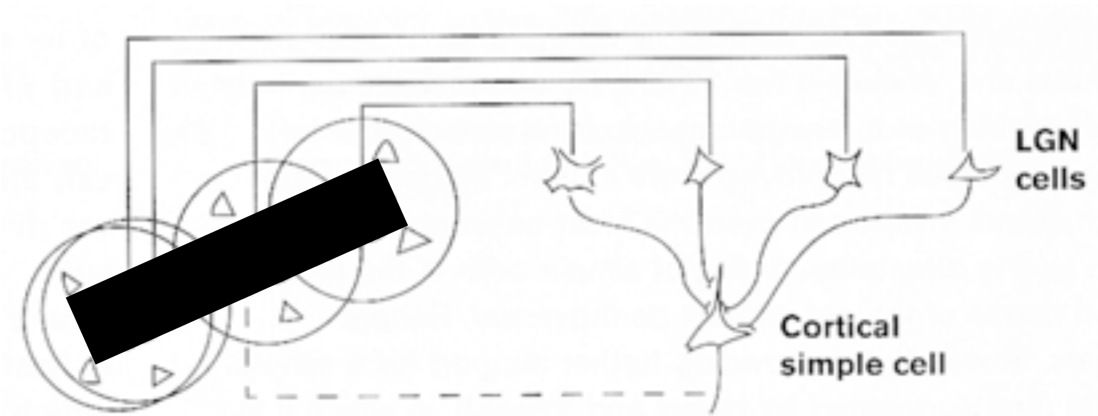


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells | Simple cells | Complex cells

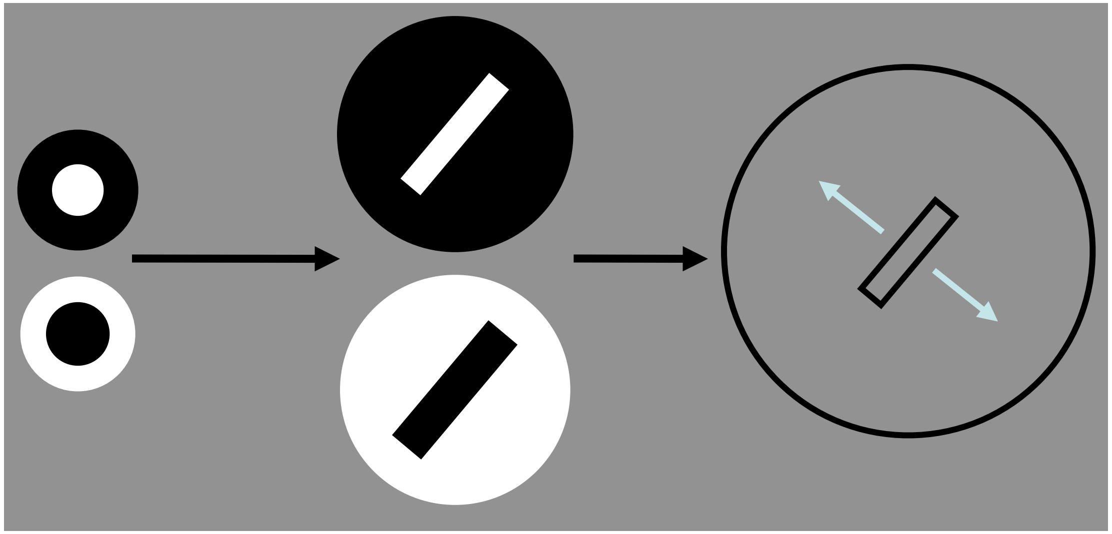




(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

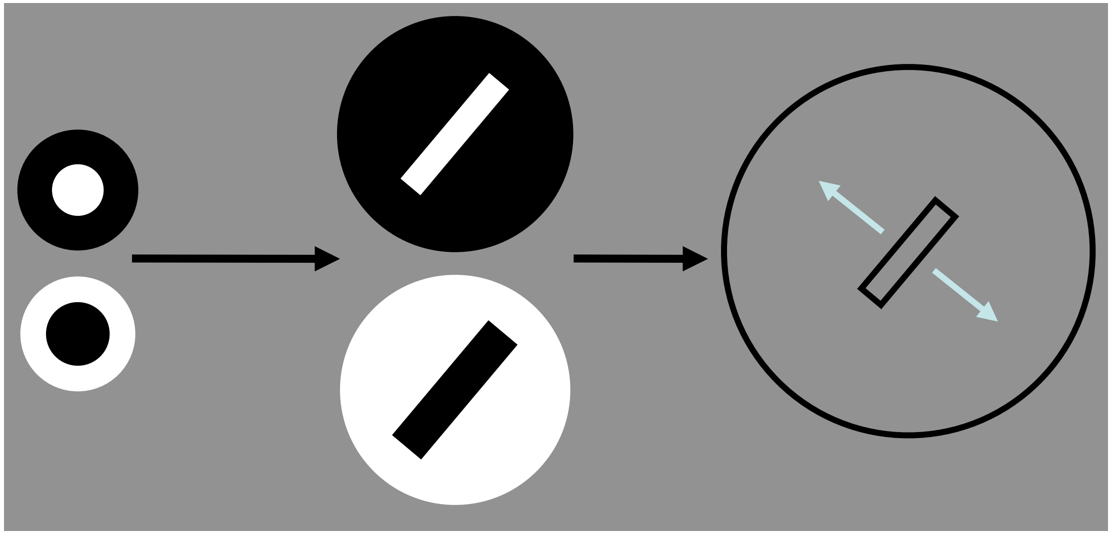
LGN-type cells | Simple cells | Complex cells

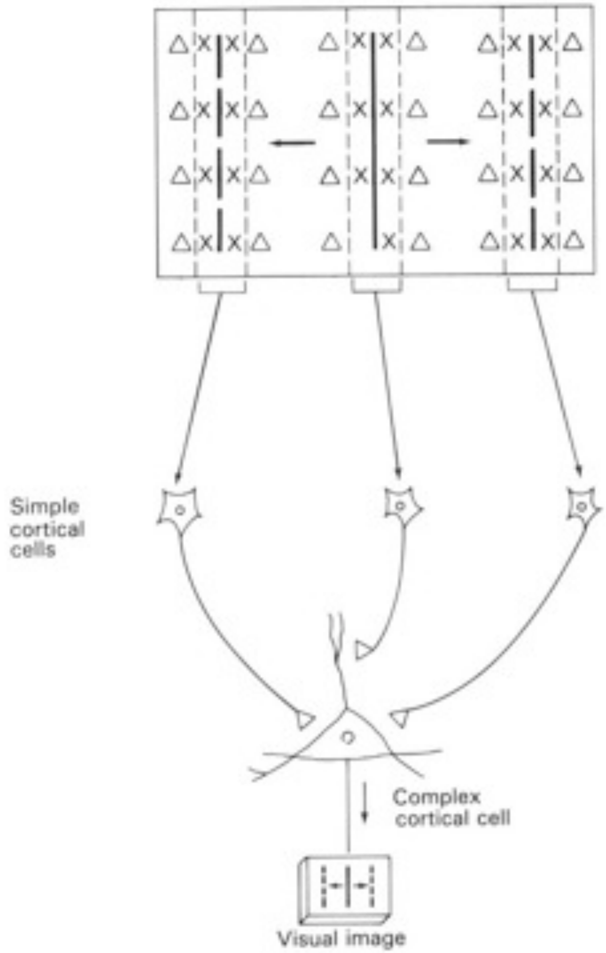


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells   Simple cells   Complex cells

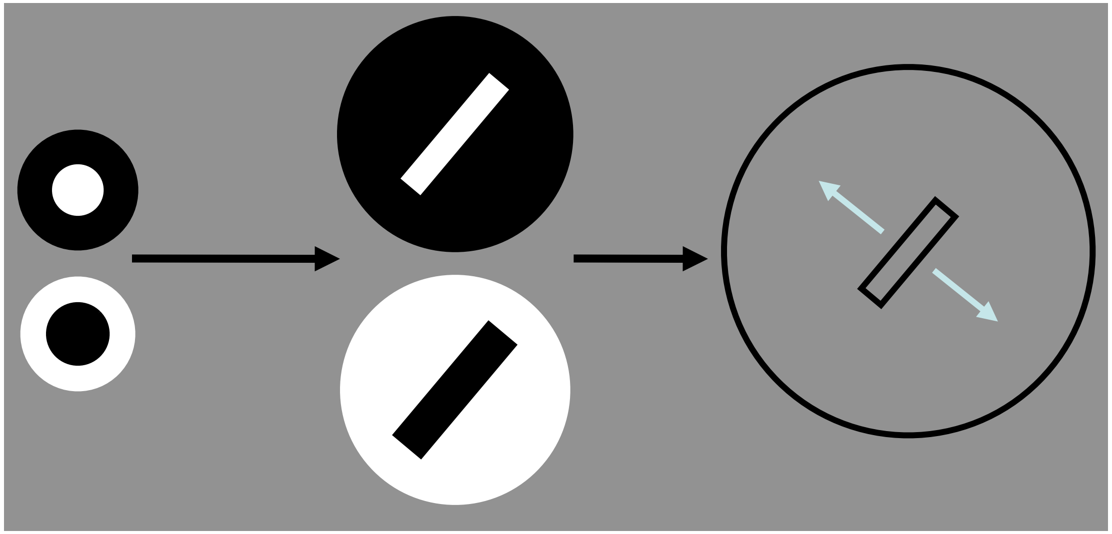


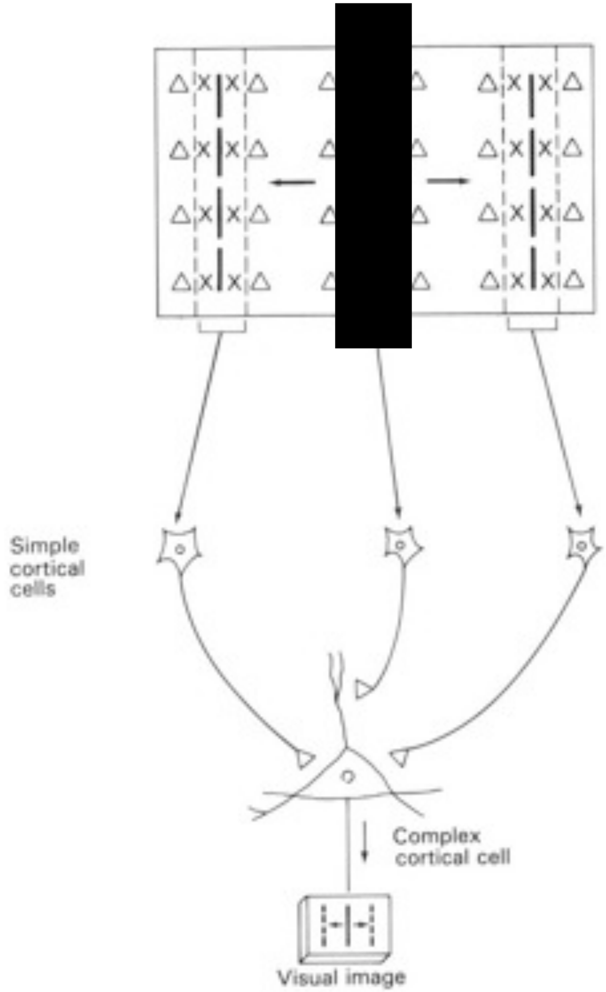


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells | Simple cells | Complex cells

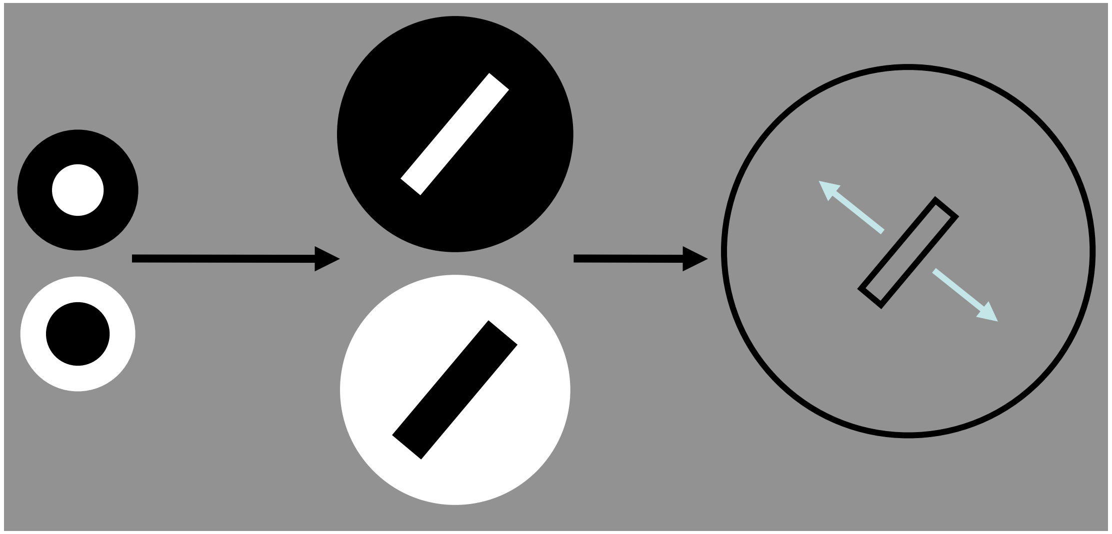


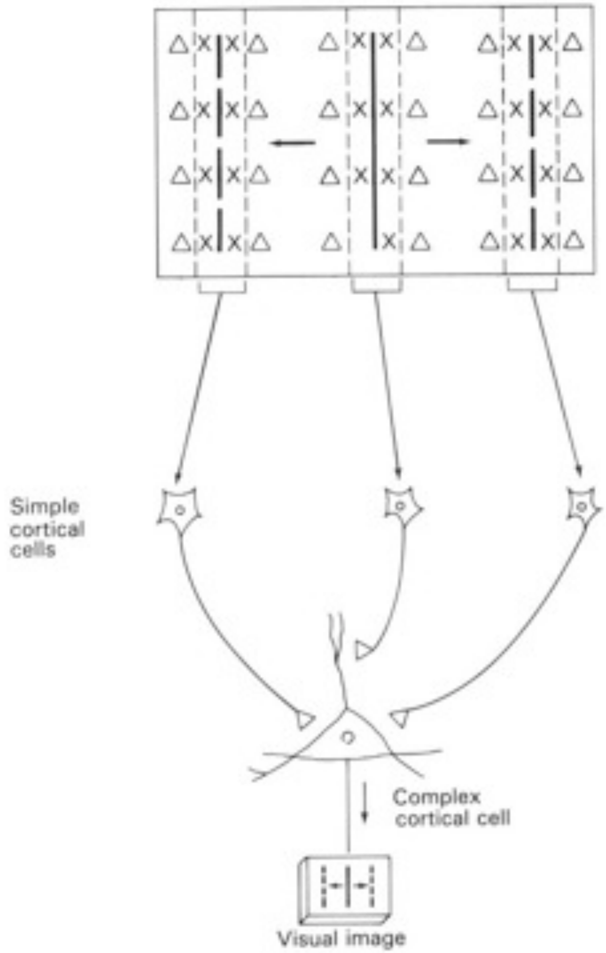


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells Simple cells Complex cells

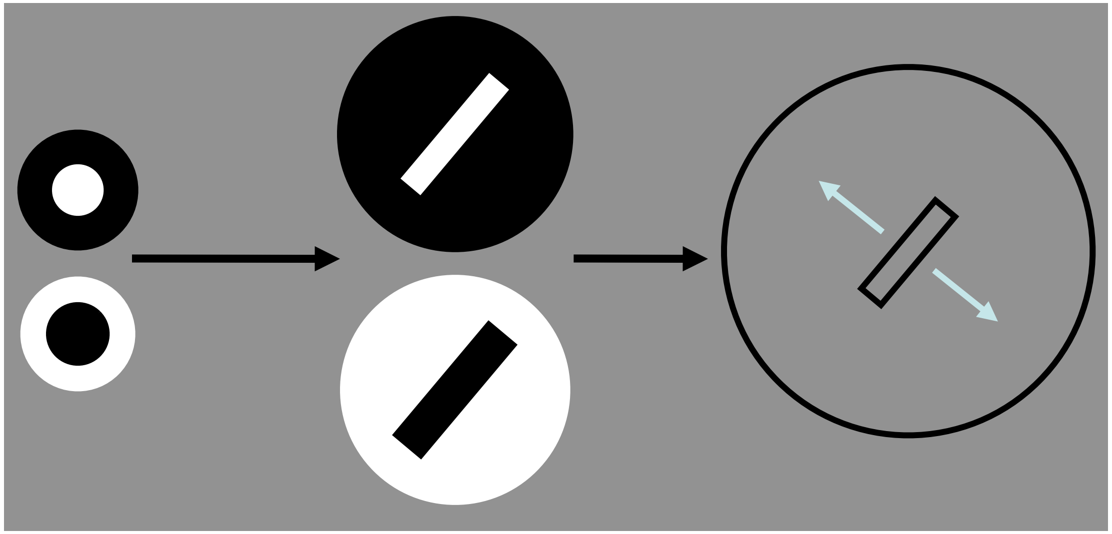


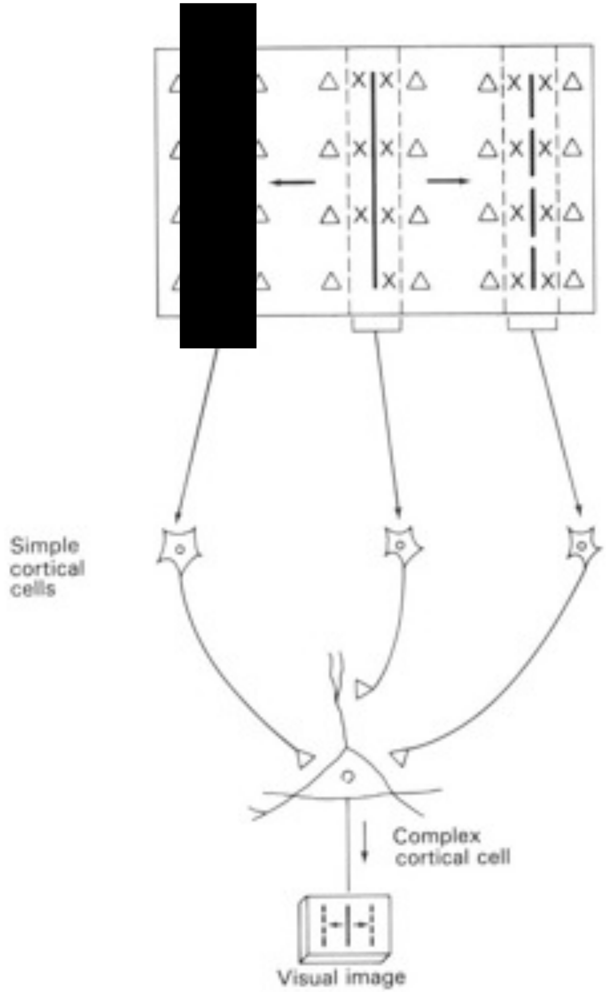


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells | Simple cells | Complex cells

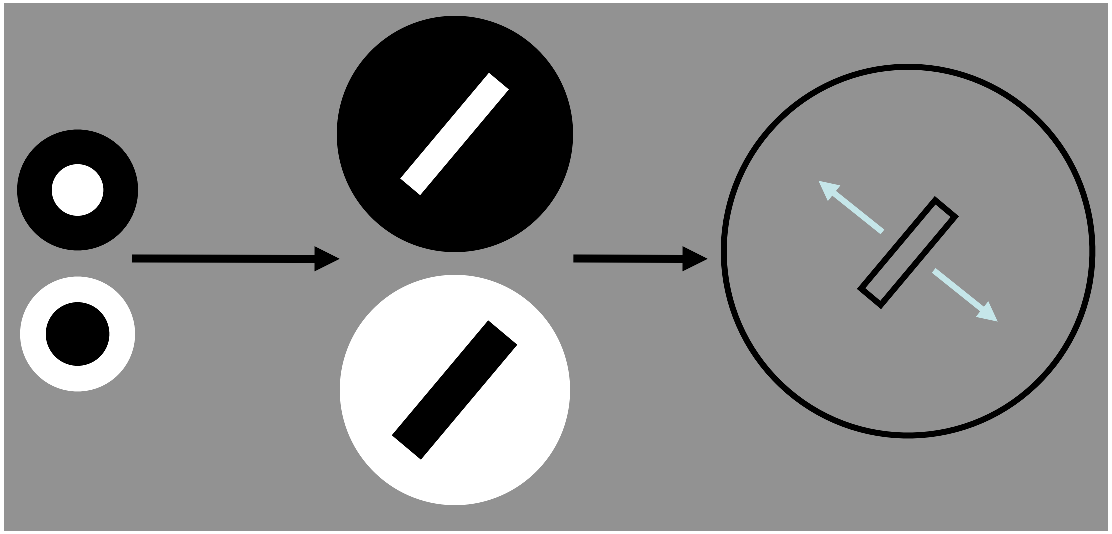


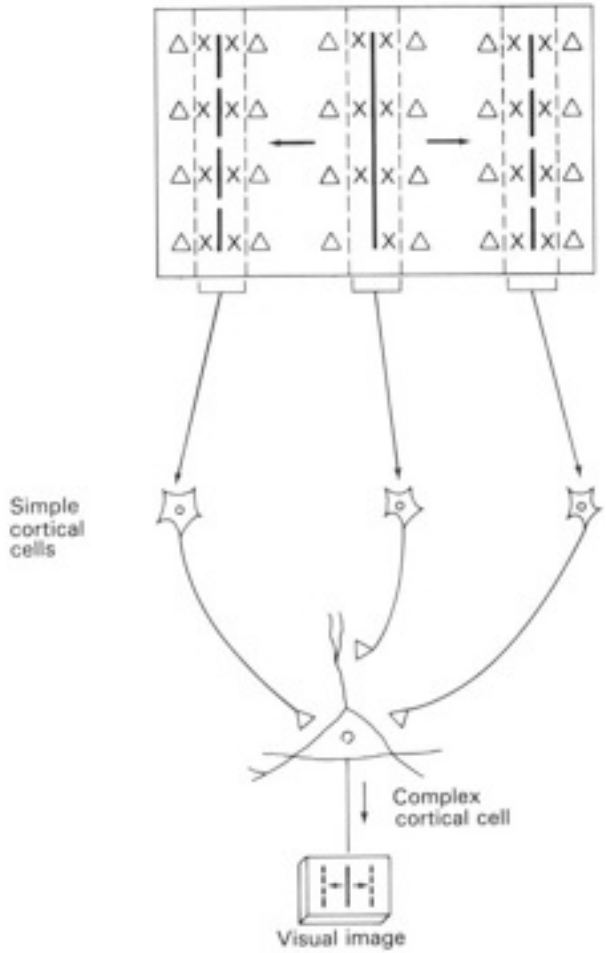


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells	Simple cells	Complex cells

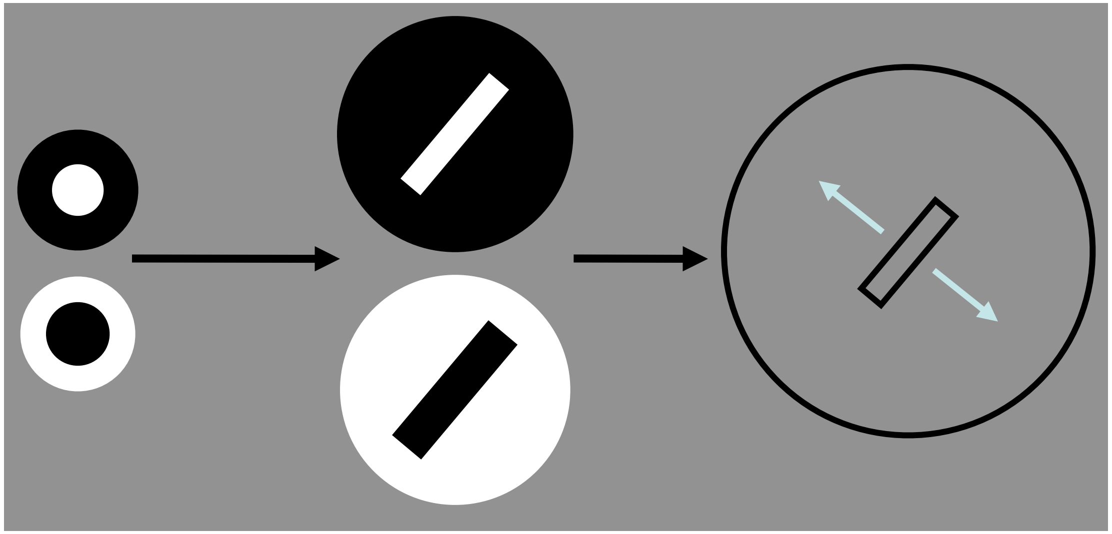


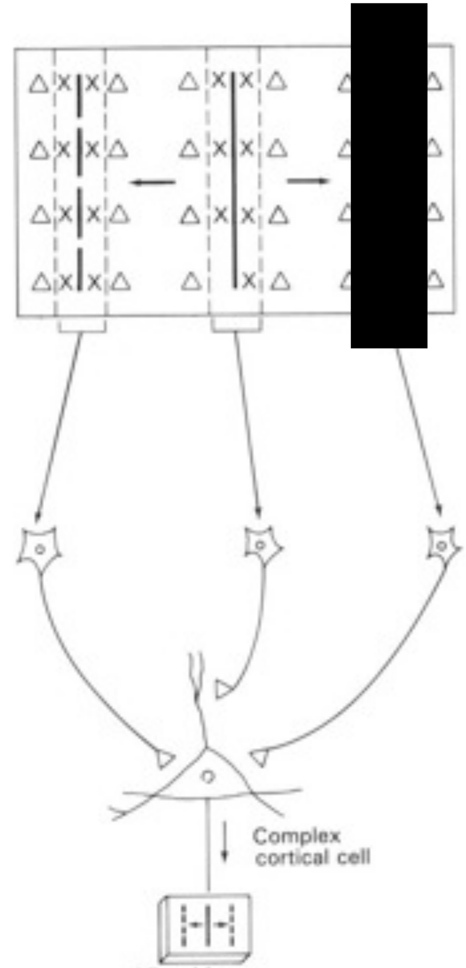


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells | Simple cells | Complex cells

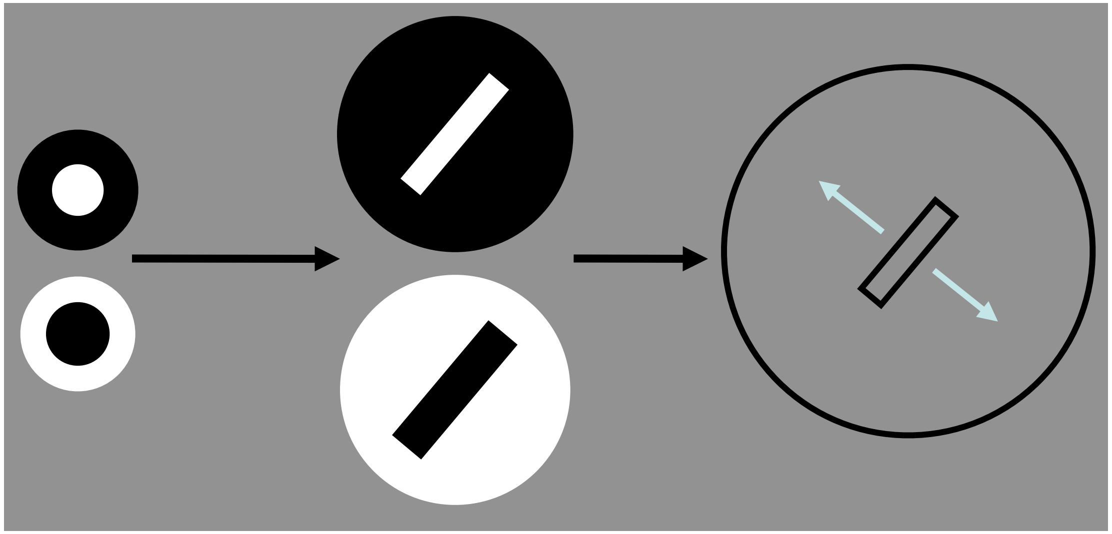


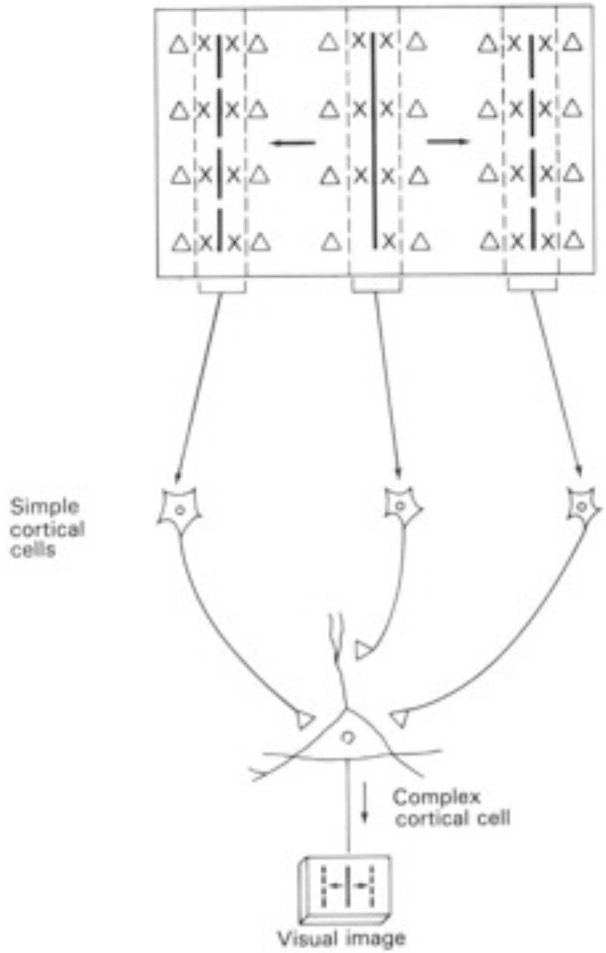


(Hubel & Wiesel 1959)

V1: hierarchy of simple and complex cells

LGN-type cells | Simple cells | Complex cells

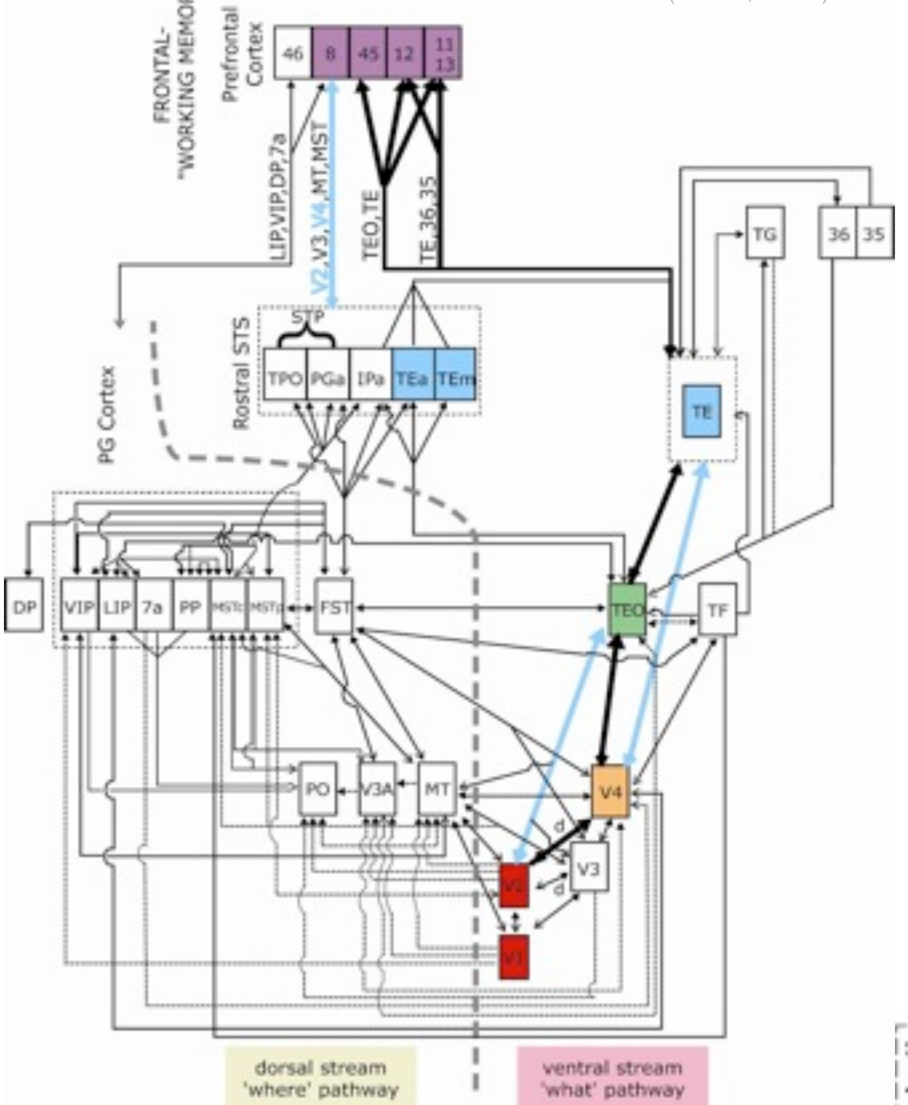


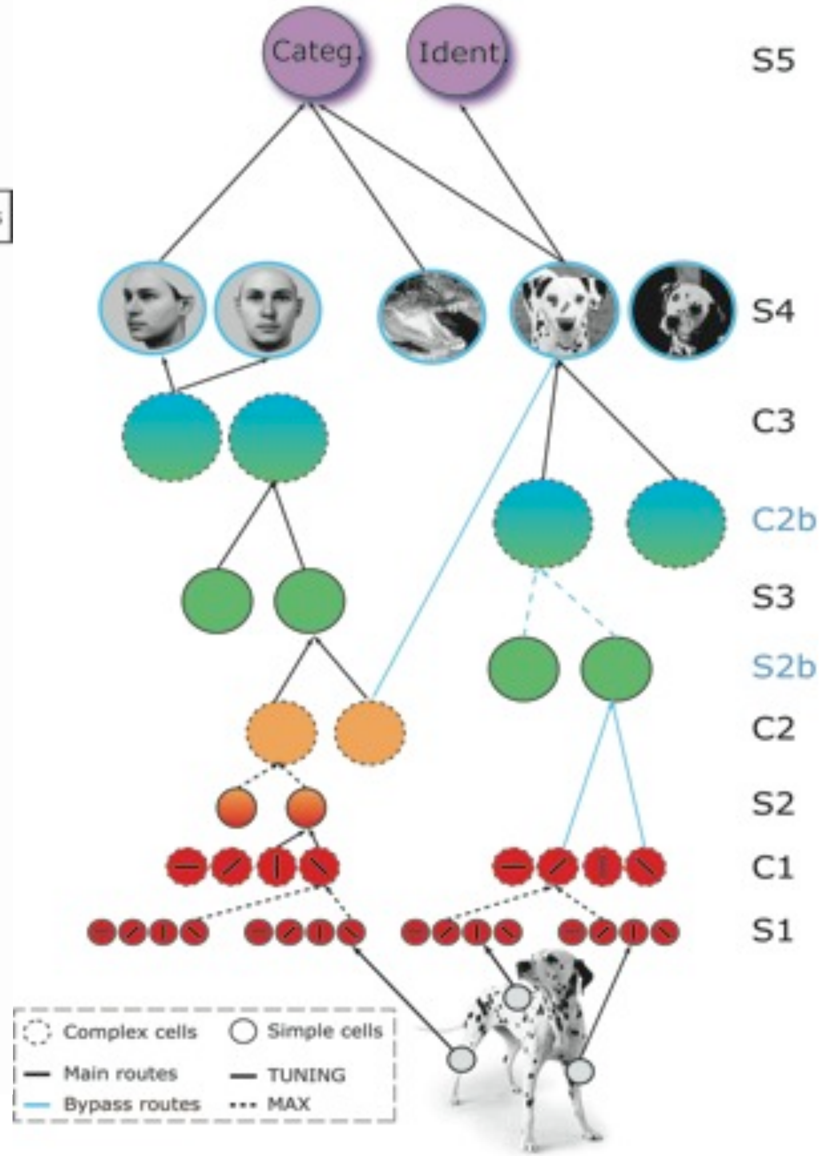


(Hubel & Wiesel 1959)

Recognition in the Ventral Stream: "classical model"

*Modified from (Gross, 1998)

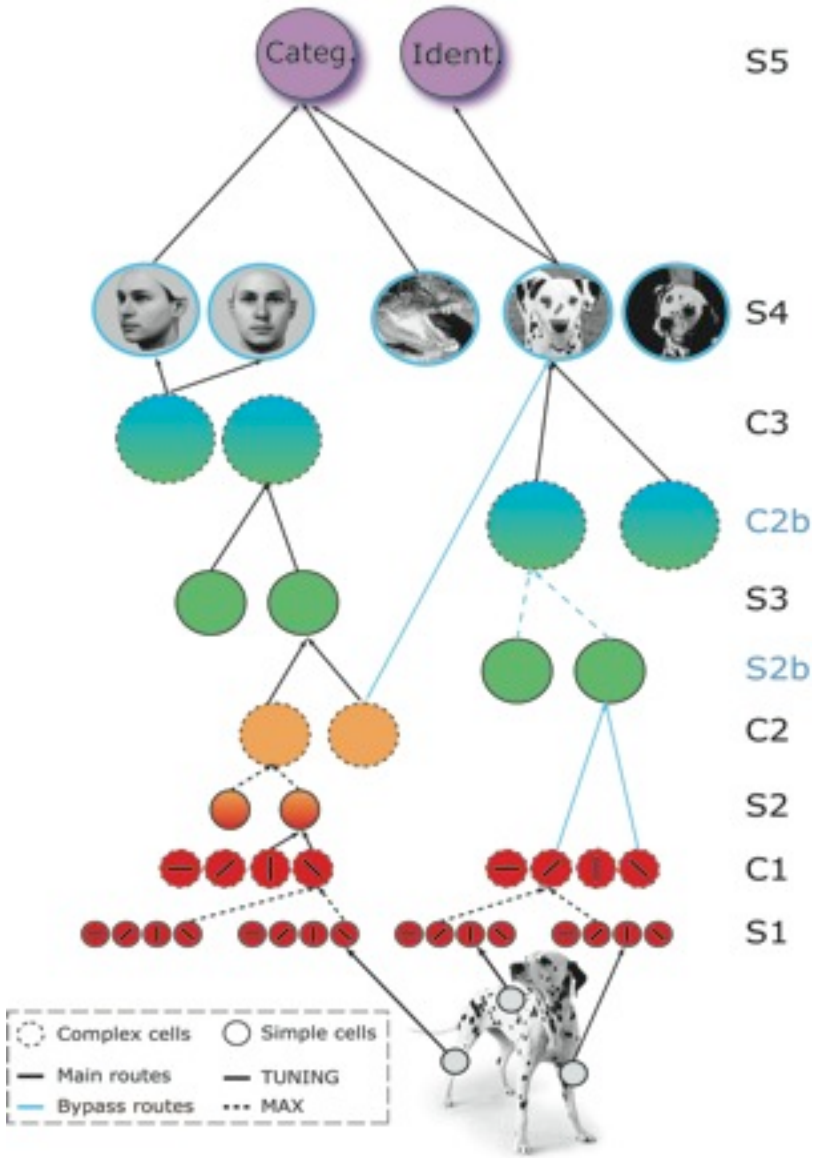




[software available online with CNS (for GPUs)]

Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007

Recognition in Visual Cortex: "classical model"



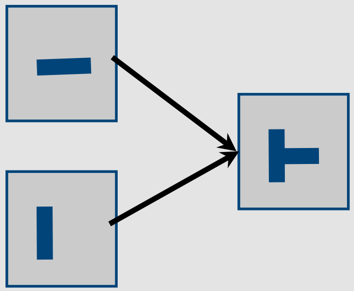
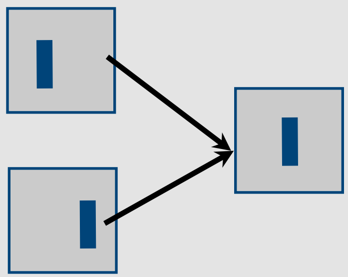
- It is in the family of "Hubel-Wiesel" models (Hubel & Wiesel, 1959: *qual.* **Fukushima**, 1980: *quant*; Oram & Perrett, 1993: *qual*; Wallis & Rolls, 1997; Riesenhuber & Poggio, 1999; Thorpe, 2002; Ullman et al., 2002; Mel, 1997; Wersing and Koerner, 2003; LeCun et al 1998: *not-bio*; Amit & Mascaro, 2003: *not-bio*; Hinton, LeCun, Bengio *not-bio;* Deco & Rolls 2006…)

- As a biological model of object recognition in the ventral stream – from V1 to PFC -- it is *perhaps* the most quantitatively faithful to known neuroscience data

[software available online]

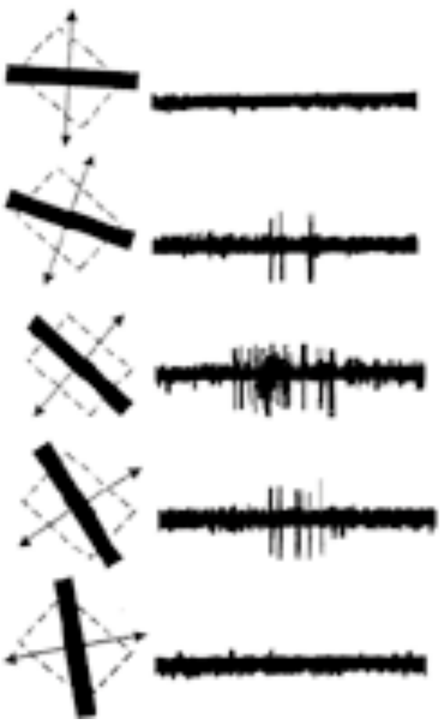
Riesenhuber & Poggio 1999, 2000; Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005; Serre Oliva Poggio 2007

Two key computations, suggested by physiology

Unit	Pooling	Computation	Operation
Simple		Selectivity / template matching	Gaussian-tuning / AND-like
Complex		Invariance	Soft-max / or-like

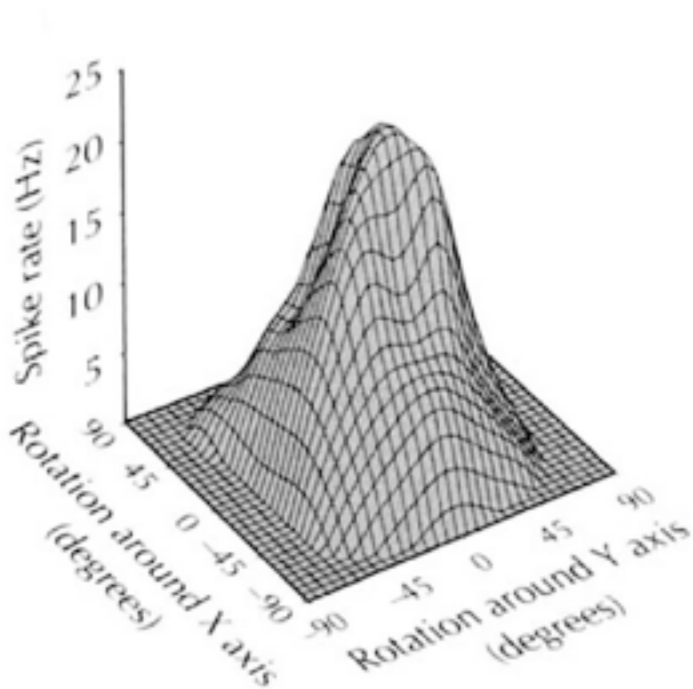
Gaussian tuning

Gaussian tuning in V1 for orientation



Hubel & Wiesel 1958

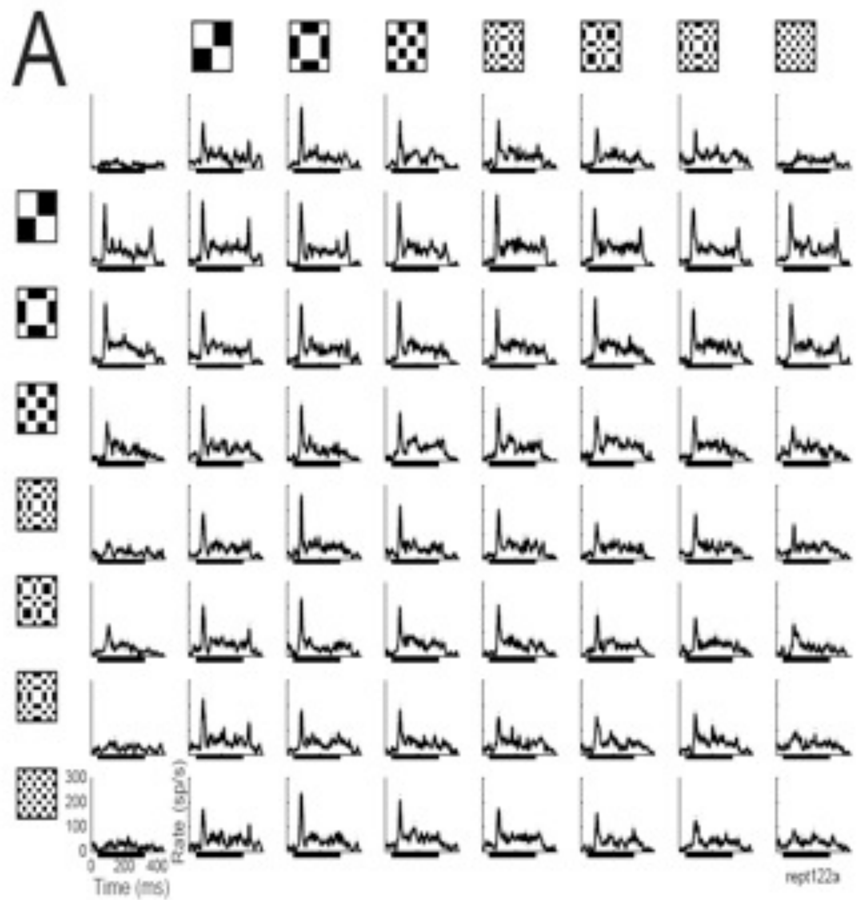
Gaussian tuning in IT around 3D views



Logothetis Pauls & Poggio 1995

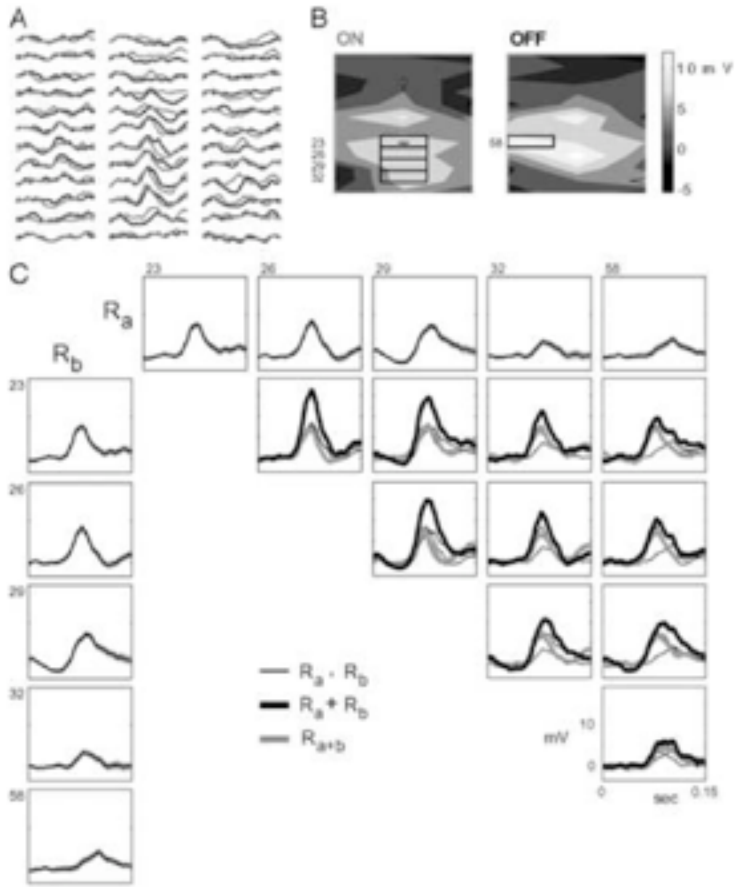
Max-like operation

Max-like behavior in V4



Gawne & Martin 2002

Max-like behavior in V1



Lampl Ferster Poggio & Riesenhuber 2004
see also Finn Prieber & Ferster 2007

Two operations (~OR, ~AND): disjunctions of conjunctions

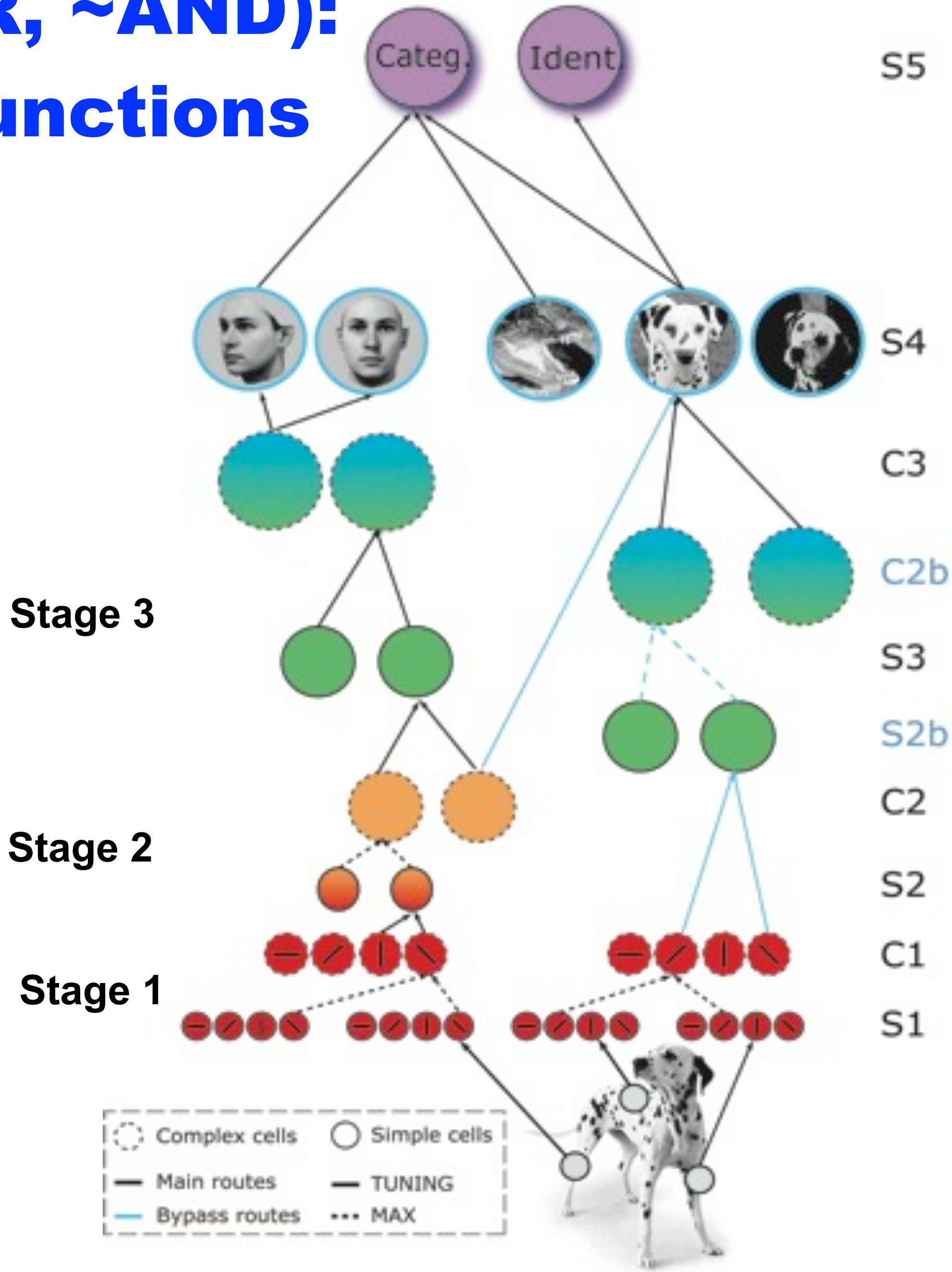
- Tuning operation (Gaussian-like, AND-like) $y = e^{-|x-w|^2}$ *or* $y \sim \frac{x \bullet w}{|x|}$
- Simple units

- Max-like operation (OR-like)

$$y = \max\{x1, x2, \ldots\}$$

- Complex units

Each operation ~microcircuits of ~100 neurons



Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like)

$$y = e^{-|x-w|^2}$$

or

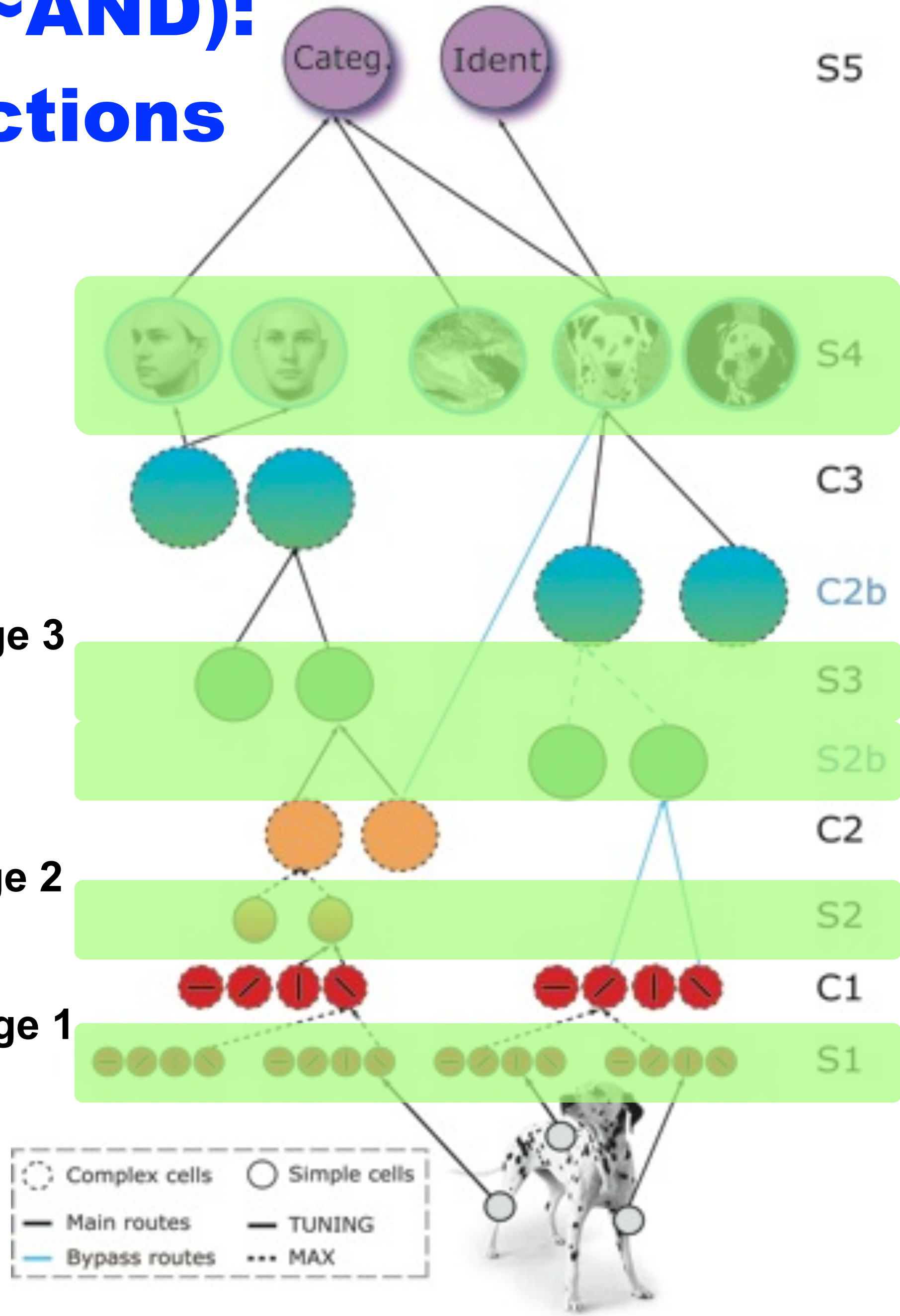
$$y \sim \frac{x \bullet w}{|x|}$$

- Simple units

- Max-like operation (OR-like)

$$y = \max\{x1, x2, \ldots\}$$

- Complex units

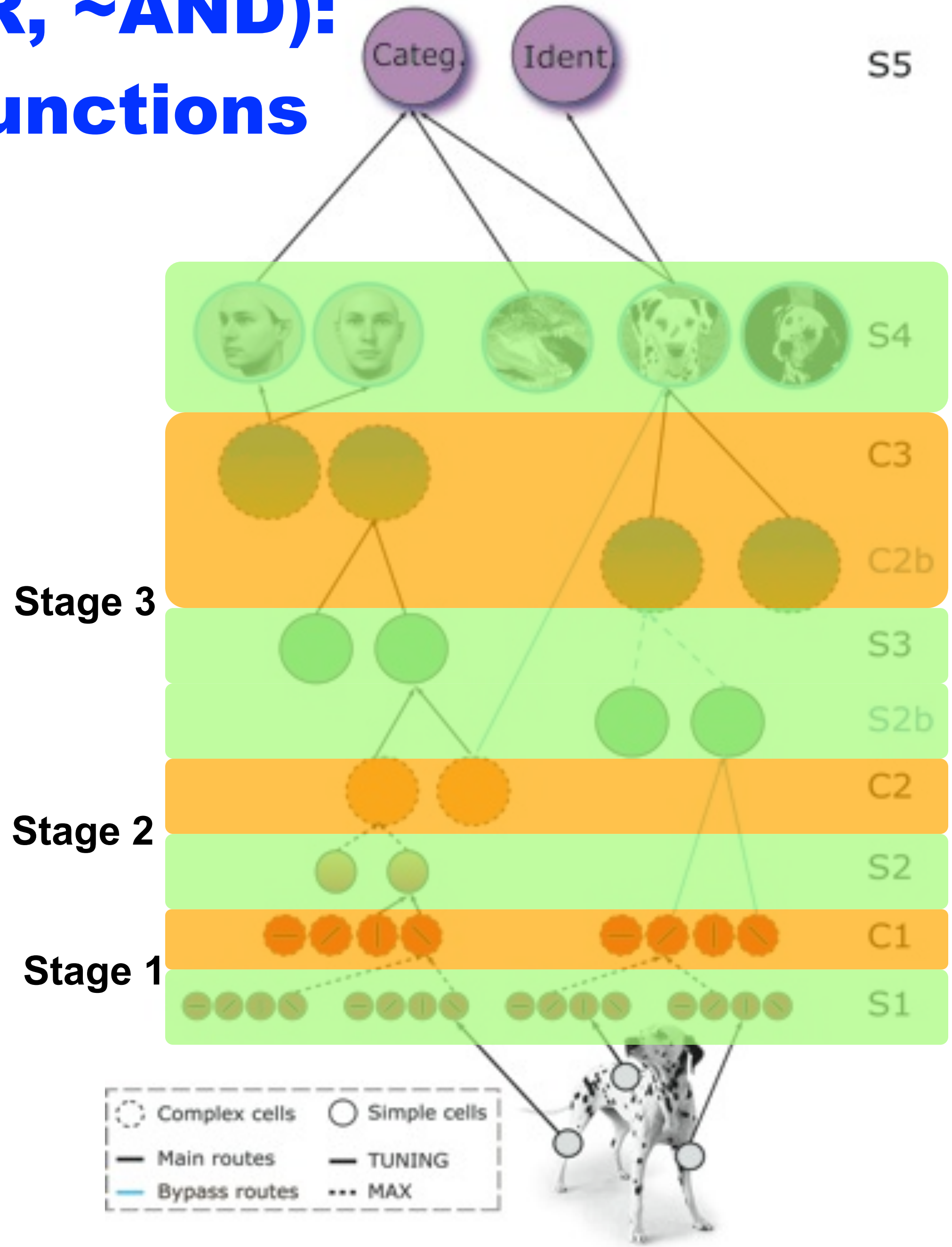


Each operation ~microcircuits of ~100 neurons

Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like) $$y = e^{-|x-w|^2}$$ *or* $$y \sim \frac{x \bullet w}{|x|}$$
- Simple units

- Max-like operation (OR-like) $$y = \max\{x1, x2, \ldots\}$$
- Complex units

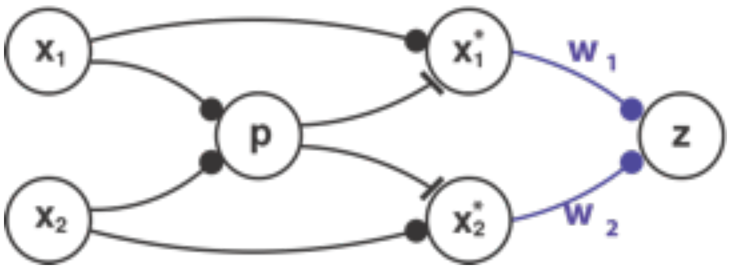


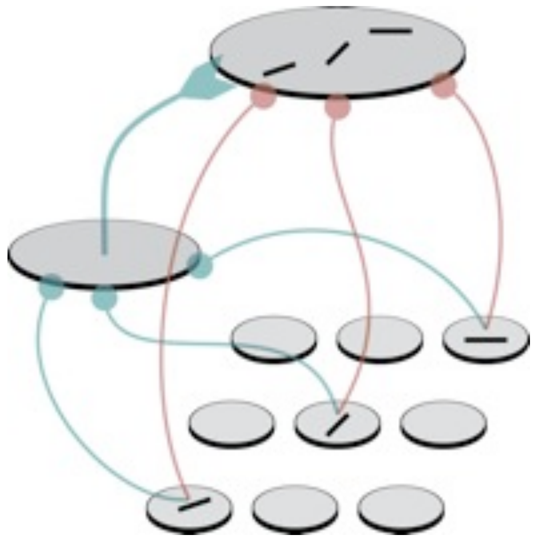
Each operation ~microcircuits of ~100 neurons

Plausible biophysical implementations

- Max and Gaussian-like tuning can be approximated with same canonical circuit using shunting inhibition. Tuning (eg "center" of the Gaussian) corresponds to synaptic weights.

$$y = \frac{\sum_{j=1}^{n} w_j^* \, x_j^p}{k + \left(\sum_{j=1}^{n} x_j^q \right)^r},$$

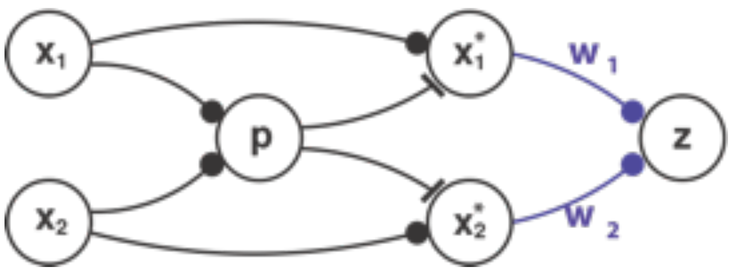


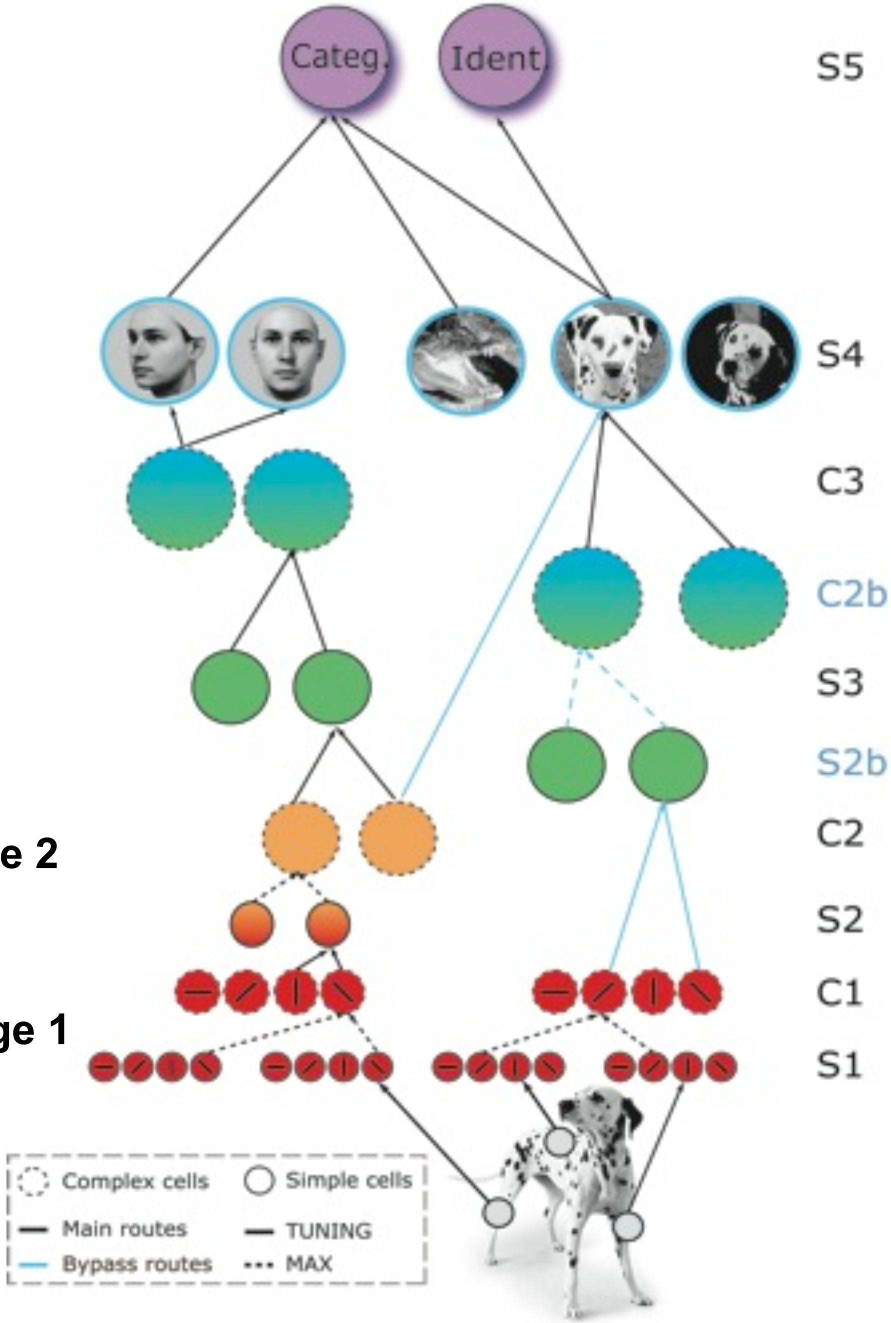


(Knoblich Koch Poggio in prep; Kouh & Poggio 2007; Knoblich Bouvrie Poggio 2007)

Recognition in Visual Cortex: circuits and biophysics

A canonical microcircuit of spiking neurons?

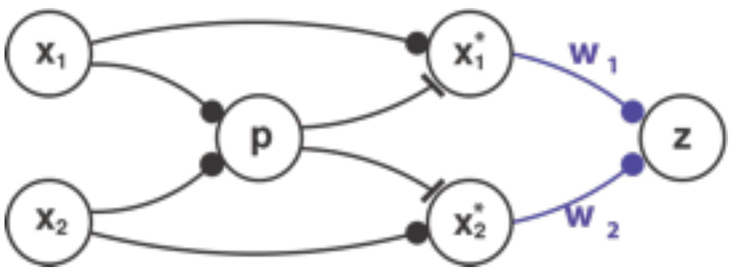


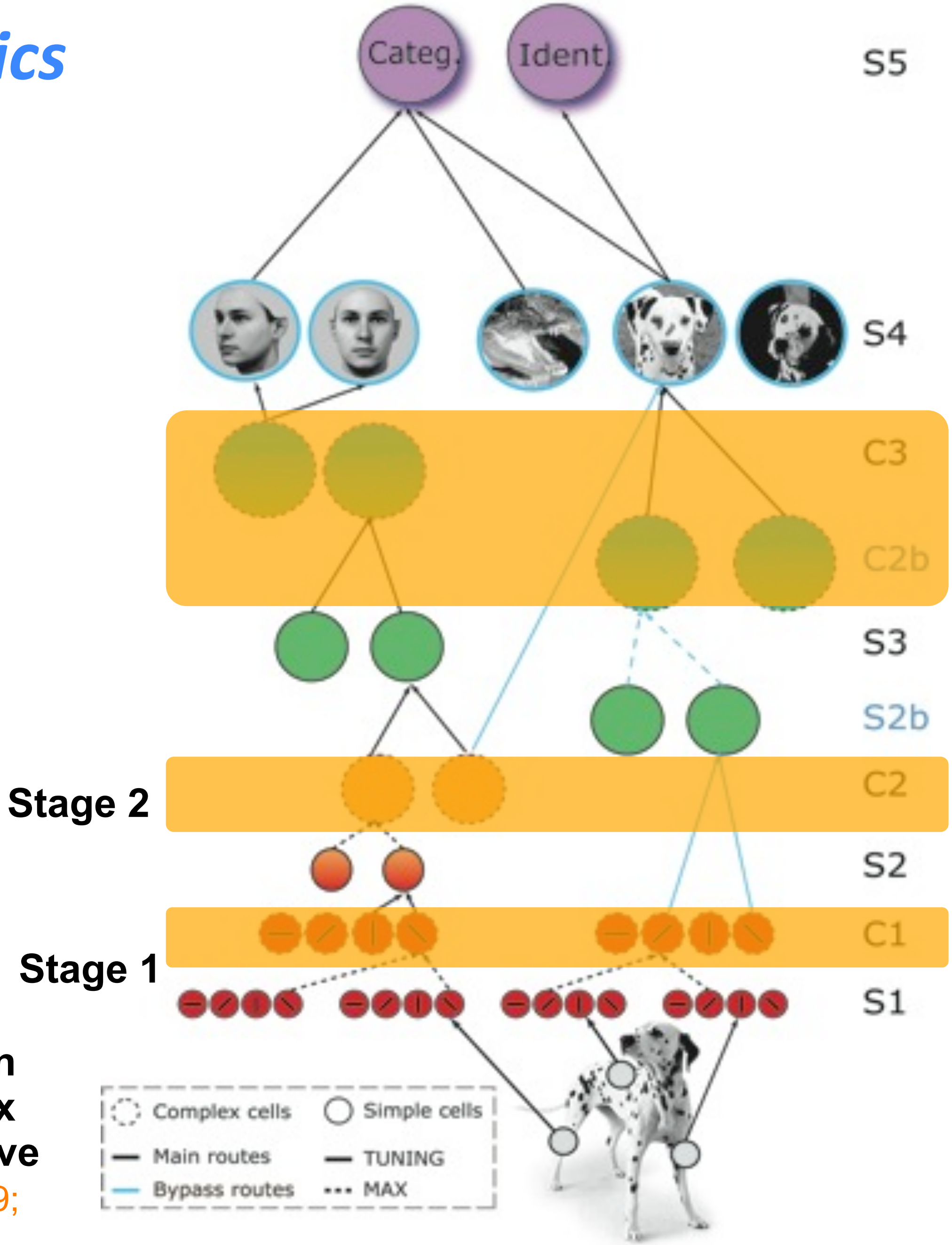


A plausible biophysical implementation for *both* Gaussian tuning (~AND) + max (~OR): normalization circuits with divisive inhibition (Kouh, Poggio, 2008; also RP, 1999; Heeger, Carandini, Simoncelli,…)

Recognition in Visual Cortex: circuits and biophysics

A canonical microcircuit of spiking neurons?

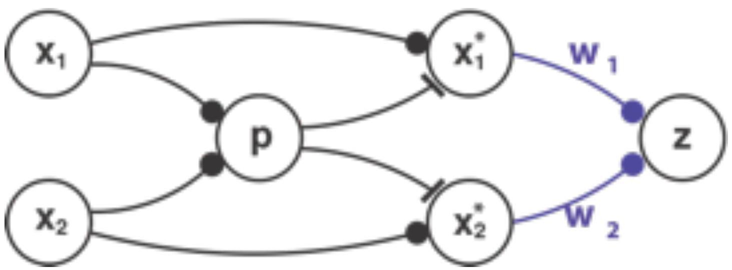


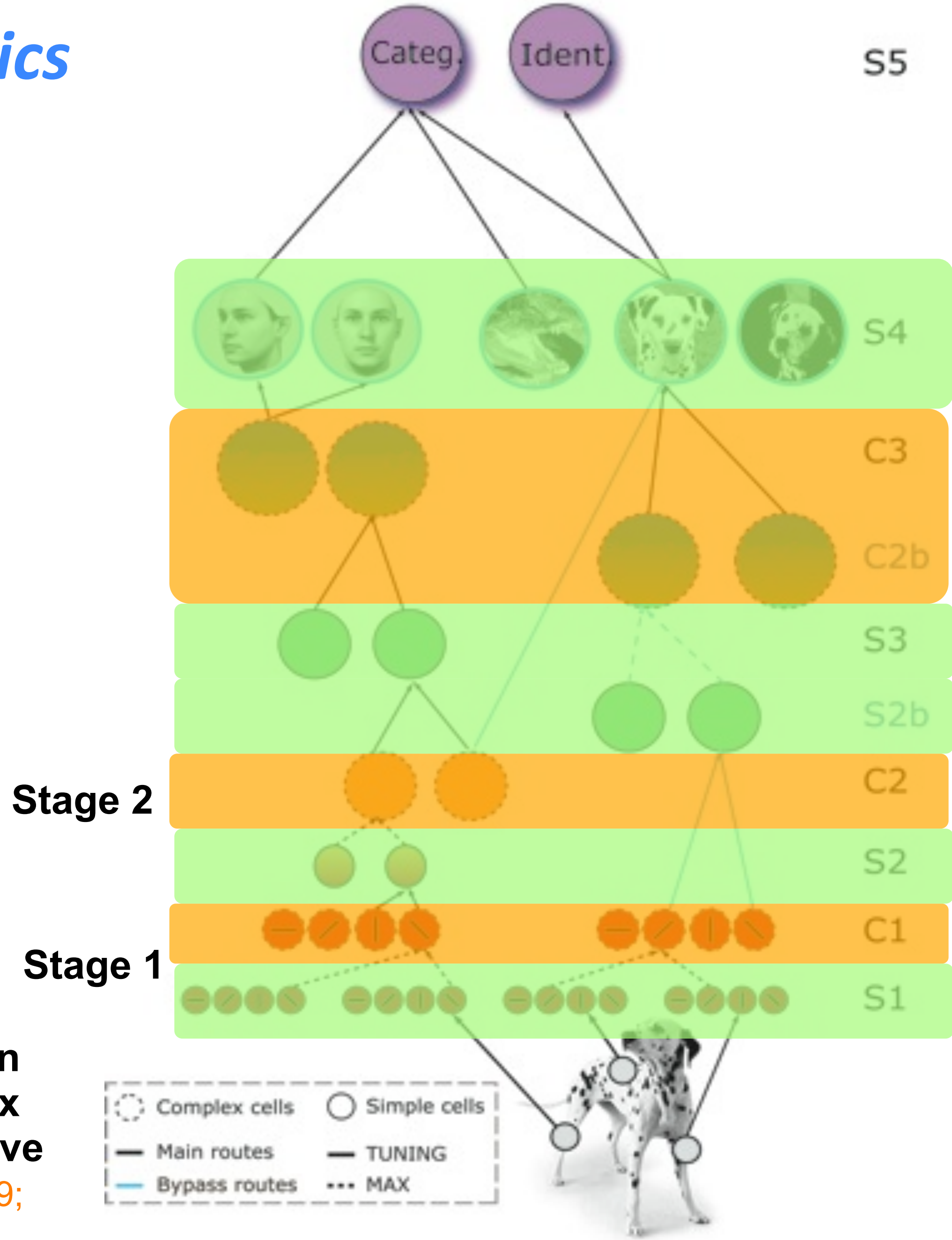


A plausible biophysical implementation for *both* Gaussian tuning (~AND) + max (~OR): normalization circuits with divisive inhibition (Kouh, Poggio, 2008; also RP, 1999; Heeger, Carandini, Simoncelli,…)

Recognition in Visual Cortex: circuits and biophysics

A canonical microcircuit of spiking neurons?





A plausible biophysical implementation for *both* Gaussian tuning (~AND) + max (~OR): normalization circuits with divisive inhibition (Kouh, Poggio, 2008; also RP, 1999; Heeger, Carandini, Simoncelli,…)

Simulation with spiking neurons and realistic synapses

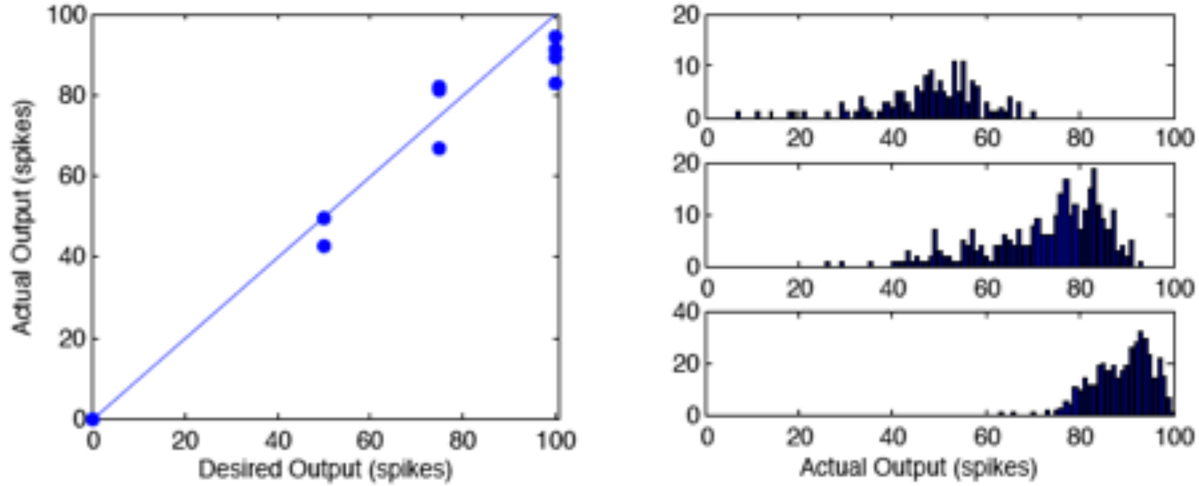


Figure 3: Mean response of max circuit depicted in Fig. 2 over 50 runs for all possible combinations of 0, 50, 75 and 100 spikes per input packet, plotted against the desired (true) maximum of the inputs (left). Histogram of all outputs (spike count in output packet) for three cases (right). The true maximum of the inputs is 50, 75 and 100 spikes, respectively (top to bottom).

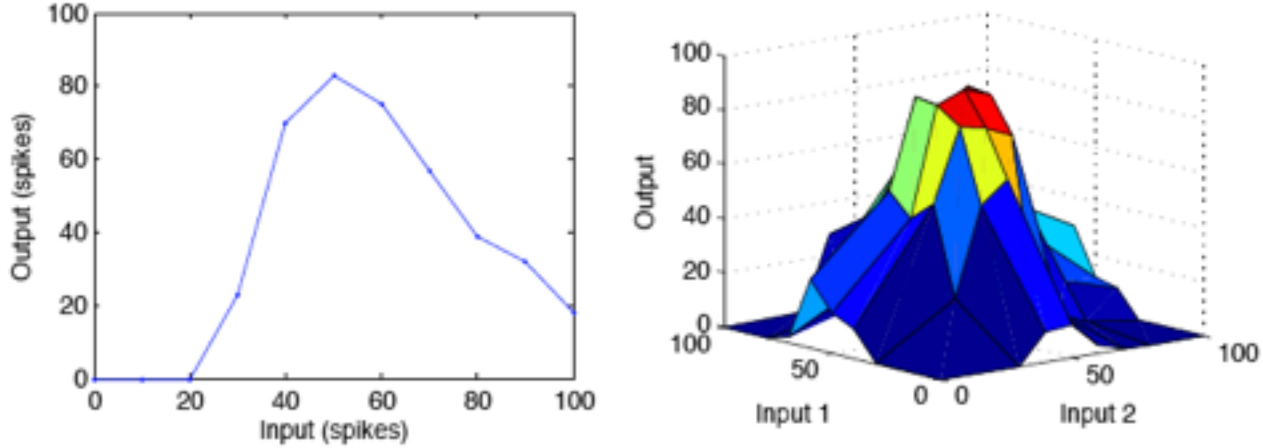


Figure 4: Output (spike count in output packet) of a one-dimensional Gaussian-like tuning circuit tuned to 50 a spike packet input (left). Output (spike count in output packet) of the two-dimensional tuning circuit depicted in Fig. 2 tuned to the combination of two 50 spike packet inputs (right).

Basic circuit is closely related to other models

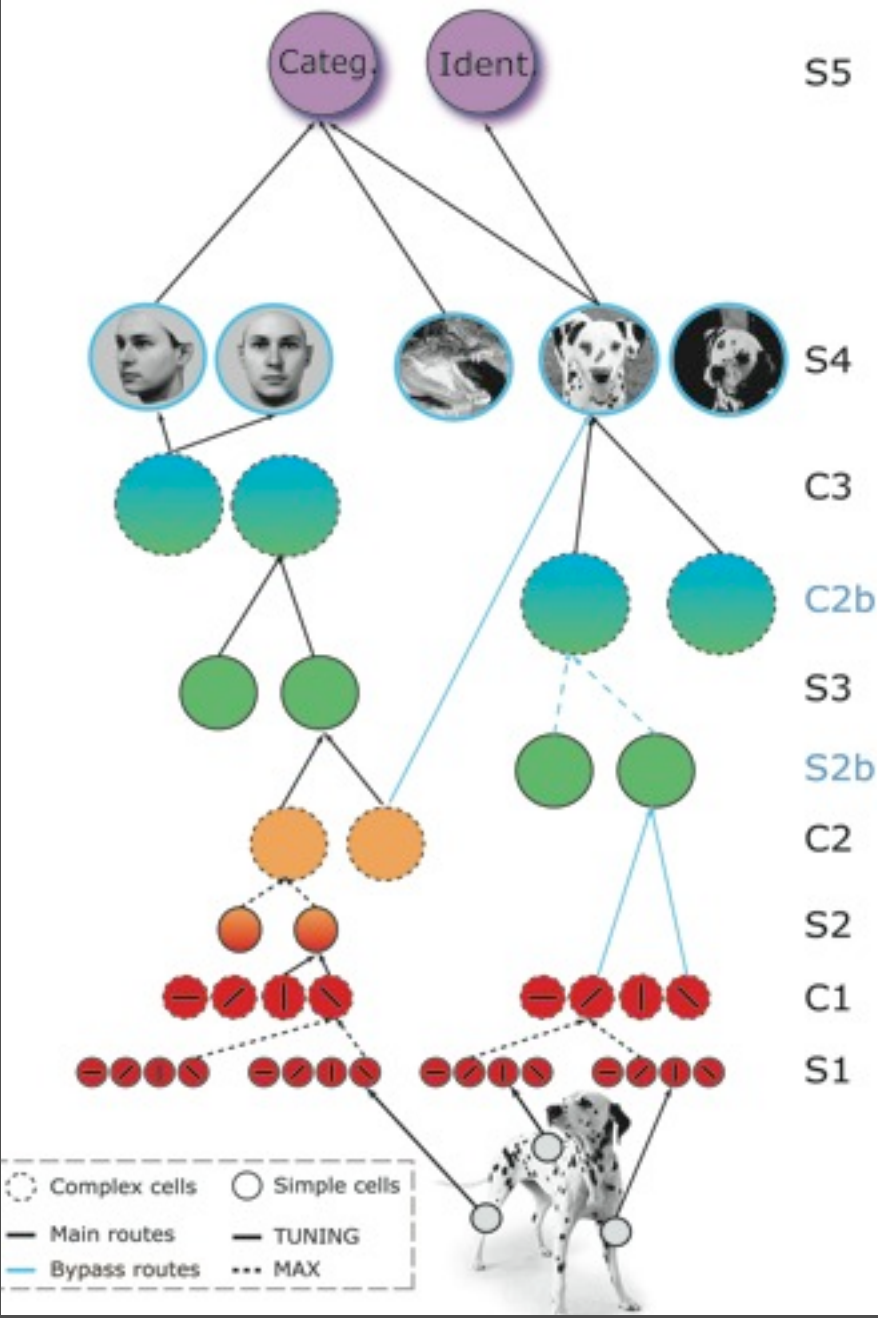
Operation	(Steady-State) Output	
Canonical	$y = \dfrac{\sum_{i=1}^{n} w_i \, x_i{}^{p}}{k + \left(\sum_{i=1}^{n} x_i{}^{q} \right)^{r}}$	(1)
Energy Model	$y = \sum_{i=1}^{2} x_i{}^{2}$	(2)
Gaussian-like	$y = \dfrac{\sum_{i=1}^{n} w_i \, x_i}{k + \sum_{i=1}^{n} x_i{}^{2}}$	(4)
Max-like	$y = \dfrac{\sum_{i=1}^{n} x_i{}^{3}}{k + \sum_{i=1}^{n} x_i{}^{2}}$	(5)

Can be implemented by shunting inhibition (Grossberg 1973, Reichardt et al. 1983, Carandini and Heeger, 1994) and spike threshold variability (Anderson et al. 2000, Miller and Troyer, 2002)

Adelson and Bergen (see also Hassenstein and Reichardt, 1956)

Of the same form as model of MT (Rust et al., Nature Neuroscience, 2007

Recognition in Visual Cortex: learning



Task-specific circuits (from IT to PFC?)

- <u>Supervised</u> learning: ~ classifier

Overcomplete dictionary of "templates" ~ image "patches" ~ ~ "parts" is learned during an **<u>unsupervised</u>** learning stage (from ~10,000 natural images) by tuning S units.

see also (Foldiak 1991; Perrett et al 1984; Wallis & Rolls, 1997; Lewicki and Olshausen, 1999; Einhauser et al 2002; Wiskott & Sejnowski 2002; Spratling 2005)

Recognition in Visual Cortex: learning (from Serre, 2007)

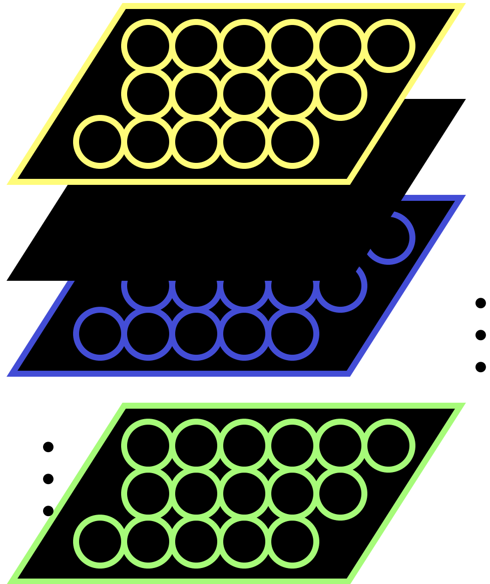
Start with S2 layer

Start with S2 layer

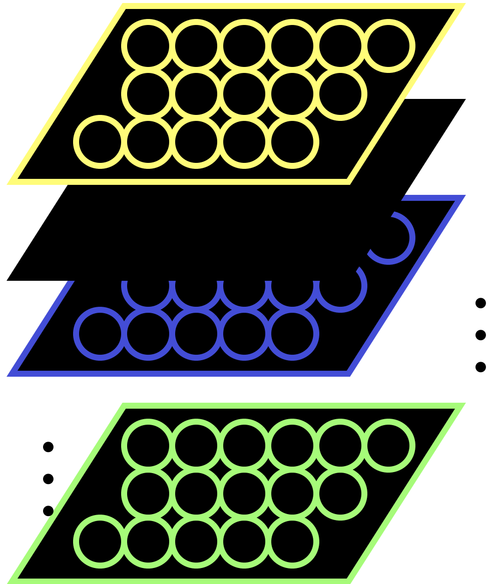
Units are organized in n
feature maps

Start with S2 layer

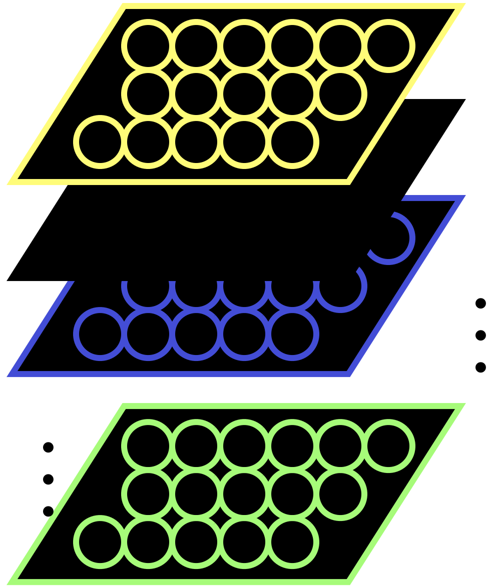
Units are organized in n feature maps



Start with S2 layer

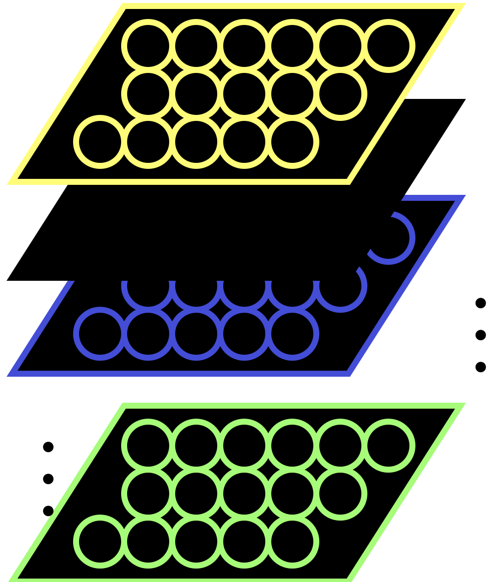


Start with S2 layer

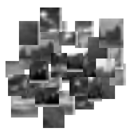


Database ~1,000 natural images

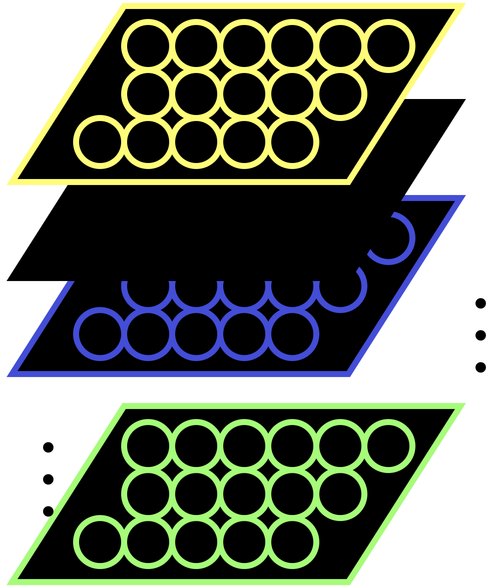
Start with S2 layer



Database ~1,000 natural images

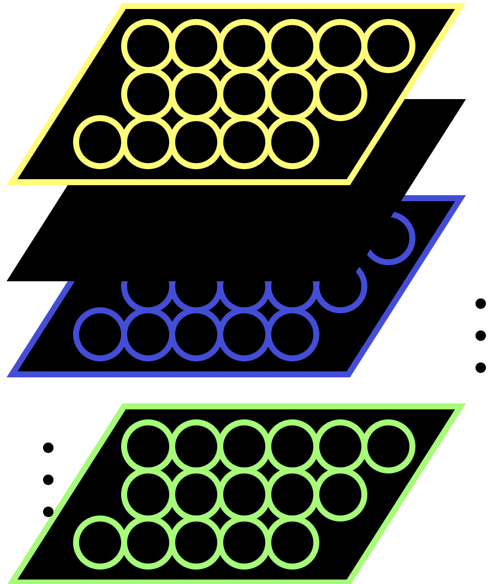


Start with S2 layer



Database ~1,000 natural images

Start with S2 layer

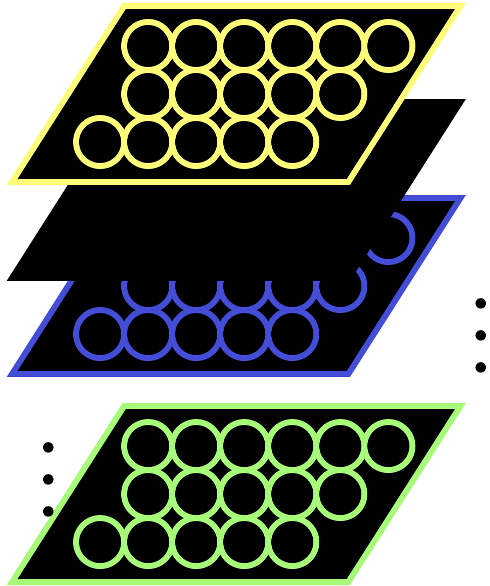


Database ~1,000 natural images

At each iteration:

- Present one image
- Learn k feature maps

Start with S2 layer

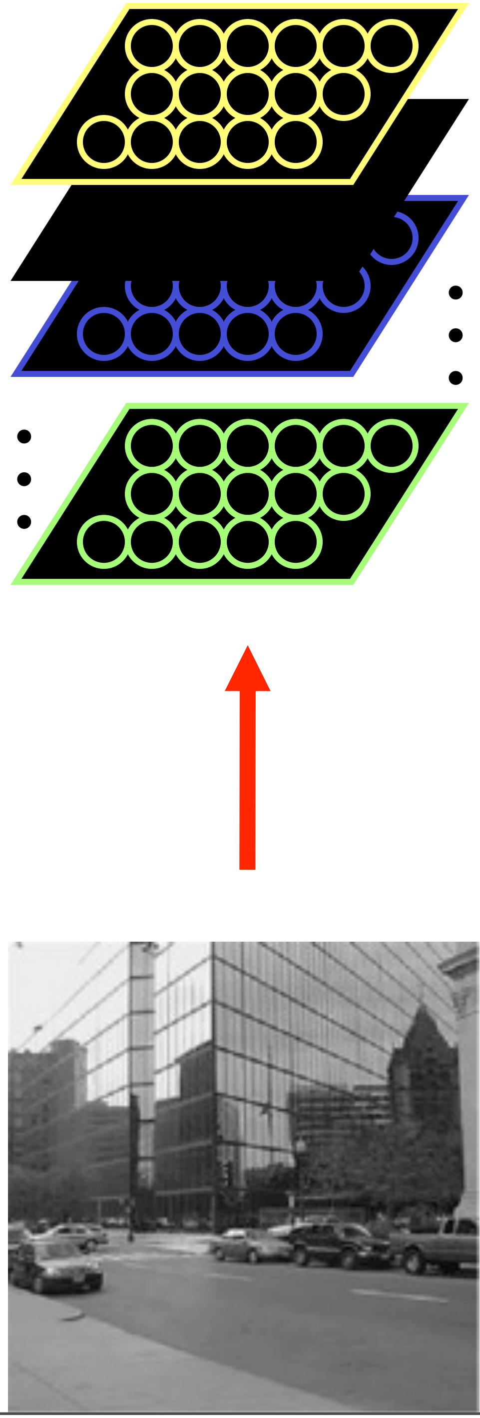


Database ~1,000 natural images

At each iteration:

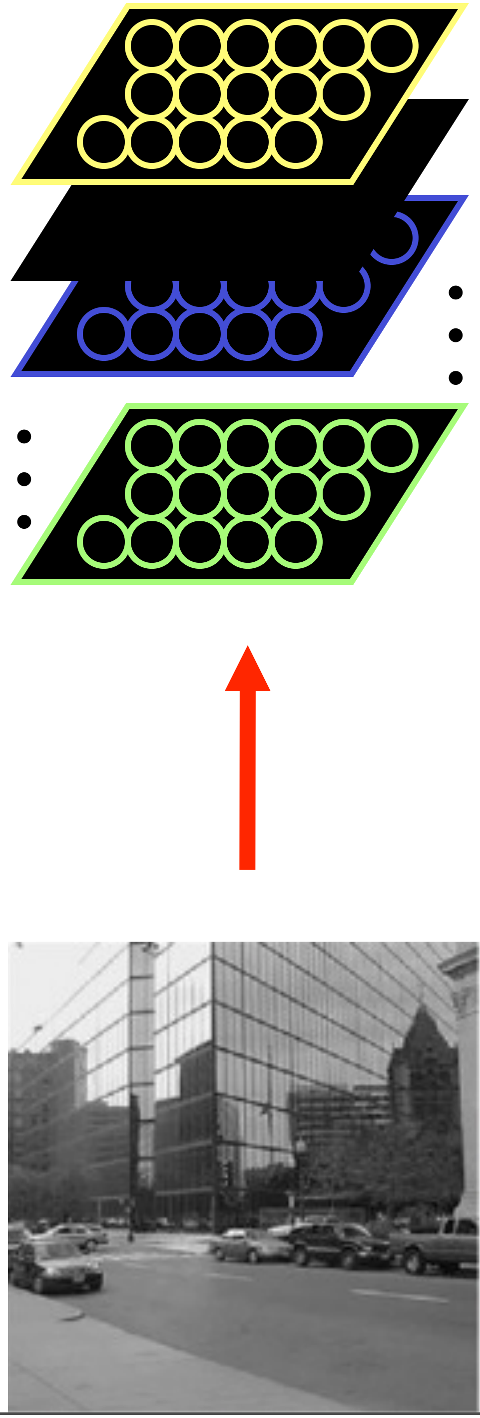
- ➢ Present one image
- ➢ Learn k feature maps

Start with S2 layer



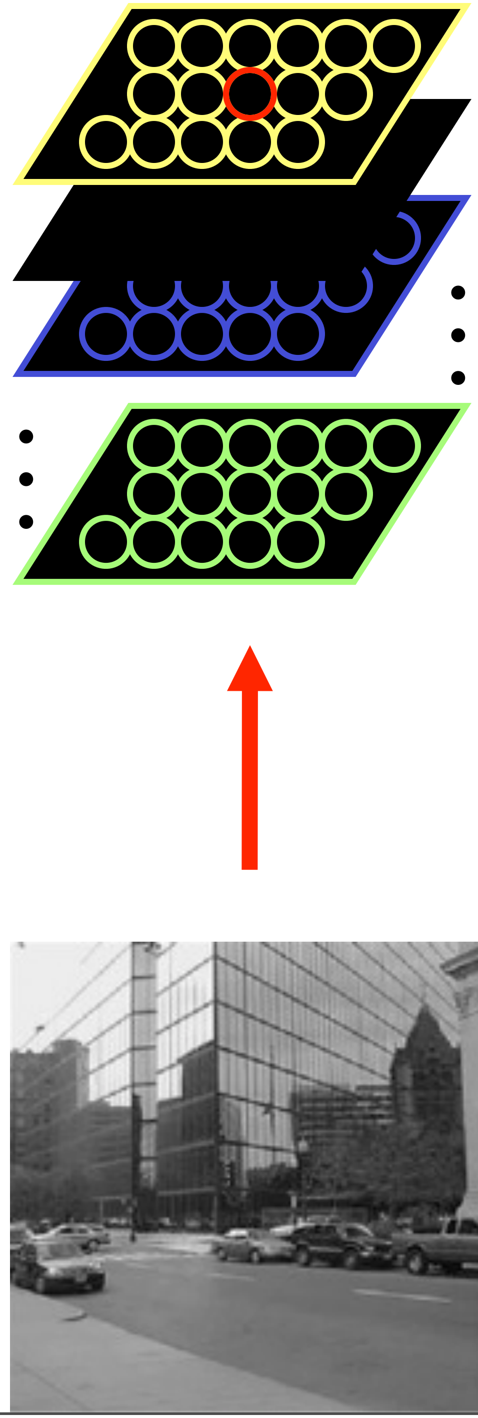
Start with S2 layer

Pick 1 unit from the first map at random



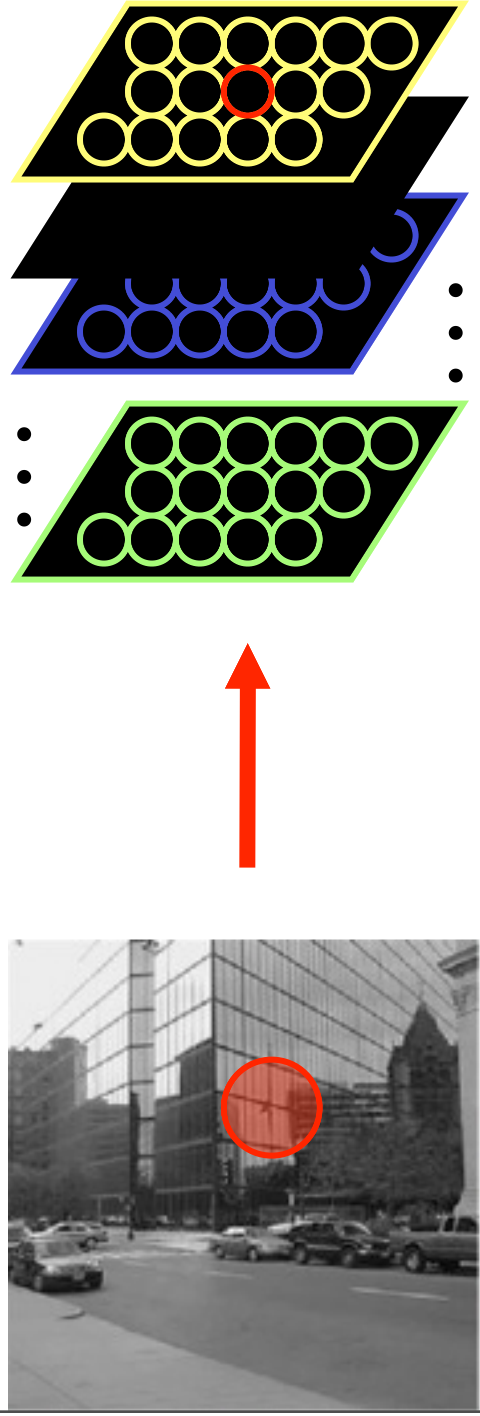
Start with S2 layer

Pick 1 unit from the first map at random



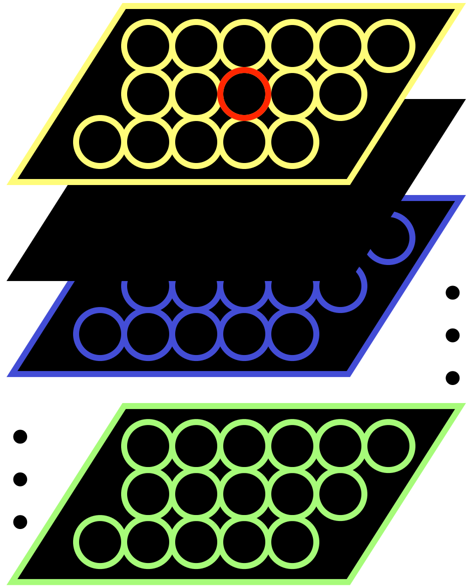
Start with S2 layer

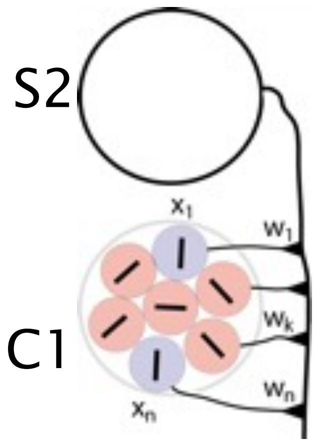
Pick 1 unit from the first map at random

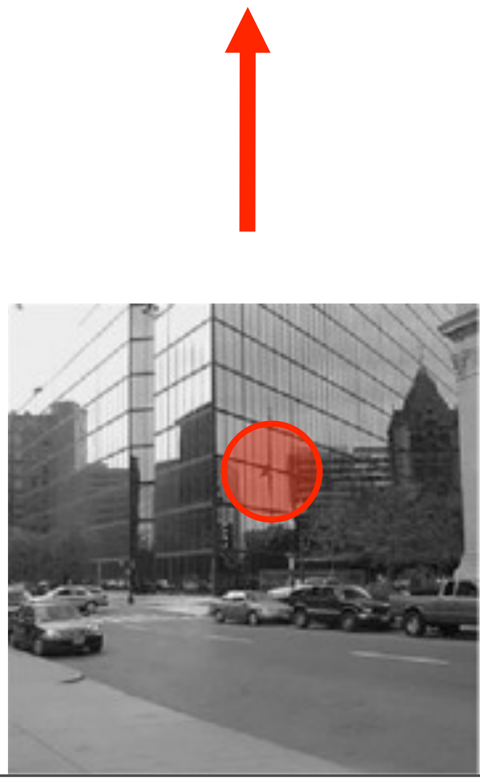


Start with S2 layer

Pick 1 unit from the first map at random



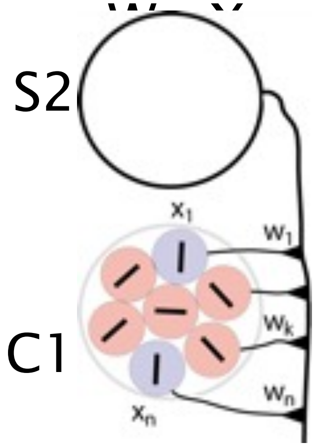


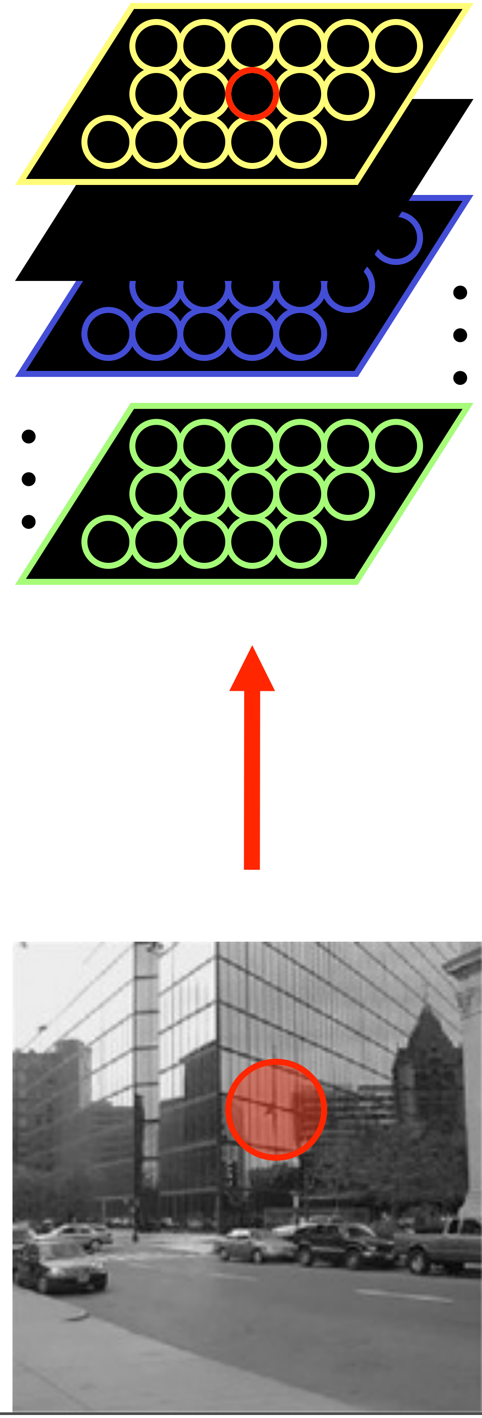


Start with S2 layer

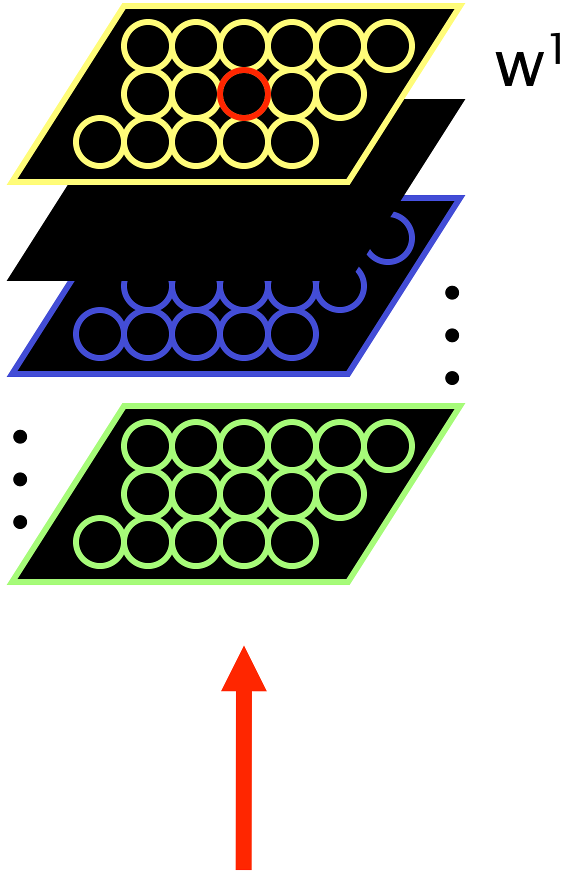
Pick 1 unit from the first map at random

Store in unit synaptic weights the precise pattern of subunits activity, i.e.





Start with S2 layer



Pick 1 unit from the first map at random

Store in unit synaptic weights the precise pattern of subunits activity, i.e.

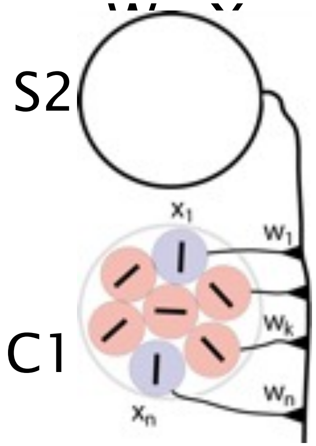
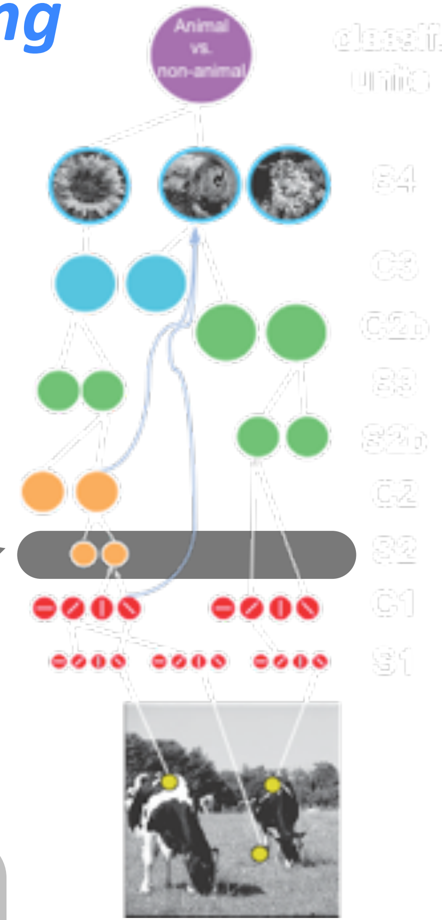


Image “moves” (looming and shifting)

Weight vector w is copied to all units in feature map 1 (across positions and scales)

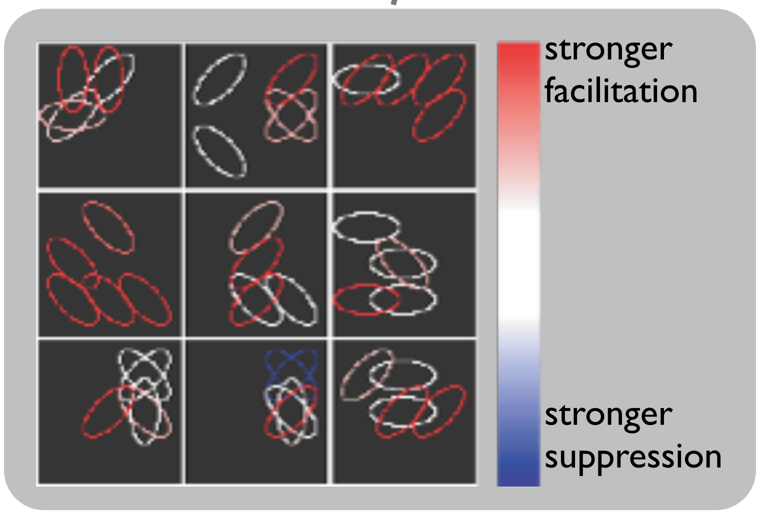
Recognition in Visual Cortex: learning



S2 units

- Features of moderate complexity (n~1,000 types)
- Combination of V1-like complex units at different orientations

- Synaptic weights ***w*** learned from natural images
- 5-10 subunits chosen at random from all possible afferents (~100-1,000)



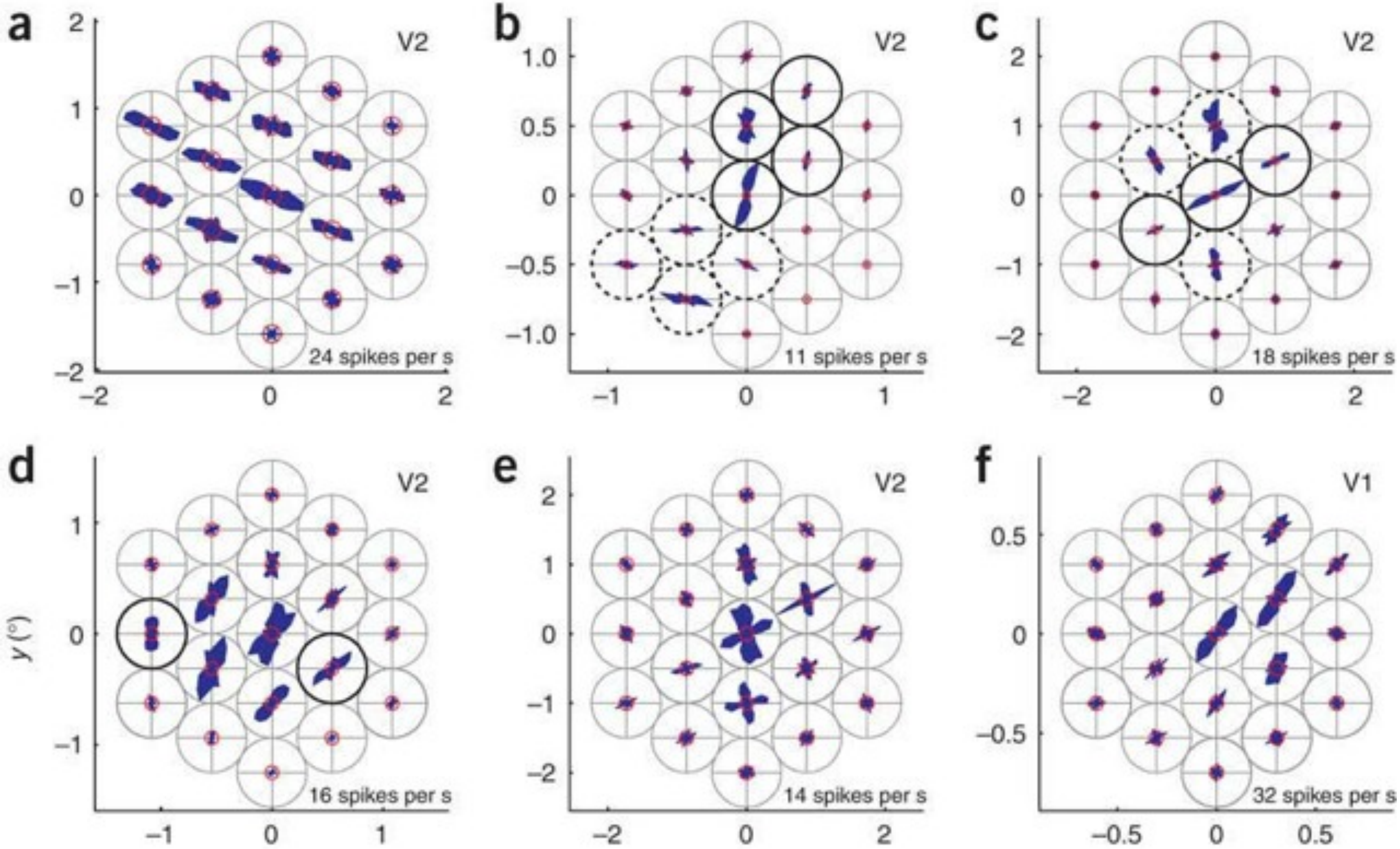
Recognition in Visual Cortex: learning

Sample S2 Units Learned *(from Serre, 2007)*

Nature Neuroscience - 10, 1313 - 1321 (2007) / Published online: 16 September 2007 | doi:10.1038/nn1975

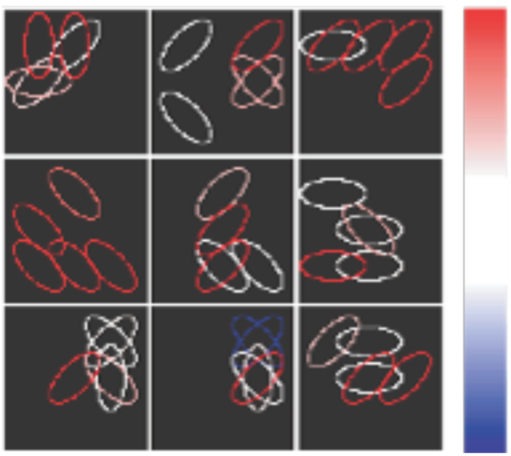
Neurons in monkey visual area V2 encode combinations of orientations

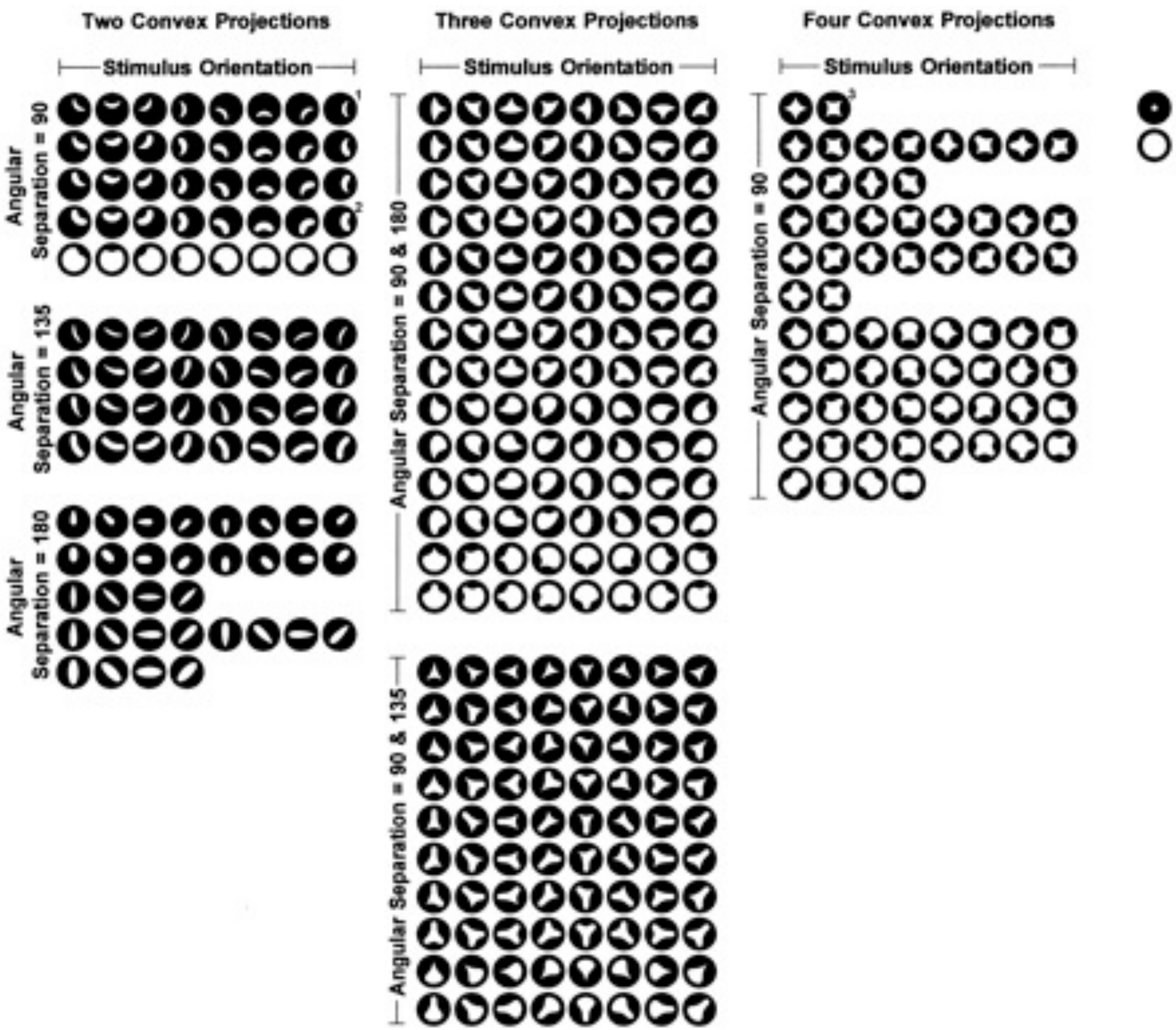
Akiyuki Anzai, Xinmiao Peng & David C Van Essen



Comparison w/ V4

Tuning for curvature and boundary conformations?



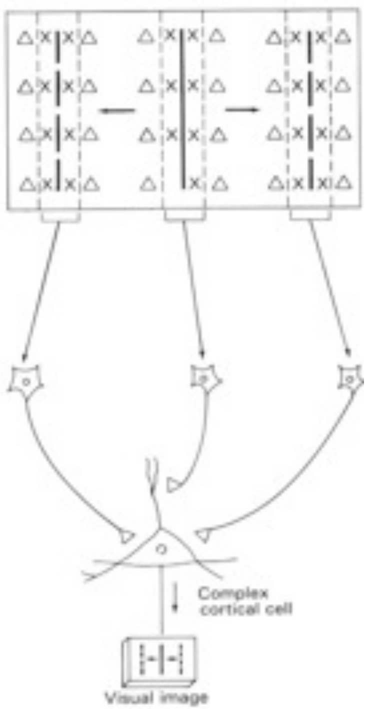


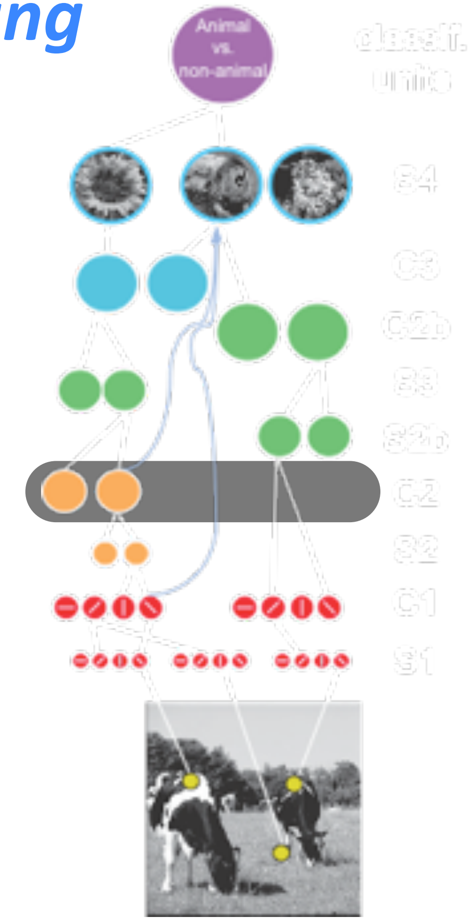
Pasupathy & Connor 2001

Recognition in Visual Cortex: learning

C2 units

- Same selectivity as S2 units but increased tolerance to position and size of preferred stimulus
- Local pooling over S2 units with same selectivity but different positions and scales
- A prediction to be tested: S2 units in V2 and C2 units in V4?

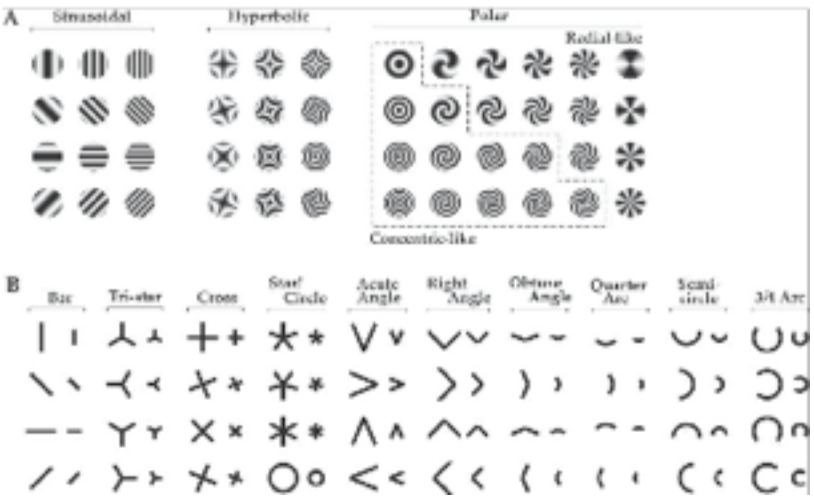


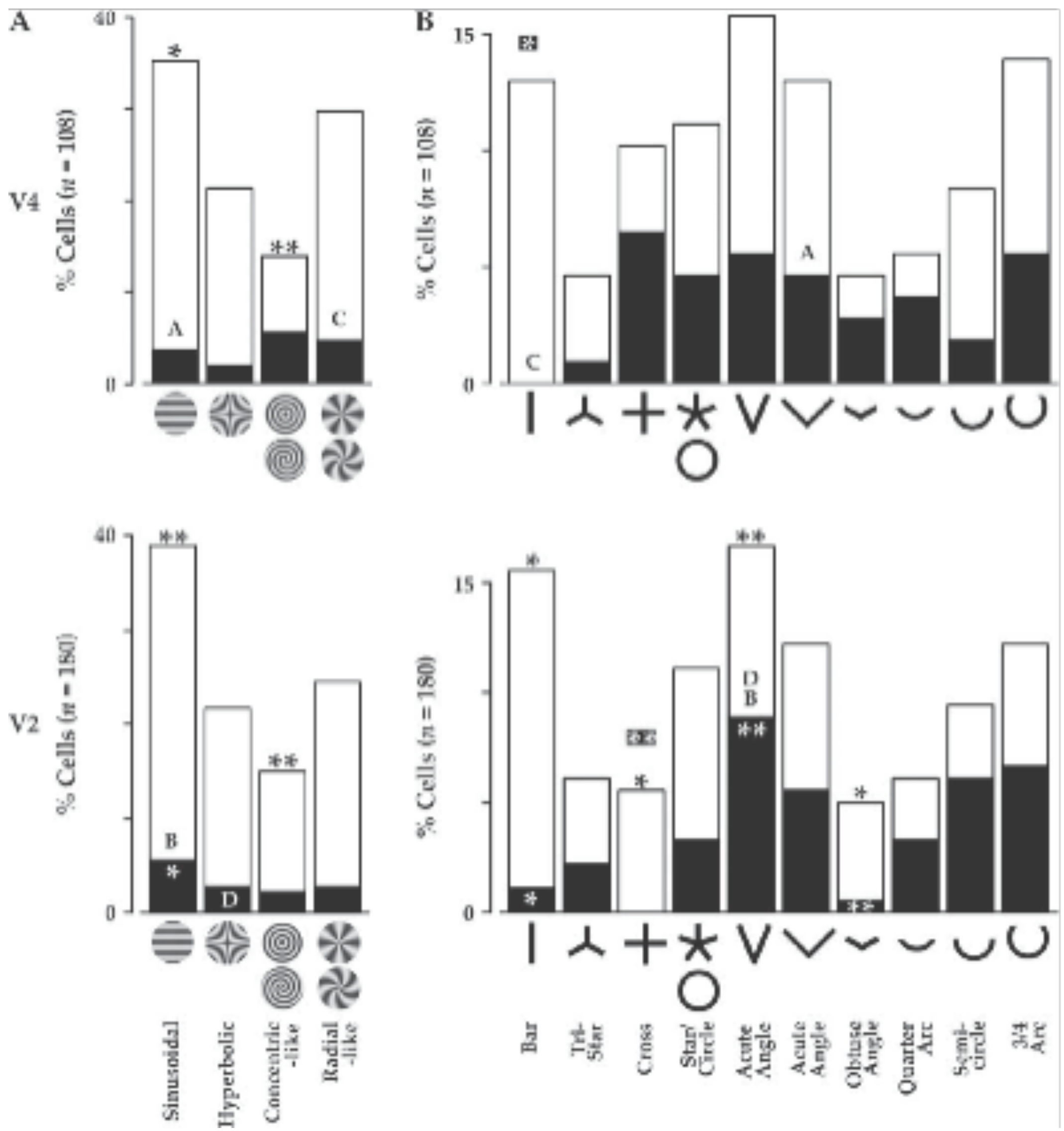


Cerebral Cortex Advance Access published online on June 19, 2006

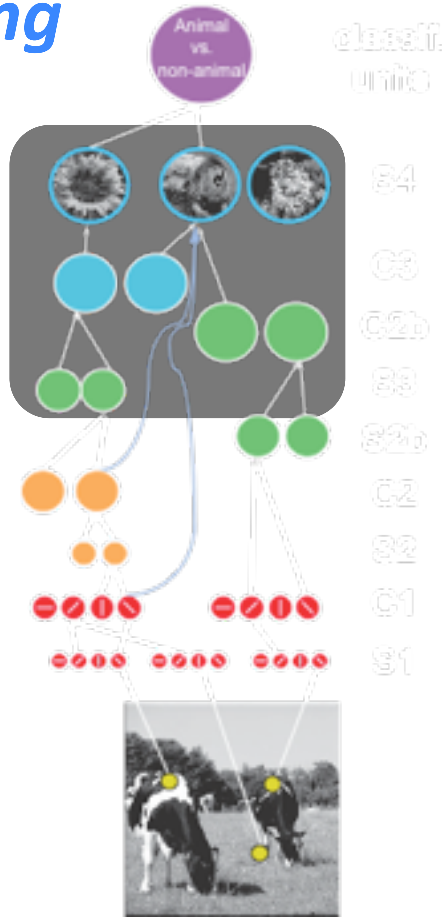
A Comparative Study of Shape Representation in Macaque Visual Areas V2 and V4

Jay Hegdé and David C. Van Essen



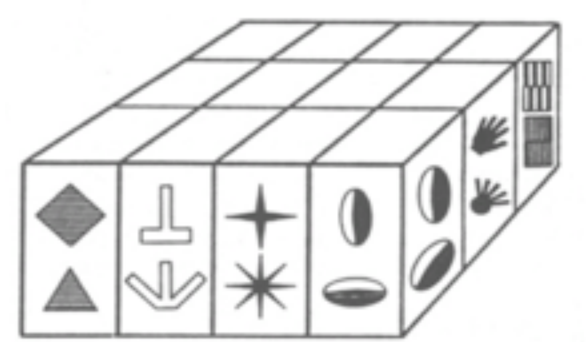


Recognition in Visual Cortex: learning

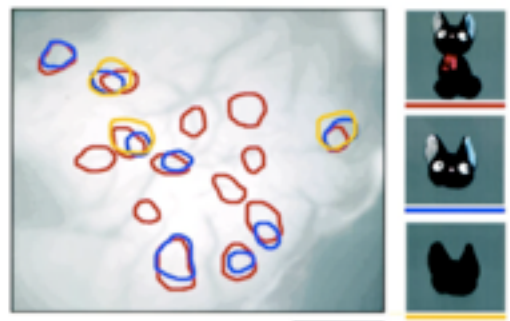


Beyond C2 units

- Units increasingly complex and invariant
- S3/C3 units:
  - Combination of V4-like units with different selectivities
  - Dictionary of ~1,000 features = num. columns in IT (Fujita 1992)



Tanaka et al.



Tsunoda et al.

A loose hierarchy

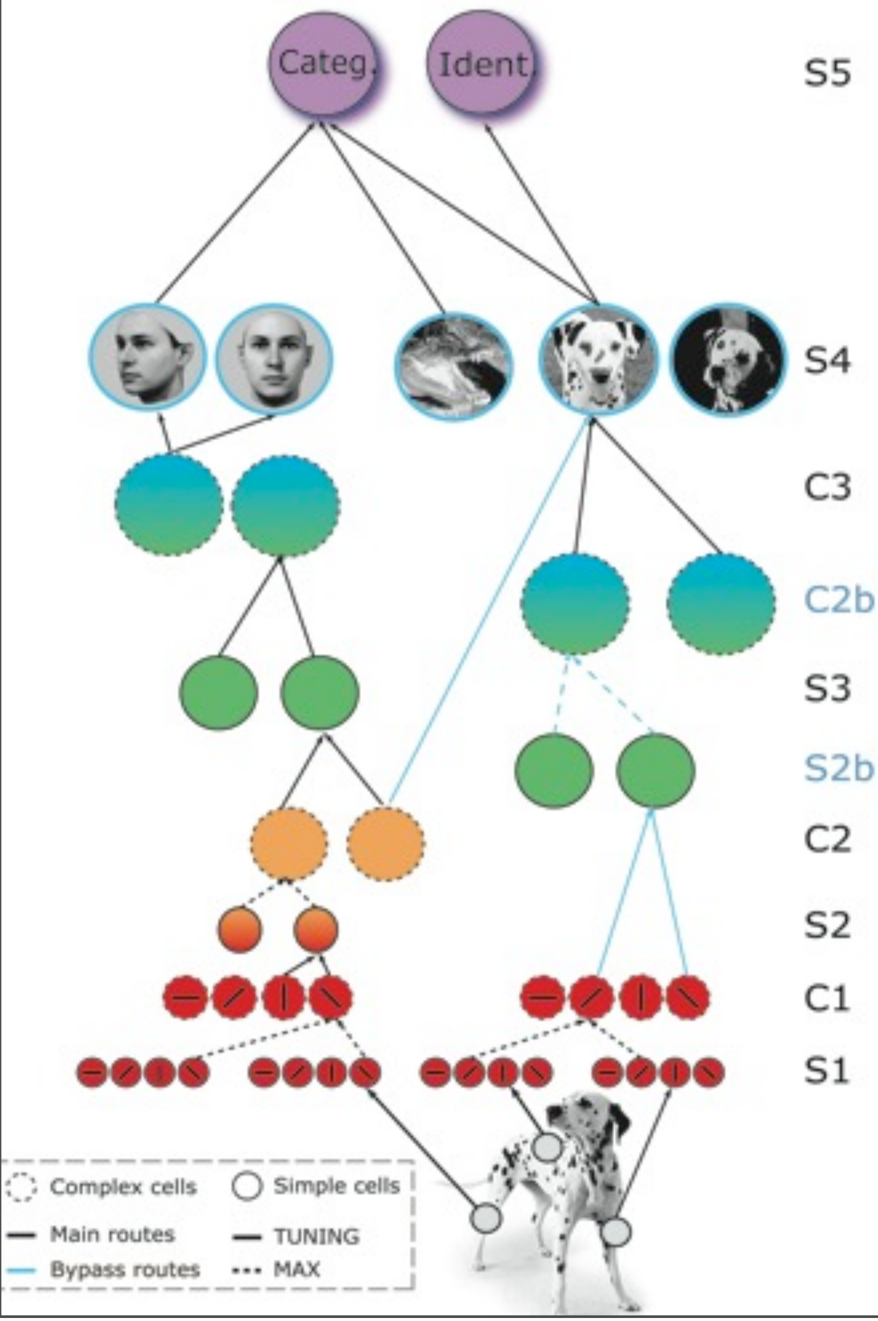
- Bypass routes along with main routes:
  - From V2 to TEO (bypassing V4) (Morel & Bullier 1990; Baizer et al 1991; Distler et al 1991; Weller & Steele 1992; Nakamura et al 1993; Buffalo et al 2005)
  - From V4 to TE (bypassing TEO) (Desimone et al 1980; Saleem et al 1992)
- “Replication” of simpler selectivities from lower to higher areas
- Rich dictionary of features – across areas -- with various levels of selectivity and invariance

Readings on the work with many relevant references

A detailed description of much of the work is in the "supermemo" at
http://cbcl.mit.edu/projects/cbcl/publications/ai-publications/2005/AIM-2005-036.pdf

Other recent publications and references can be found at
http://cbcl.mit.edu/publications/index-pubs.html

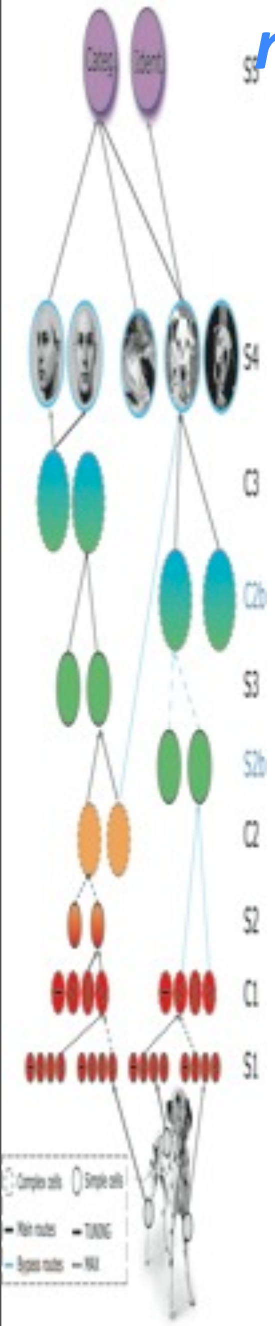
Model: testable at different levels



The most recent version of this straightforward class of models is consistent with many data at different levels -- from the *computational to the biophysical level.*

Being testable across all these levels is a high bar and an important one (too easy to develop models that explain one phenomenon or one area or one illusion...these models overfit the data, they are not scientific)

Recognition in Visual Cortex: model accounts for physiology+ psychophysics



Hierarchical Feedforward Models: is consistent with or predict neural data

V1:

Simple and complex cells tuning (Schiller et al 1976; Hubel & Wiesel 1965; Devalois et al 1982)

MAX-like operation in subset of complex cells (Lampl et al 2004)

V2:

Subunits and their tuning (Anzai, Peng, Van Essen 2007)

V4:

Tuning for two-bar stimuli (Reynolds Chelazzi & Desimone 1999)

MAX-like operation (Gawne et al 2002)

Two-spot interaction (Freiwald et al 2005)

Tuning for boundary conformation (Pasupathy & Connor 2001, Cadieu, Kouh, Connor et al., 2007)

Tuning for Cartesian and non-Cartesian gratings (Gallant et al 1996)

IT:

Tuning and invariance properties (Logothetis et al 1995, paperclip objects)

Differential role of IT and PFC in categorization (Freedman et al 2001, 2002, 2003)

Read out results (Hung Kreiman Poggio & DiCarlo 2005)

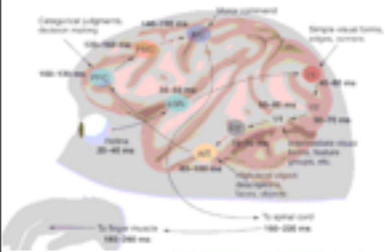
Pseudo-average effect in IT (Zoccolan Cox & DiCarlo 2005; Zoccolan Kouh Poggio & DiCarlo 2007)

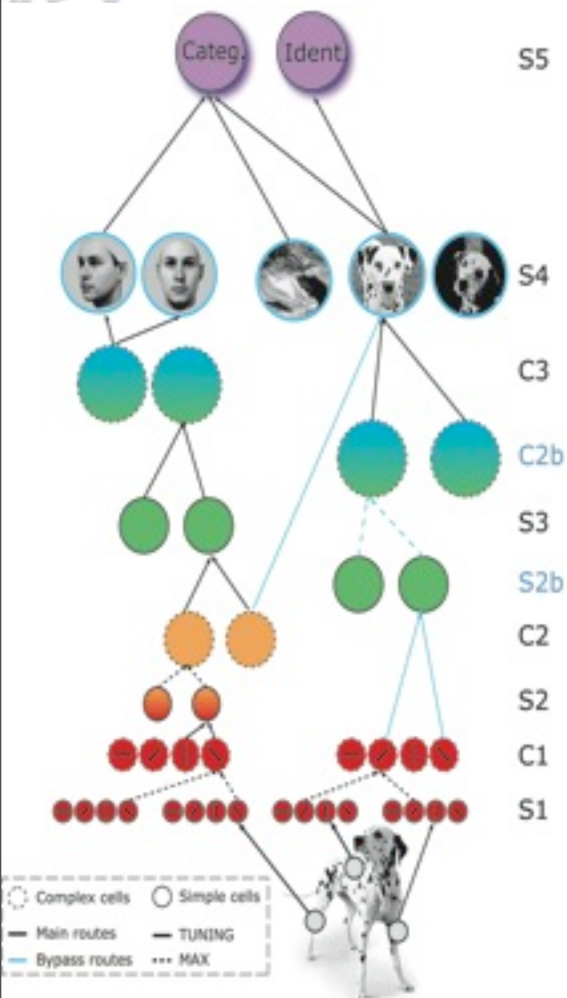
Human:

Rapid categorization (Serre Oliva Poggio 2007)

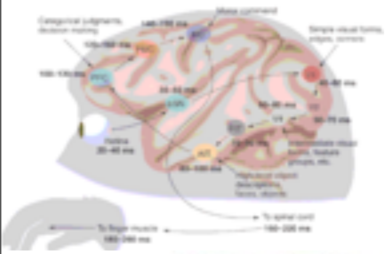
Face processing (fMRI + psychophysics) (Riesenhuber et al 2004; Jiang et al 2006)

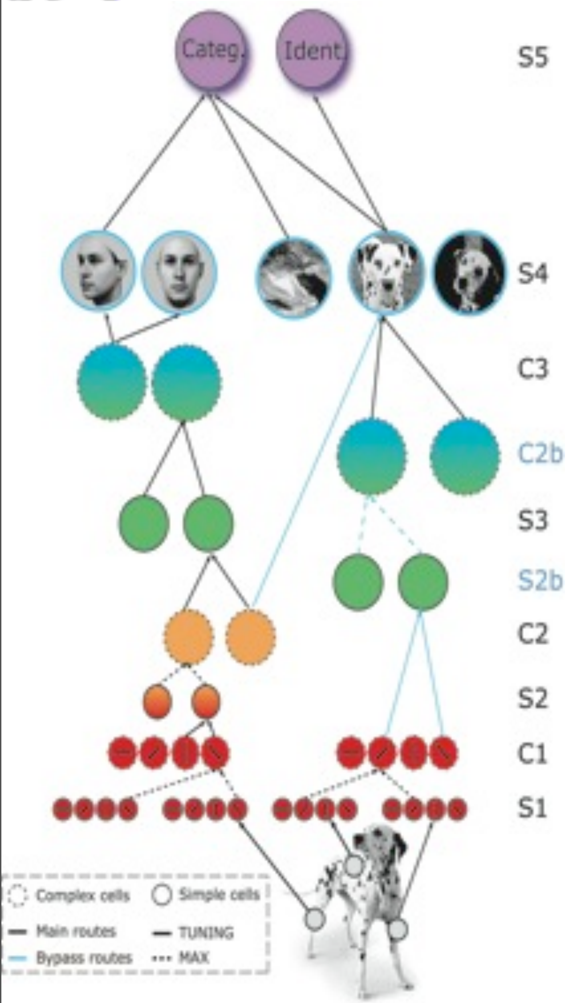
Recognition in Visual Cortex: model accounts for phychophysics



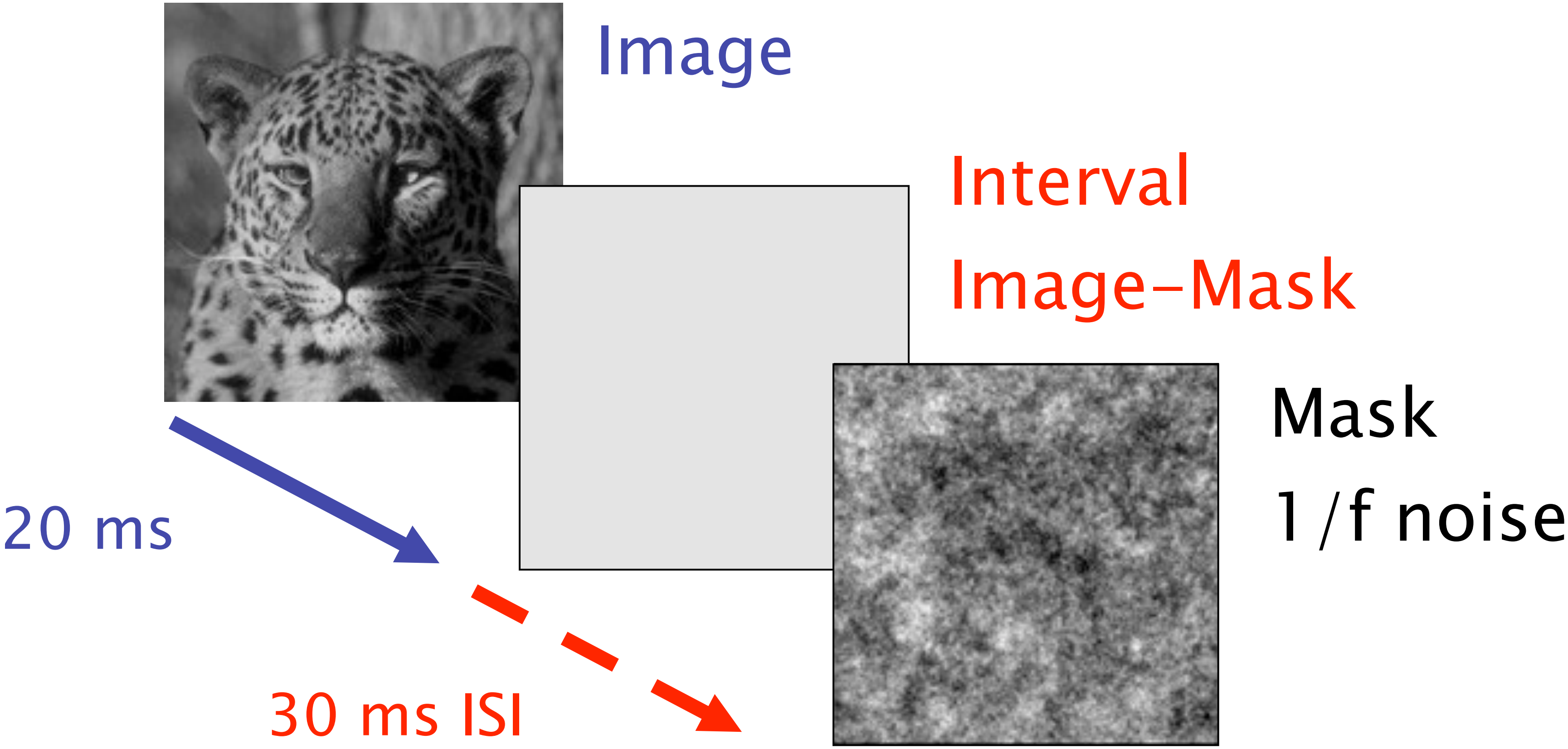


Recognition in Visual Cortex: model accounts for phychophysics





Hierarchical feedforward models of the ventral stream

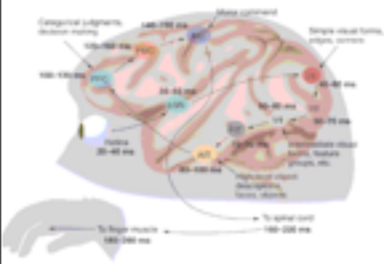


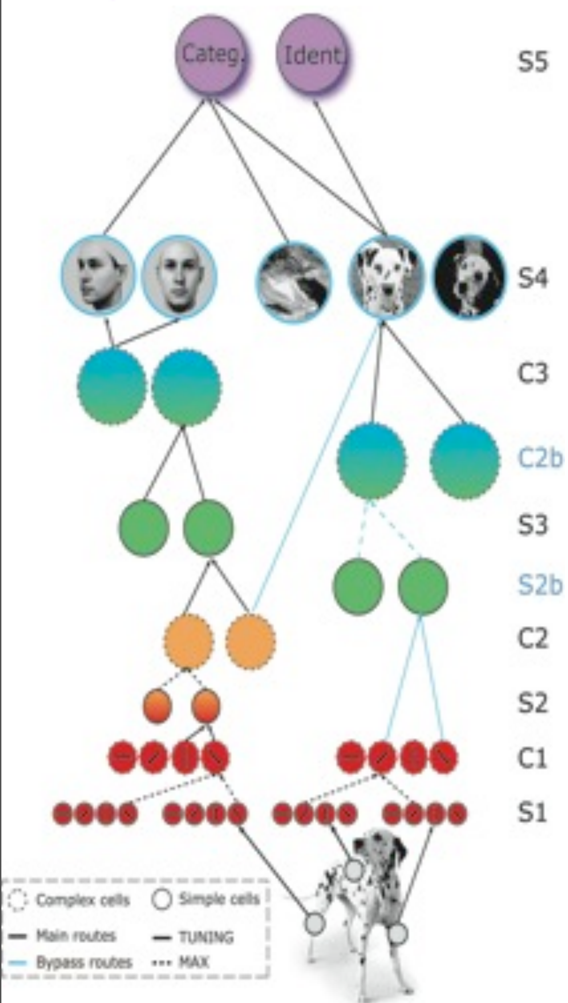
Animal present or not ?

Rapid Categorization: mask should force visual cortex to operate in feedforward mode

Thorpe et al 1996; Van Rullen & Koch 2003; Bacon-Mace et al 2005

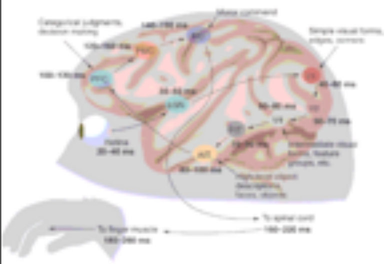
Hierarchical feedforward models of the ventral stream

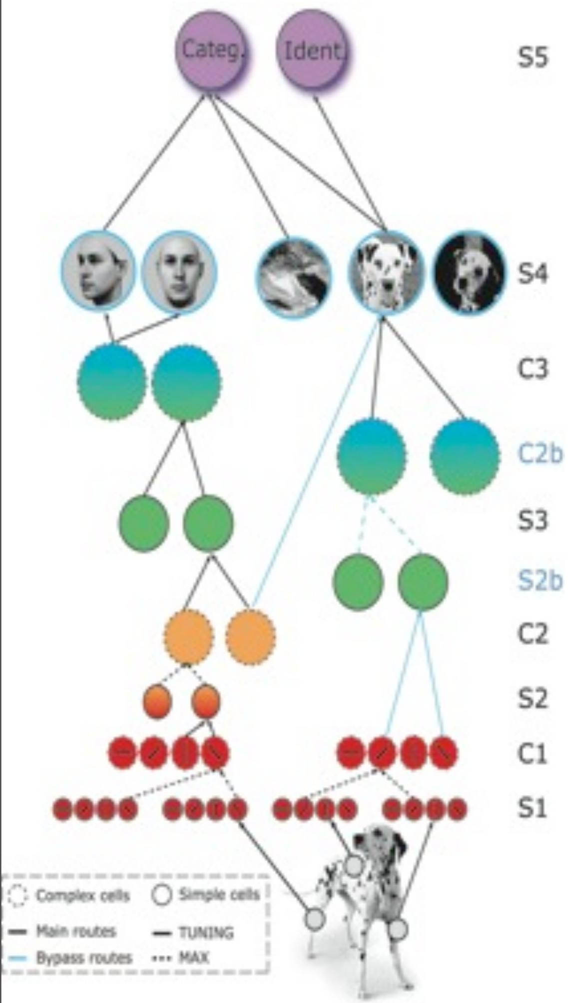




Rapid Categorization

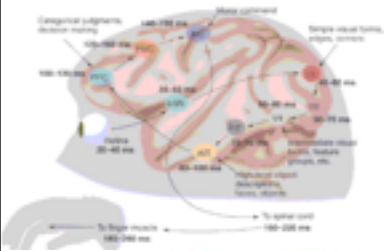
Hierarchical feedforward models of the ventral stream

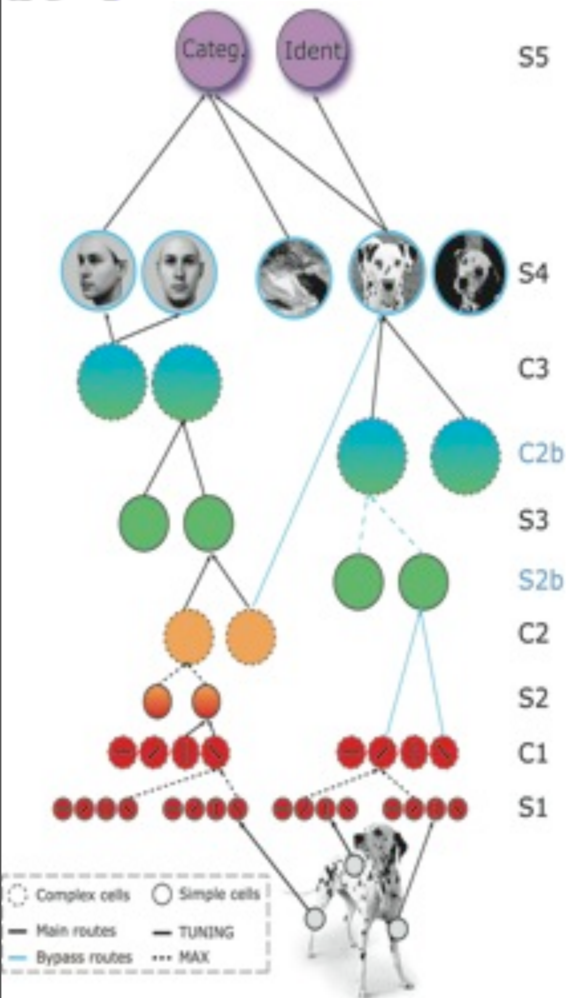




Rapid Categorization

Recognition in Visual Cortex: model accounts for phychophysics





Feedforward Models: "predict" rapid categorization (82% model vs. 80% humans)

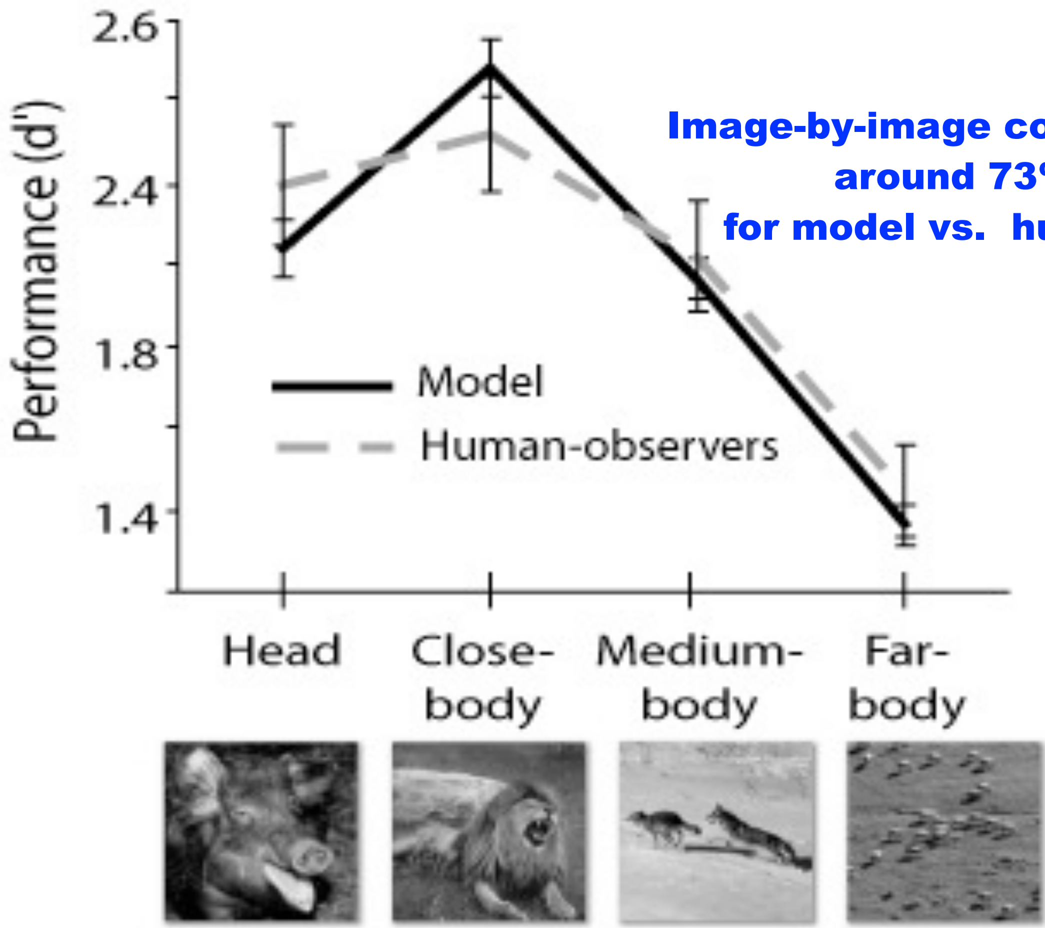
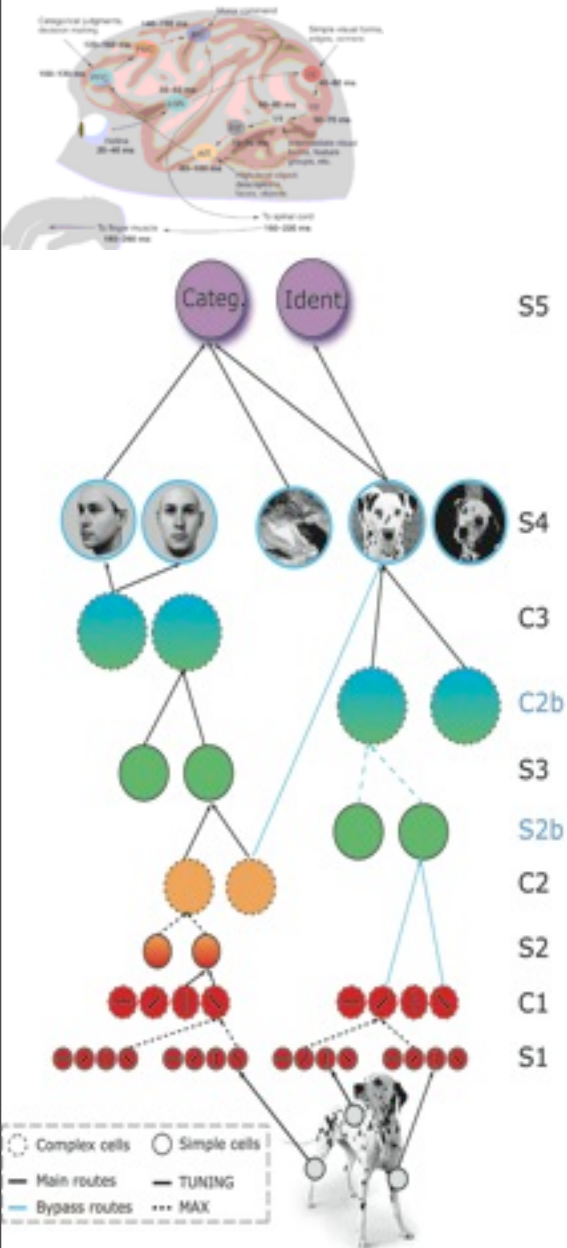
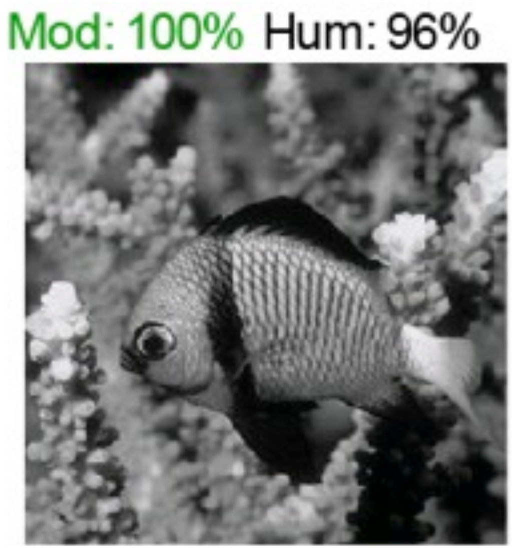


Image-by-image correlation: around 73% for model vs. humans)

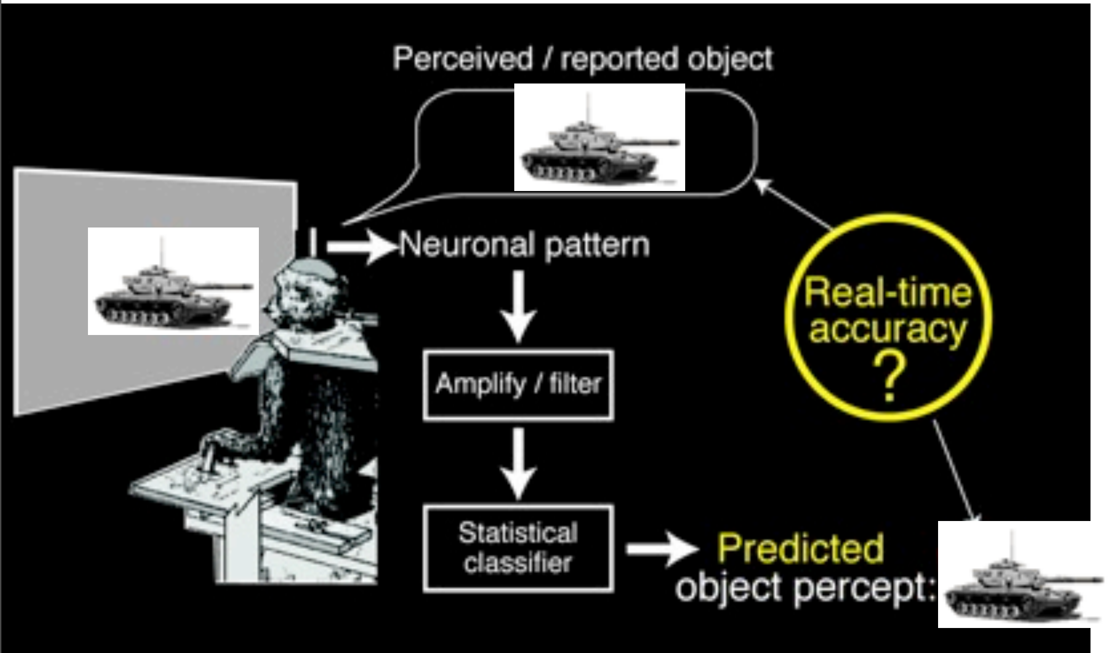
Hierarchical model of recognition in visual cortex

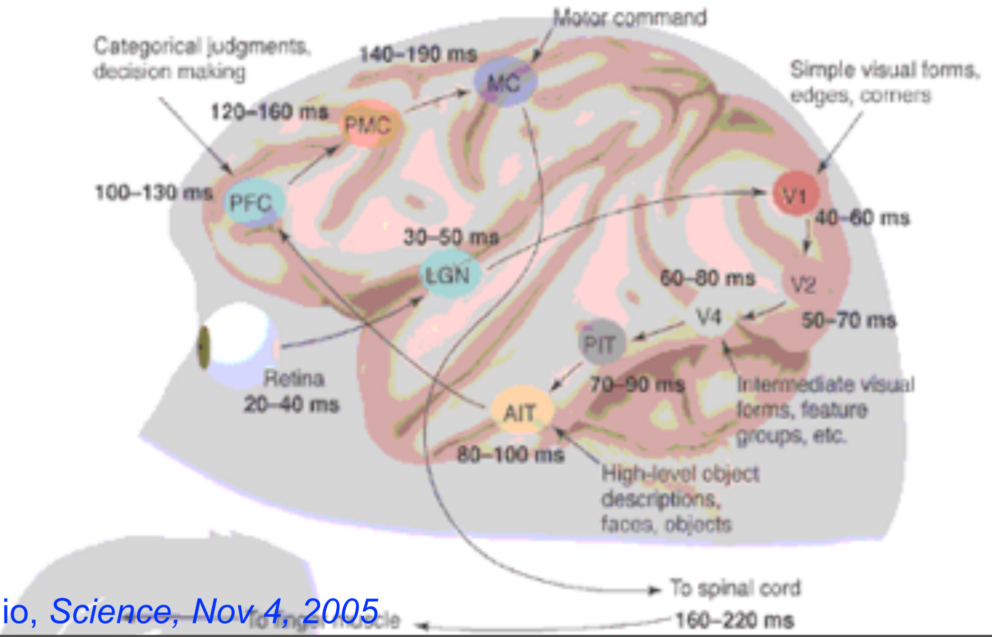


- Image-by-image correlation:
  - Heads: ρ=0.71
  - Close-body: ρ=0.84
  - Medium-body: ρ=0.71
  - Far-body: ρ=0.60



Agreement of model w| IT Readout data

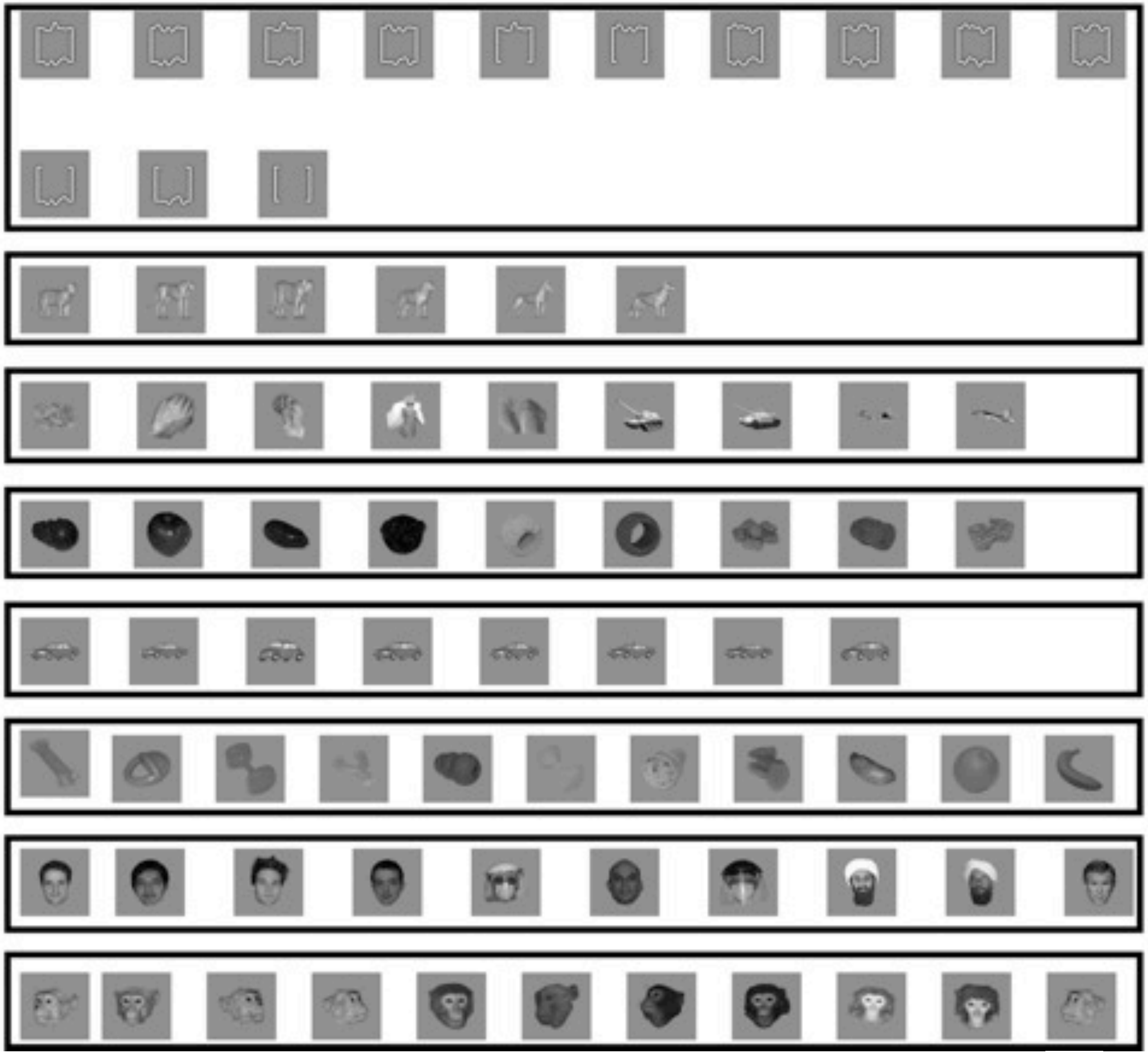




Chou Hung, Gabriel Kreiman, James DiCarlo, Tomaso Poggio, *Science, Nov 4, 2005*

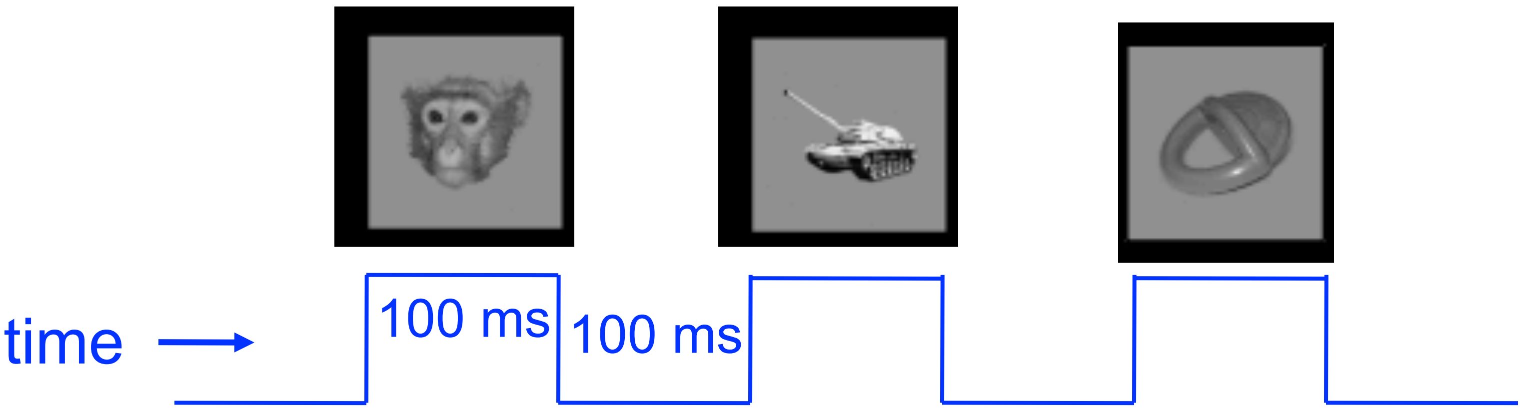
Reading-out the neural code in AIT

77 objects,
8 classes



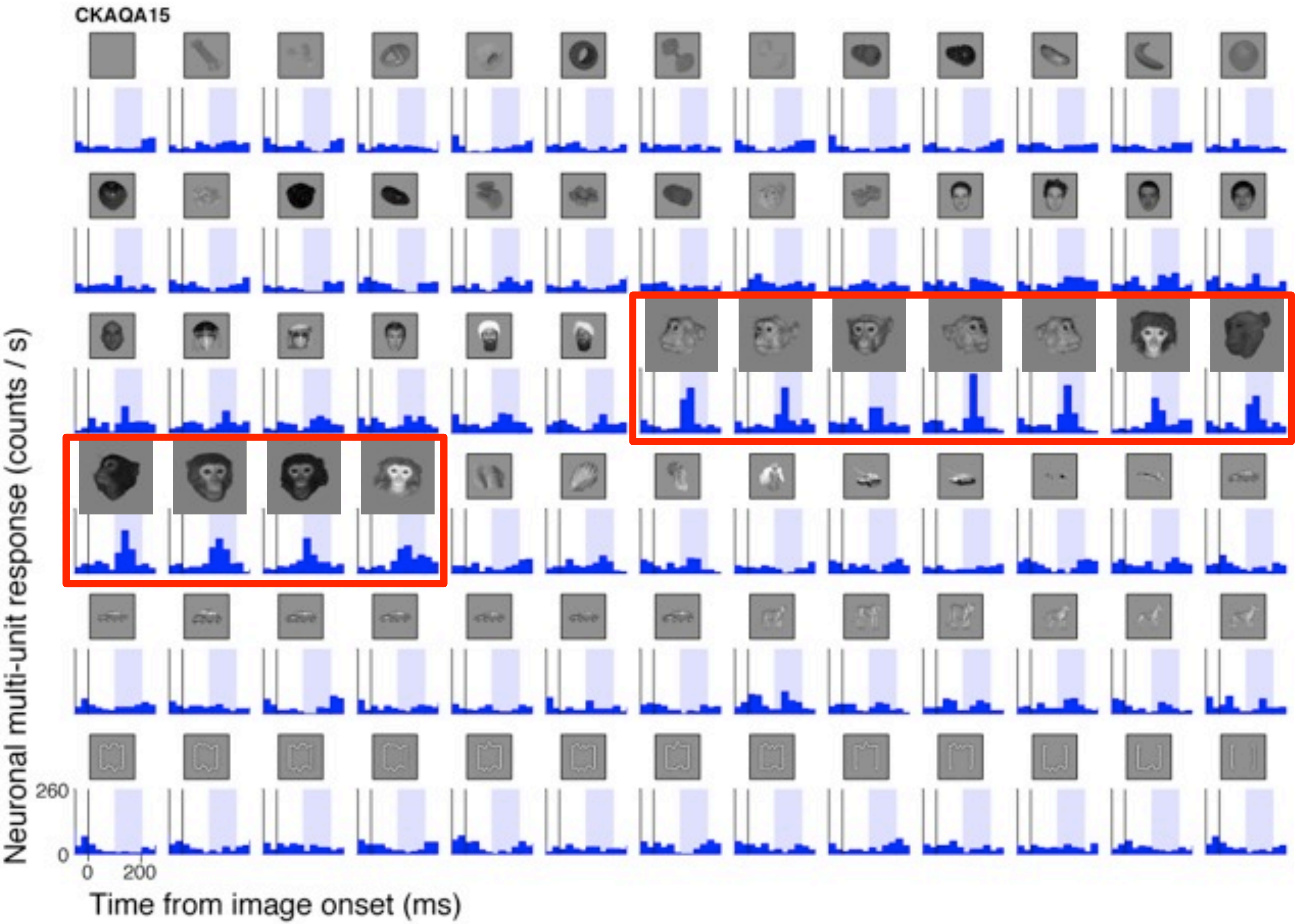
Chou Hung, Gabriel Kreiman, James DiCarlo, Tomaso Poggio, *Science, Nov 4, 2005*

Recording at each recording site during passive viewing



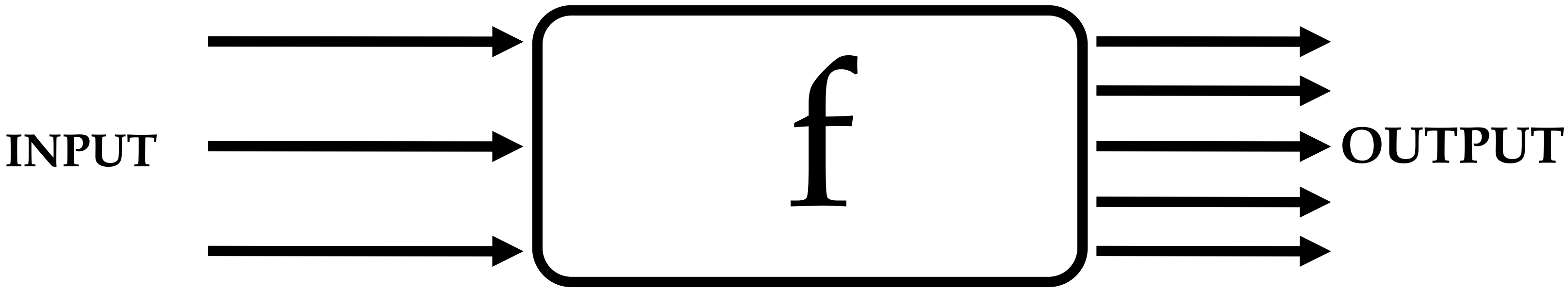
- 77 visual objects
- 10 presentation repetitions per object
- presentation order randomized and counter-balanced

Agreement of model w| IT Readout data



Chou Hung, Gabriel Kreiman, James DiCarlo, Tomaso Poggio, *Science, Nov 4, 2005*

Training a classifier on neuronal activity.



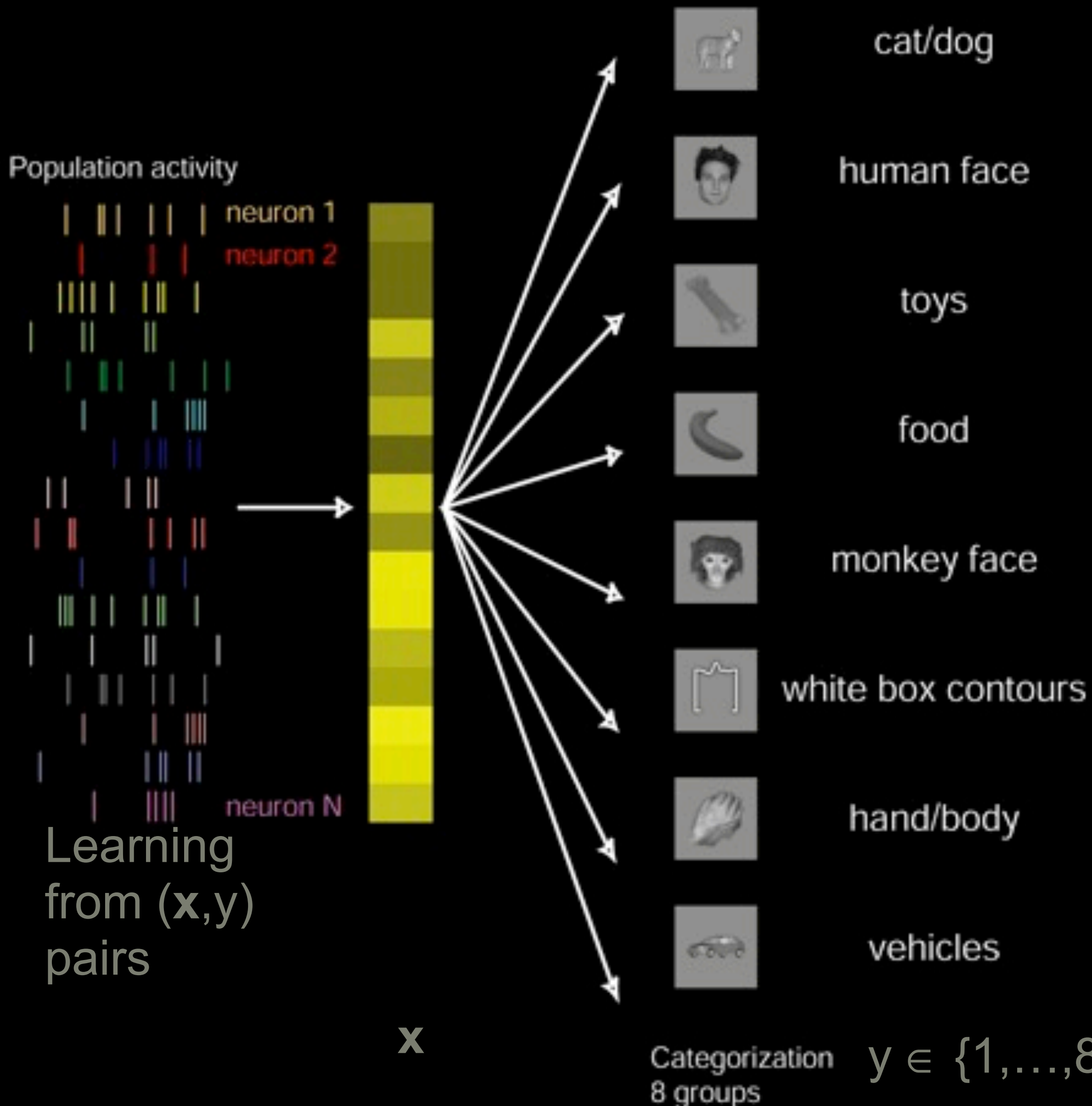
From a set of data (vectors of activity of n neurons (x) and object label (y)

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\}$$

Find (by training) a classifier eg a function f such that $f(x) = \hat{y}$

is a **good predictor** of object label y for a **future** neuronal activity x

Decoding the Neural Code ...
population response (using a classifier)



Learning from (**x**,y) pairs

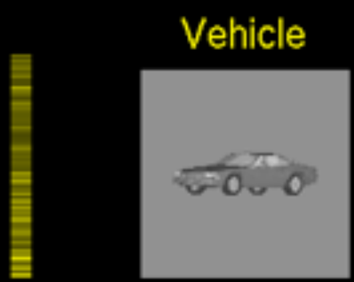
x

$y \in \{1,\ldots,8\}$

Hung*, Kreiman, Poggio, DiCarlo. *Science 2005*

From neuronal population activity…

…a classifier can decode and guess what the monkey was seeing…



Categorization

- Toy
- Body
- Human Face
- Monkey Face
- Vehicle
- Food
- Box
- Cat/Dog

Video speed: 1 frame/sec
Actual presentation rate: 5 objects/sec

80% accuracy in read-out from ~200 neurons

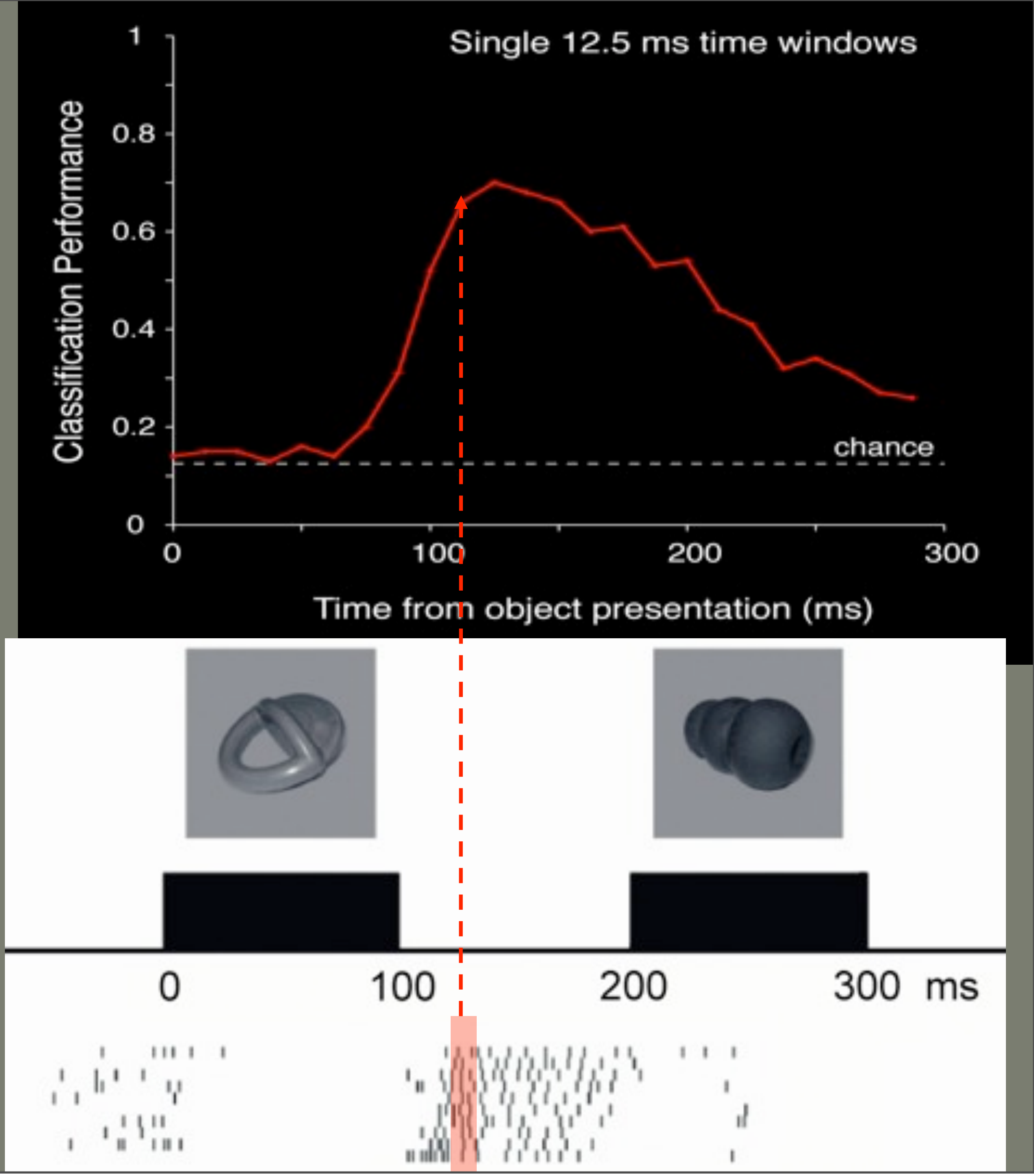
So…experimentally we can decode the brain’s code and
read-out from neural activity what the monkey is seeing

We can also read-out with similar results from the model !!!

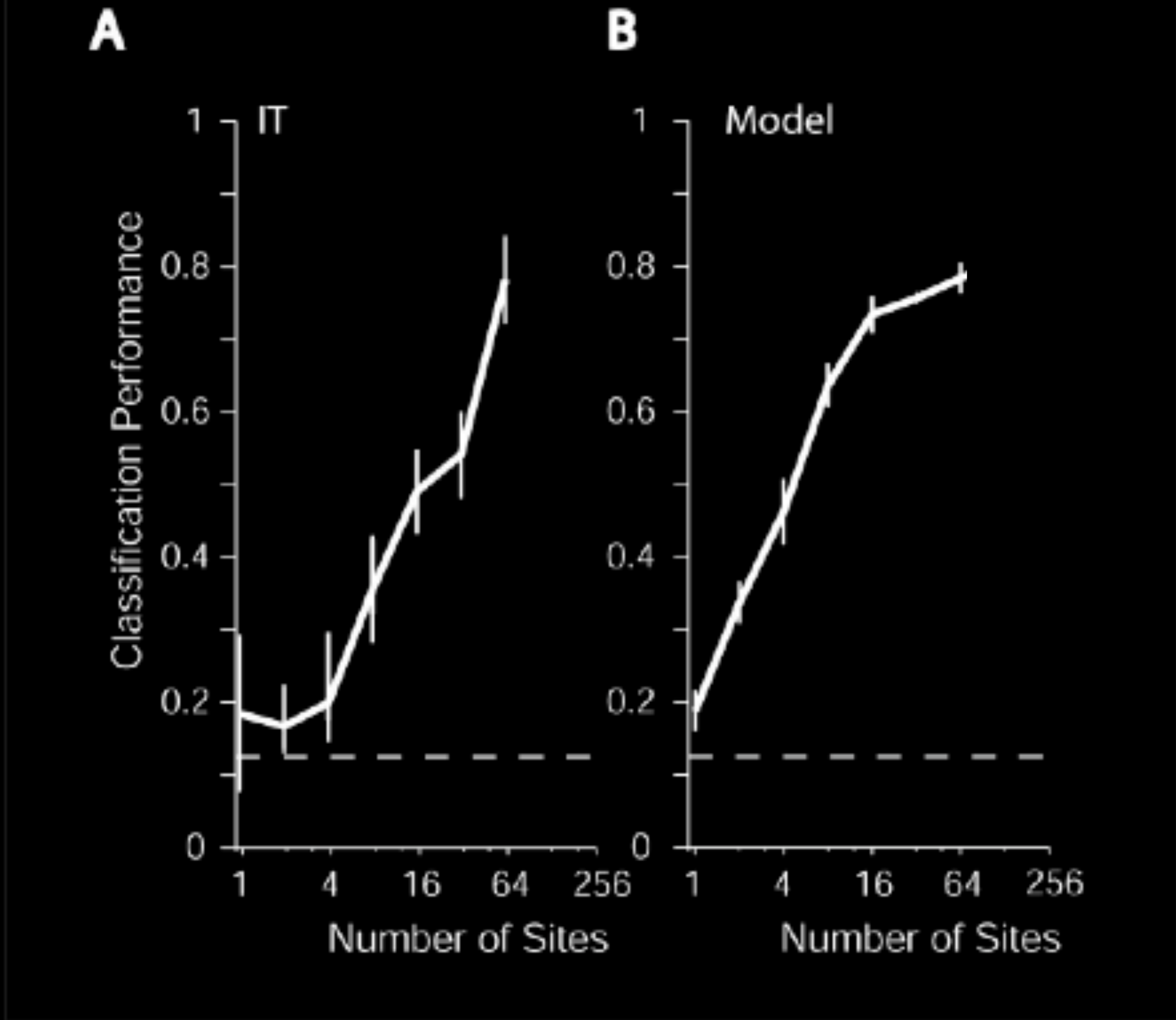
A result (C. Hung, et al., 2005): very rapid read-out of object information rapid (80-100 ms from onset of stimulus)

Information represented by population of neurons over very short times (over 12.5ms bin)

Very strong constraint on neural code (not firing rate). Consistent with our IF circuits for max and tuning



It turns out that the model agrees with IT data: we can decode from model units as well as from IT

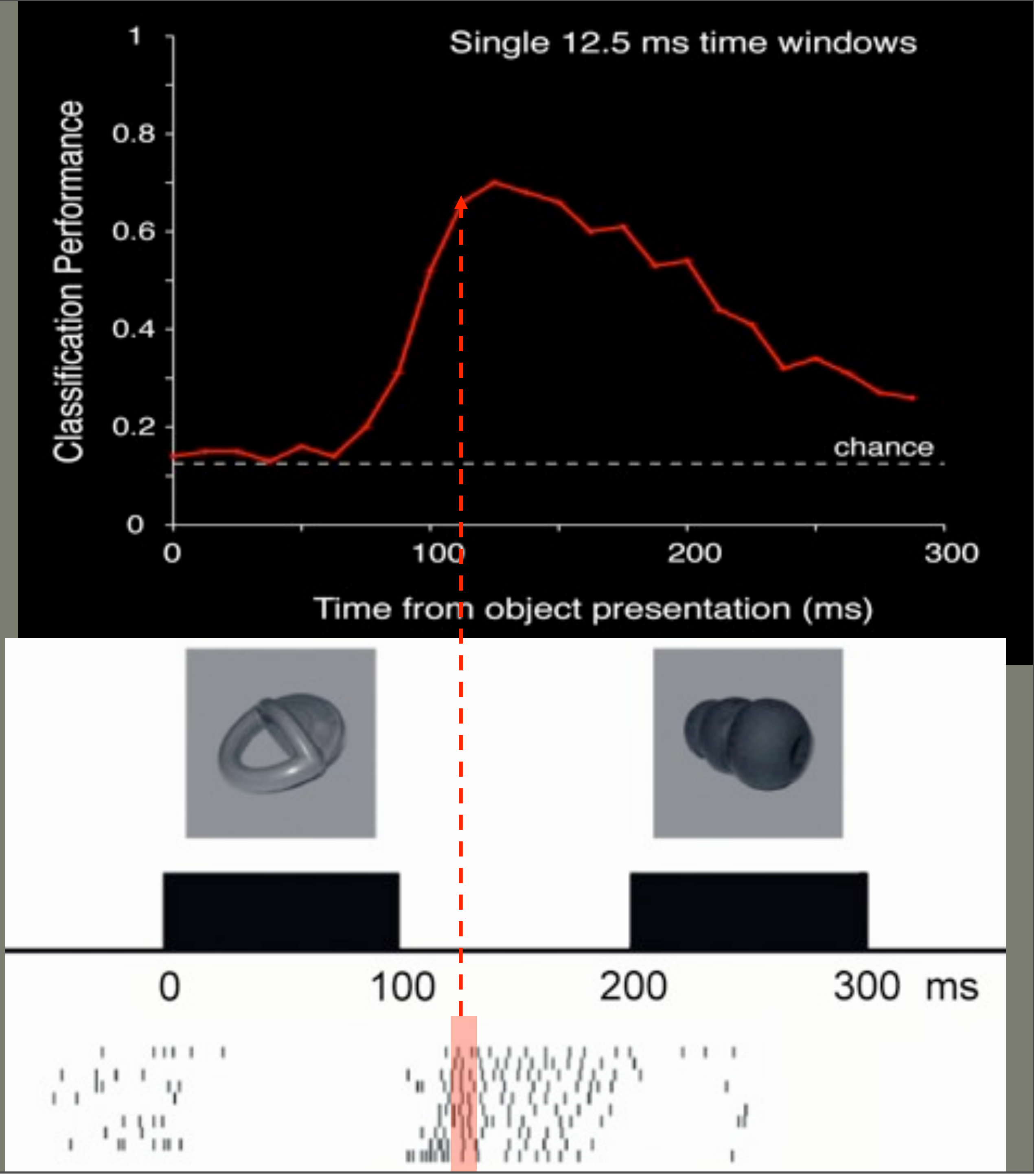


Serre, Kouh, Cadieu, Knoblich, Kreiman, Poggio. *MIT AI Memo* 2005

A result (C. Hung, et al., 2005): very rapid read-out of object information rapid (80-100 ms from onset of stimulus)

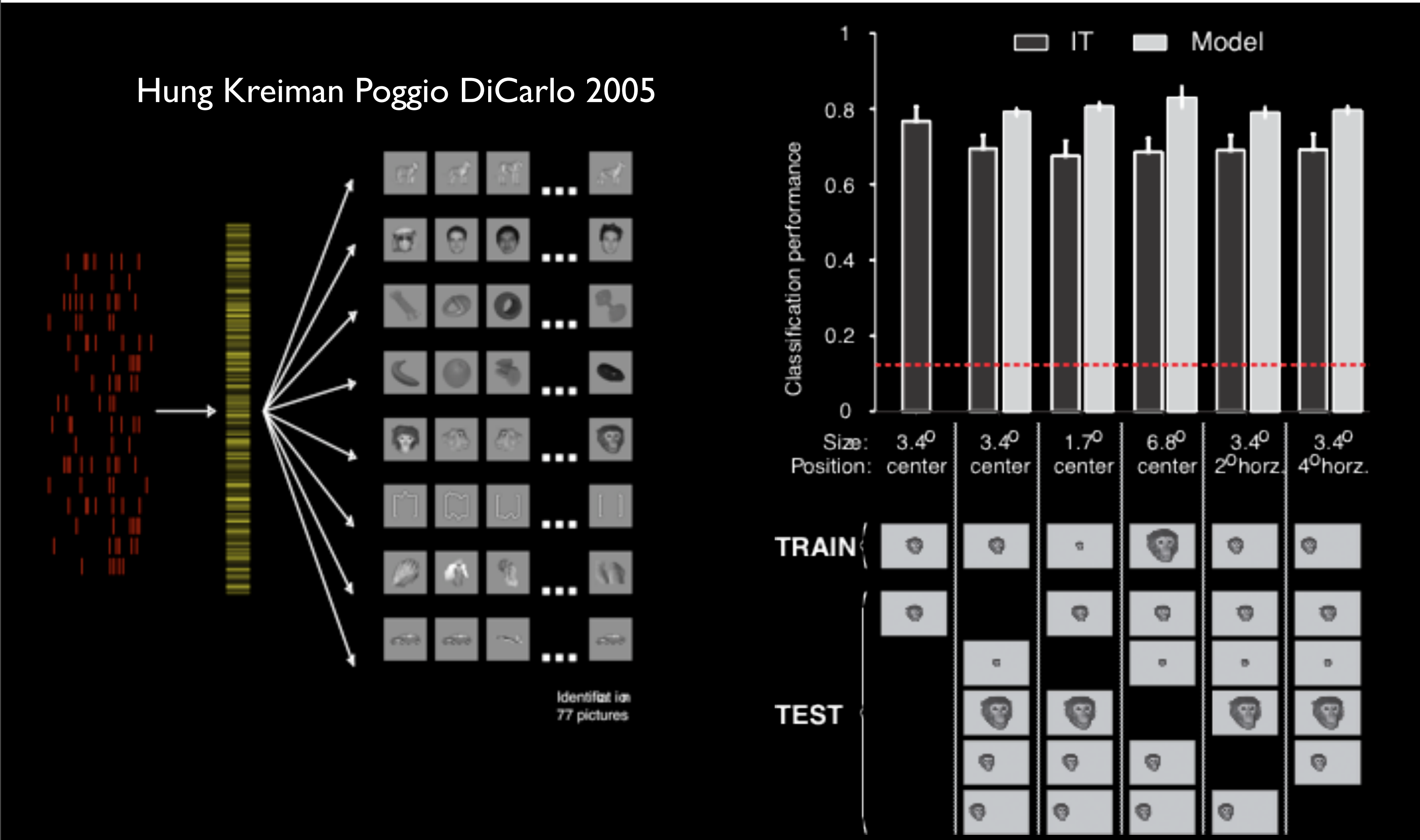
Information represented by population of neurons over very short times (over 12.5ms bin)

Very strong constraint on neural code (not firing rate). Consistent with our IF circuits for max and tuning



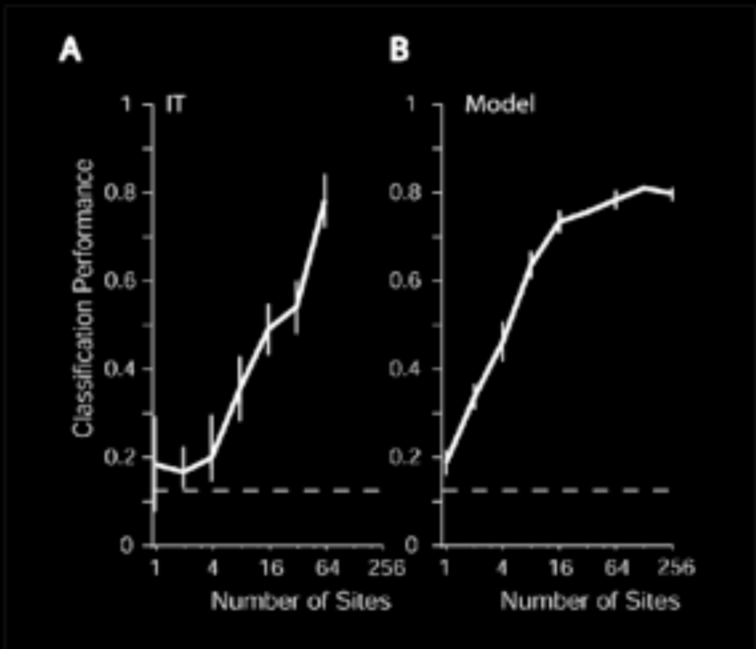
Agreement of model w| IT Readout data

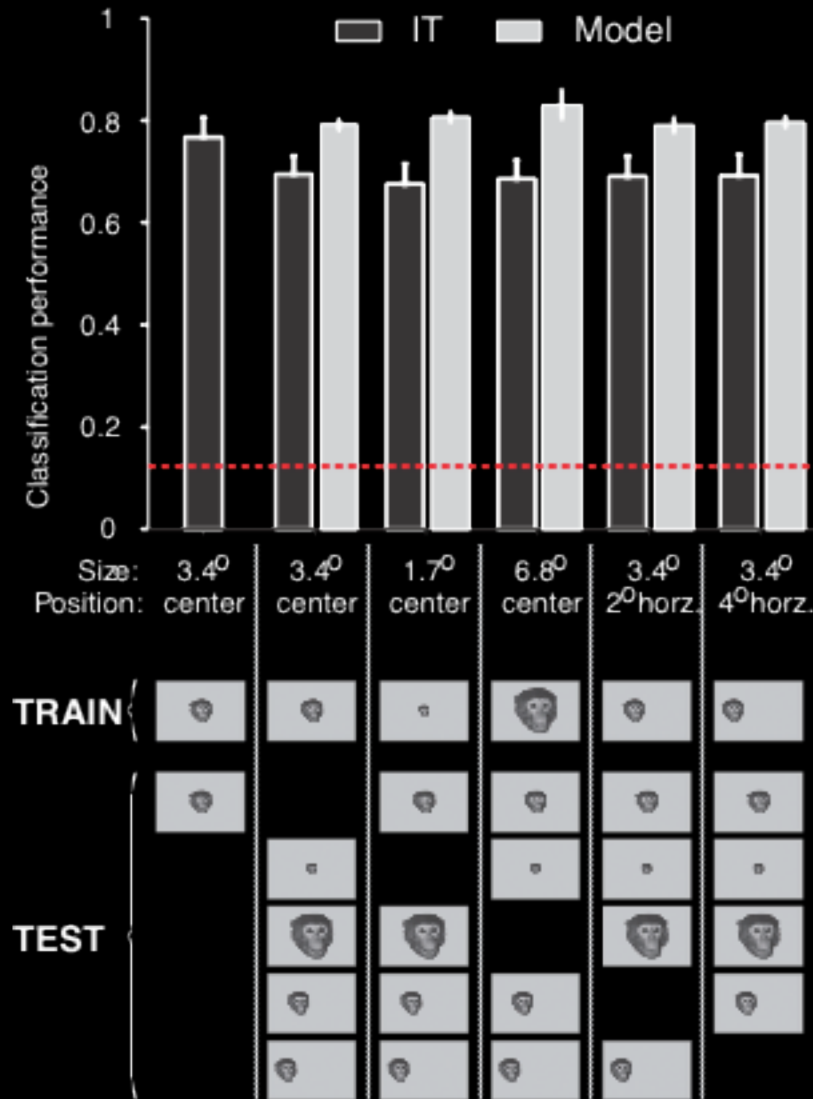
Reading out category and identity invariant to position and scale



Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005

Agreement of Model w| IT Readout data

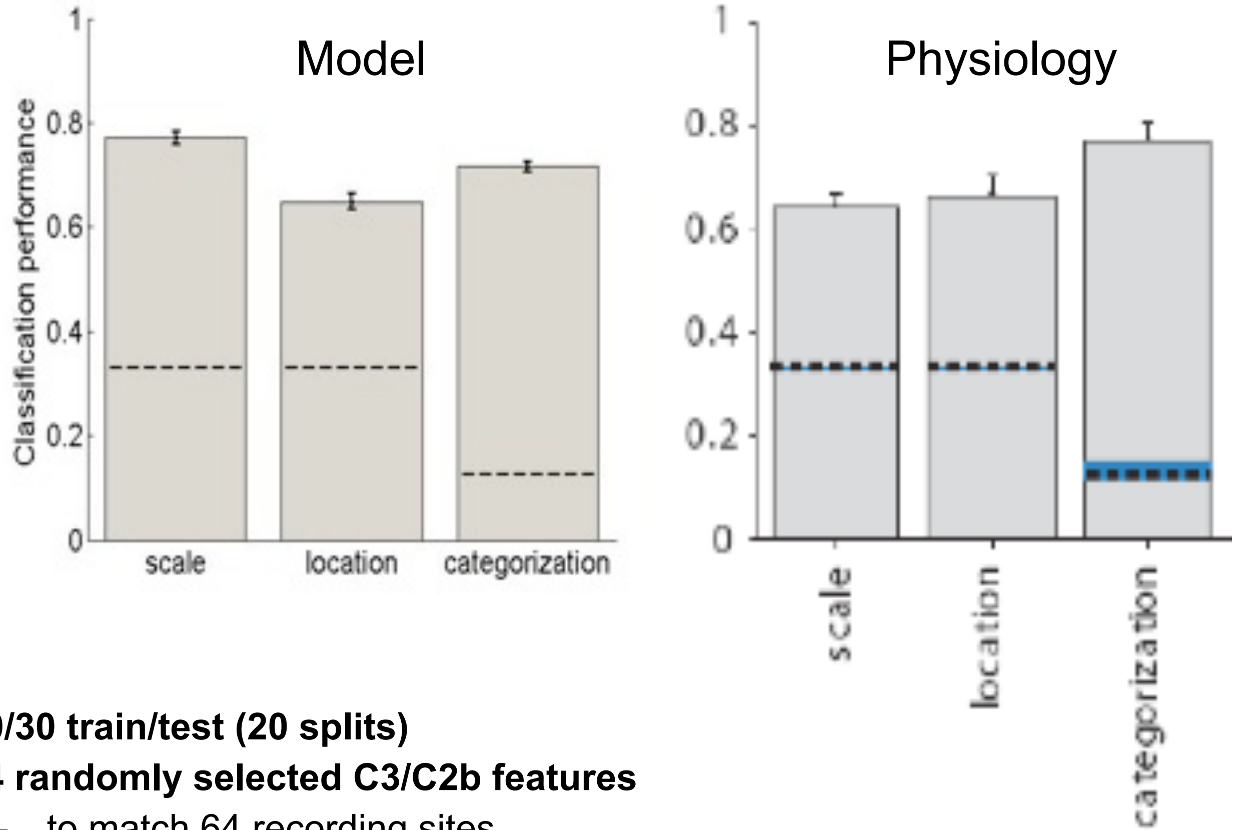




Reading out category and identity "*invariant*" to position and scale

Hung, et al. 2005; Serre et al., 2005

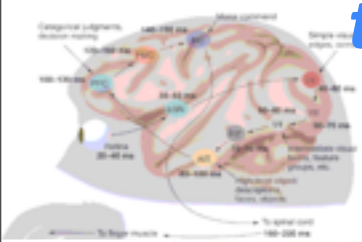
Reading Out Scale and Position Information: comparing the model to Hung et al.

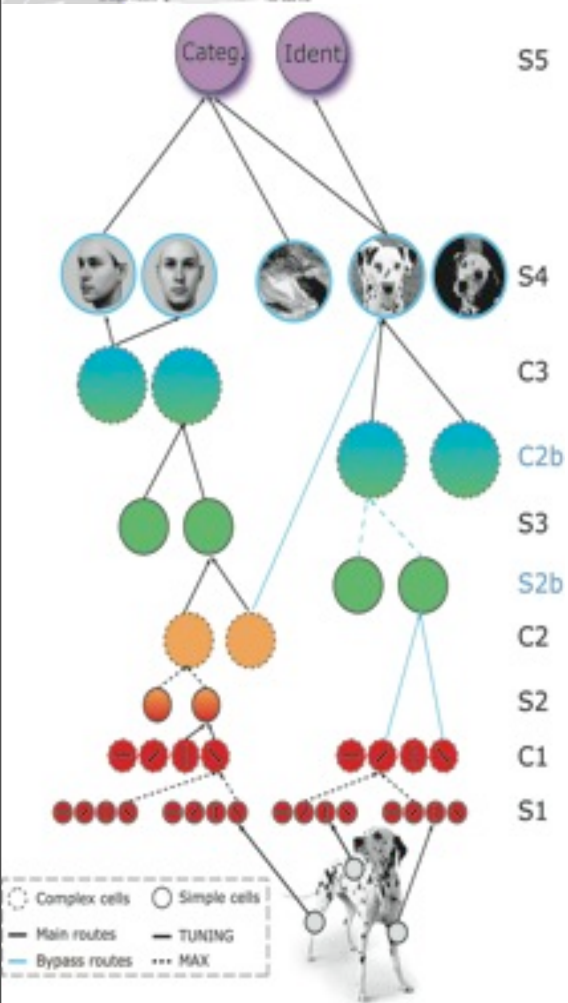


- **70/30 train/test (20 splits)**
- **64 randomly selected C3/C2b features**
  - to match 64 recording sites
- **Scale:** 77.2 ± 1.25% vs. ~63% (physiology)
- **Location:** 64.9 ± 1.44% vs. ~65% (physiology)
- **Categorization:** 71.6 ± 0.91% vs. ~77% (physiology)

Tan, Serre, Poggio, 2008

Recognition in Visual Cortex: testing computational performance

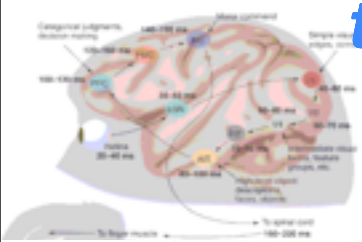


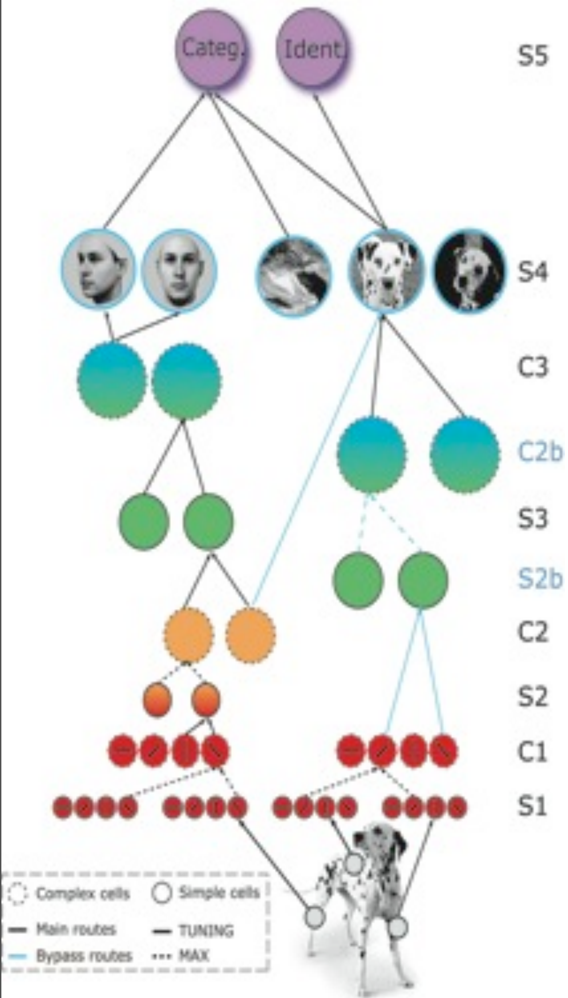


Models of the *ventral stream* in cortex
perform well compared to
engineered computer vision systems (in 2006)
on several databases

Bileschi, Wolf, Serre, Poggio, 2007

Recognition in Visual Cortex: testing computational performance

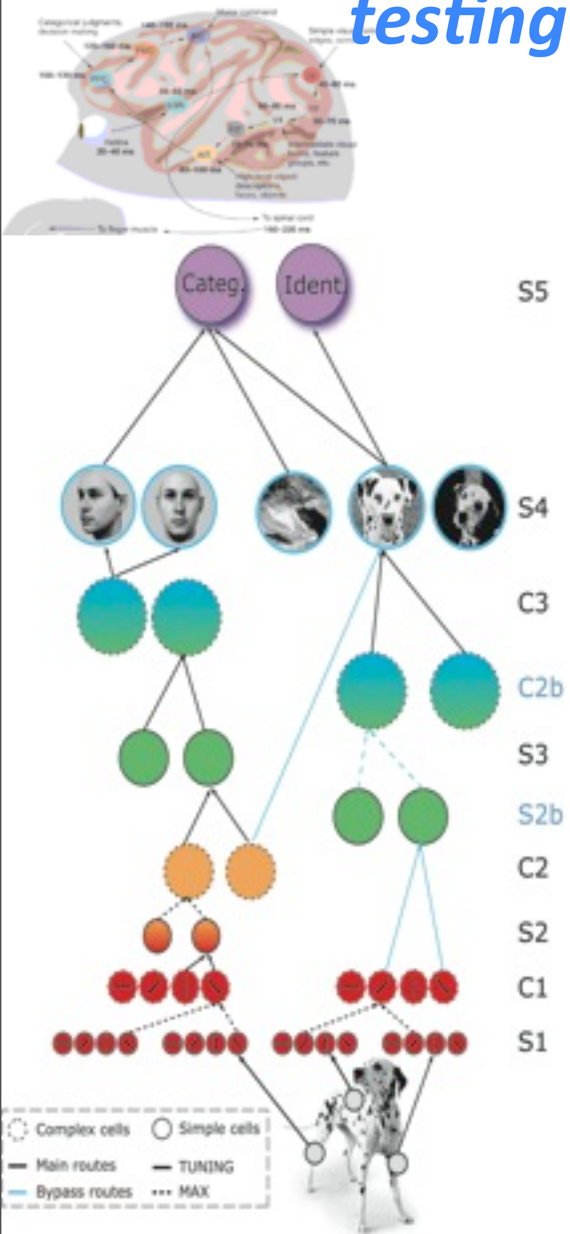




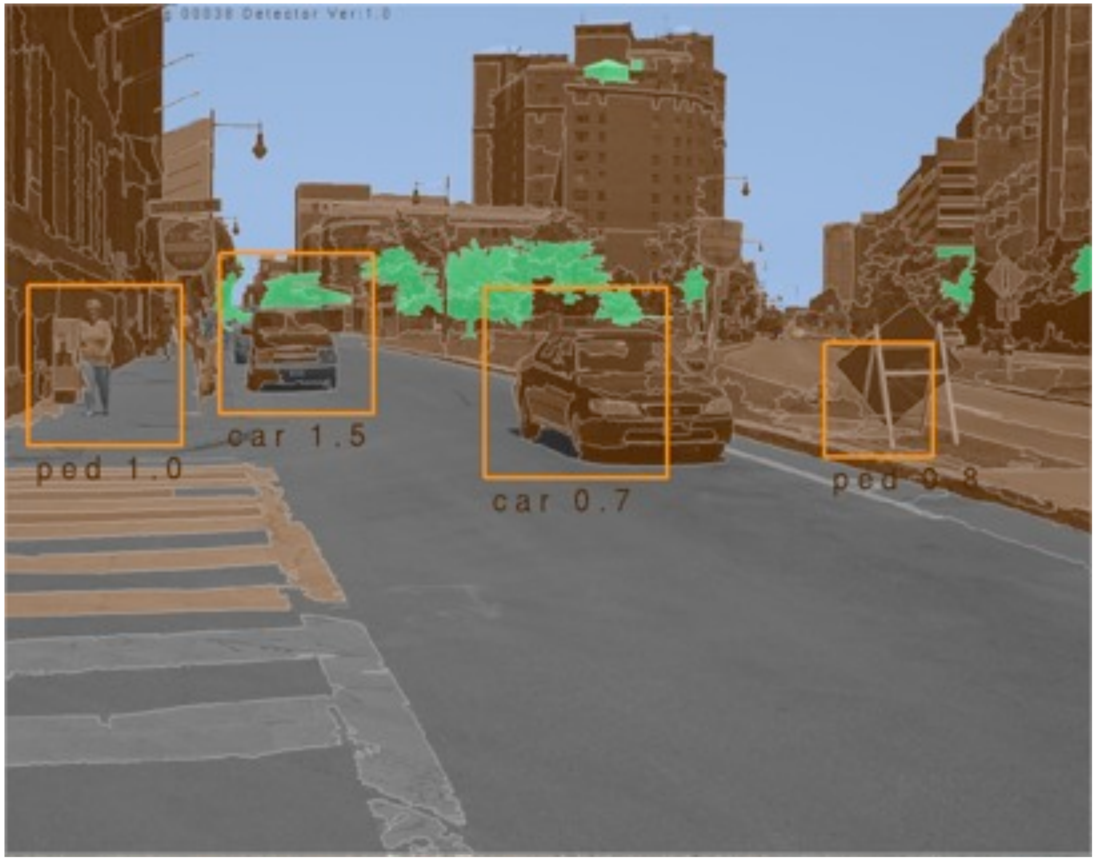
Models of the *ventral stream* in cortex
perform well compared to
engineered computer vision systems (in 2006)
on several databases

Bileschi, Wolf, Serre, Poggio, 2007

Recognition in Visual Cortex: testing computational performance

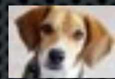


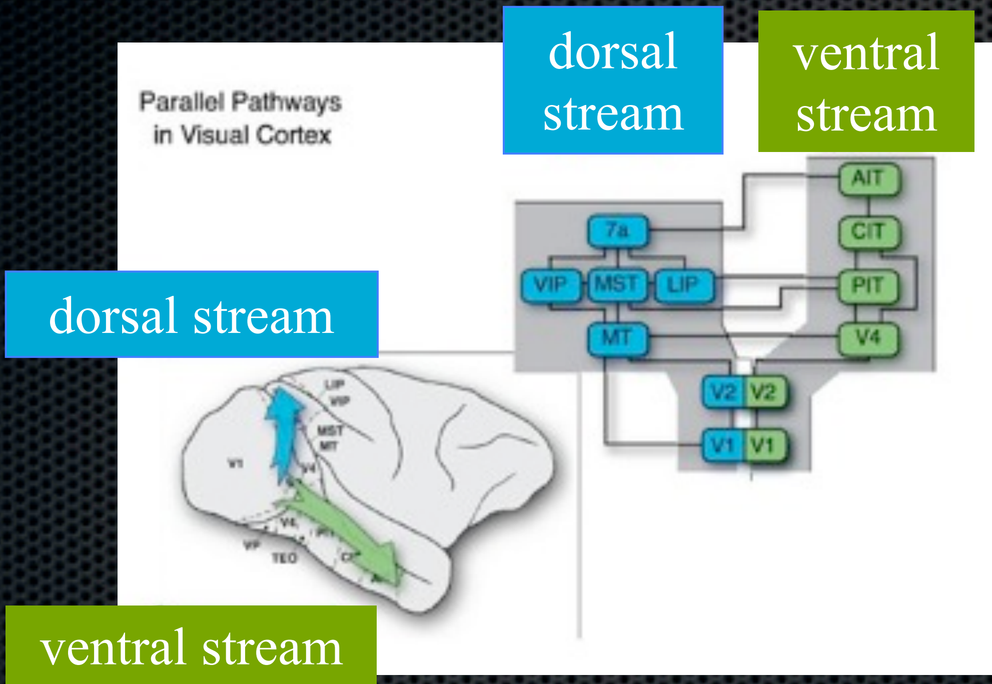
Models of the <u>*ventral stream*</u> in cortex
perform well compared to
engineered computer vision systems (in 2006)
on several databases



Bileschi, Wolf, Serre, Poggio, 2007

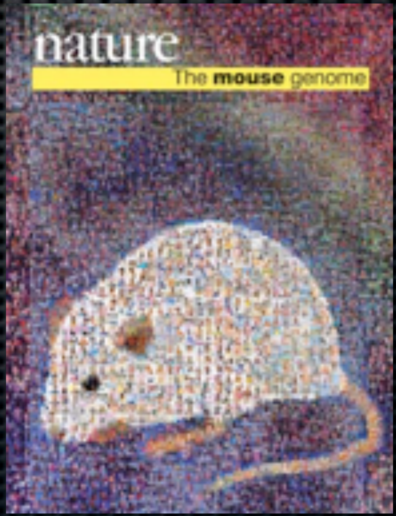
Model extension to the dorsal stream: Recognition of actions





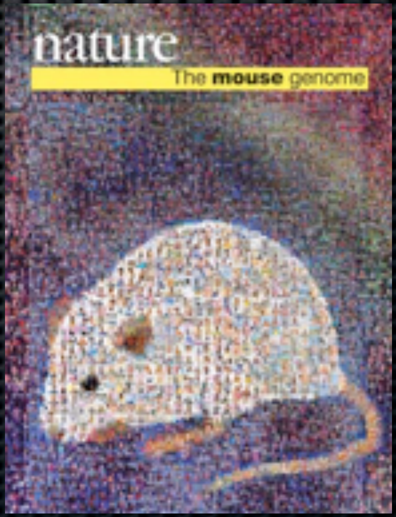
Thomas Serre, Hueihan Jhuang & Tomaso Poggio collaboration with David Sheinberg at Brown University

Quantitative automatic phenotyping



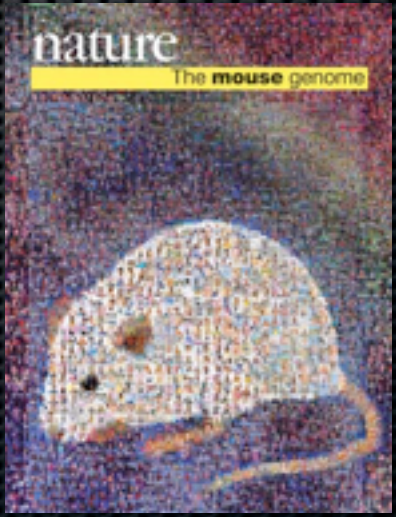
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:



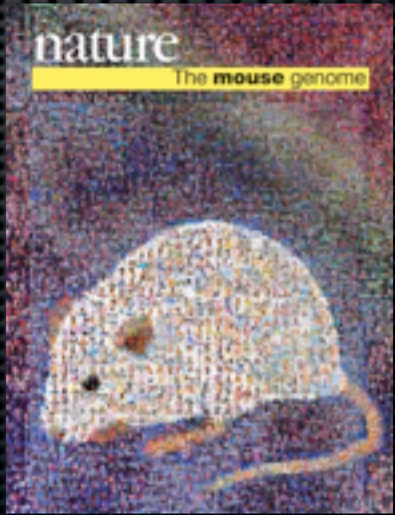
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes



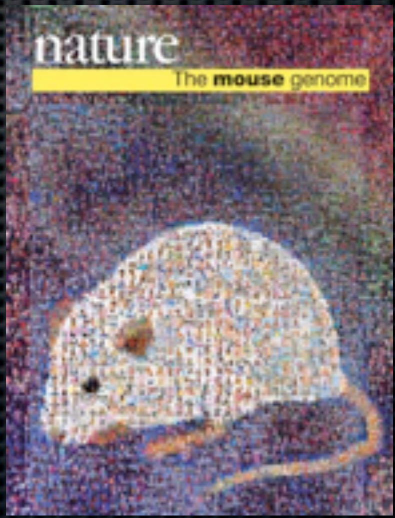
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases



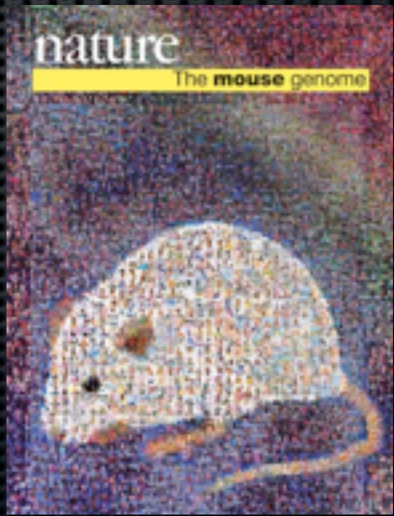
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs



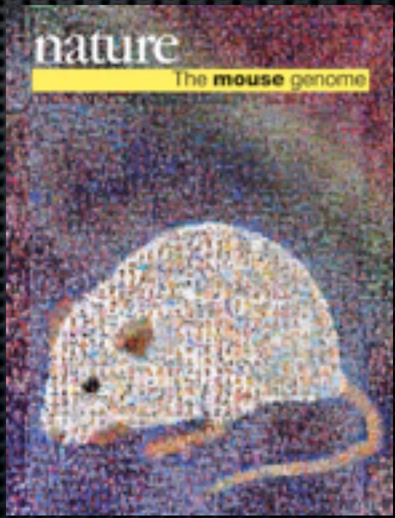
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:



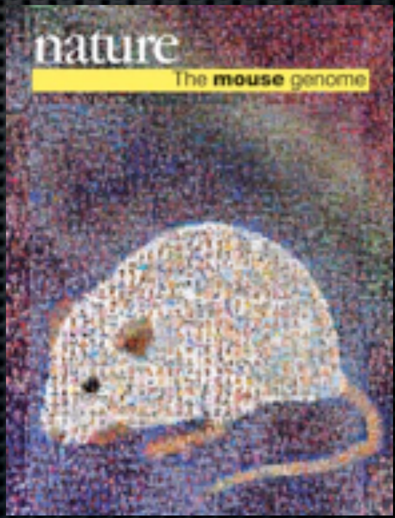
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:
  - Limit subjectivity of human intervention



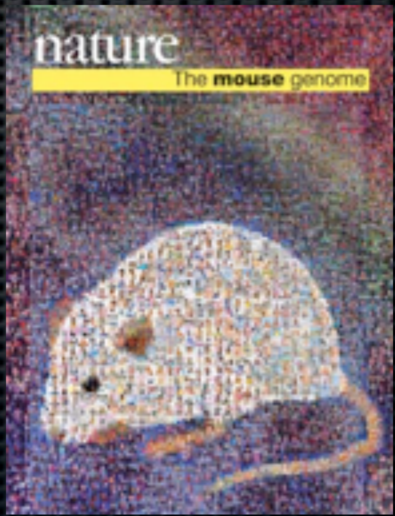
Quantitative automatic phenotyping

- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:
  - Limit subjectivity of human intervention
  - 24/7 home-cage analysis of behavior



Quantitative automatic phenotyping

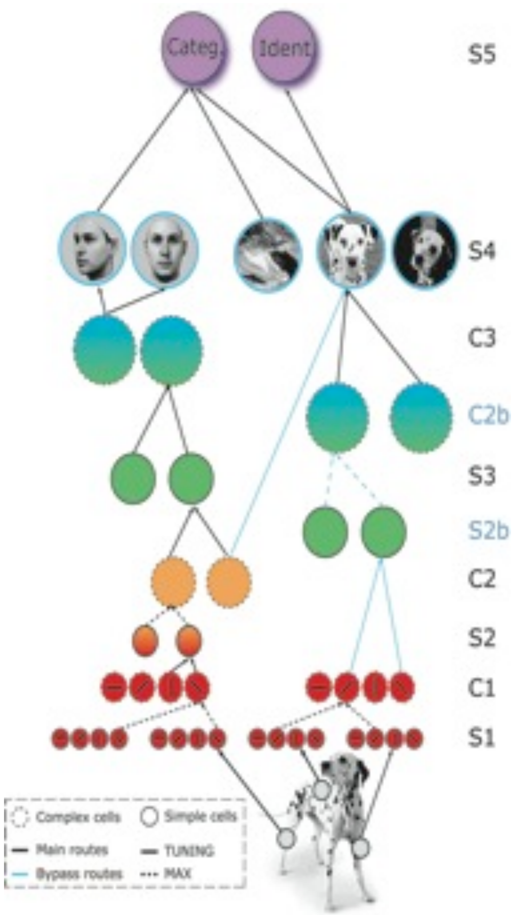
- Behavioral analyses of mouse behavior needed to:
  - Assess functional roles of genes
  - Validate models of mental diseases
  - Help assess efficacy of drugs
- Automated quant system to help:
  - Limit subjectivity of human intervention
  - 24/7 home-cage analysis of behavior
  - 24/7 monitoring of animal well-being



Recognition in Visual Cortex: testing computational performance

Models of the *dorsal stream* in cortex lead to better systems for action recognition in videos: automatic phenotyping of mice.

Hierarchical model of recognition: action recognition, ventral + dorsal stream (Giese and Poggio 2003);

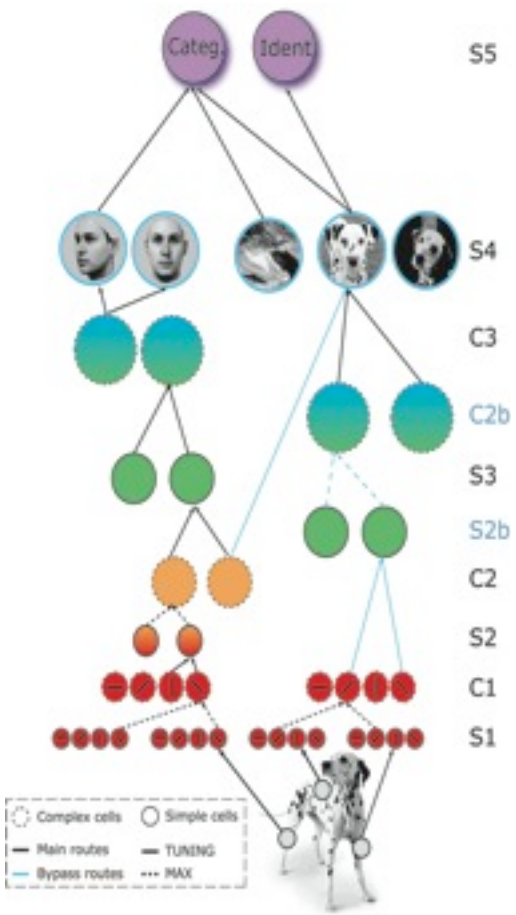


Jhuang , Garrote, Yu, Khilnani, Poggio, Mutch, Steele, Serre, Nature Communicatons, 2010

Recognition in Visual Cortex: testing computational performance

Models of the *dorsal stream* in cortex lead to better systems for action recognition in videos: automatic phenotyping of mice.

Hierarchical model of recognition: action recognition, ventral + dorsal stream (Giese and Poggio 2003);



Jhuang , Garrote, Yu, Khilnani, Poggio, Mutch, Steele, Serre, Nature Communicatons, 2010

Recognition in Visual Cortex: testing computational performance

Models of cortex lead to better systems for action recognition in videos: automatic phenotyping of mice

Performance

human agreement	72%
proposed system	77%
commercial system	61%
chance	12%

Jhuang , Garrote, Yu, Khilnani, Poggio, Mutch Steele, Serre, Nature Communicatons, 2010

Recognition in Visual Cortex: testing computational performance

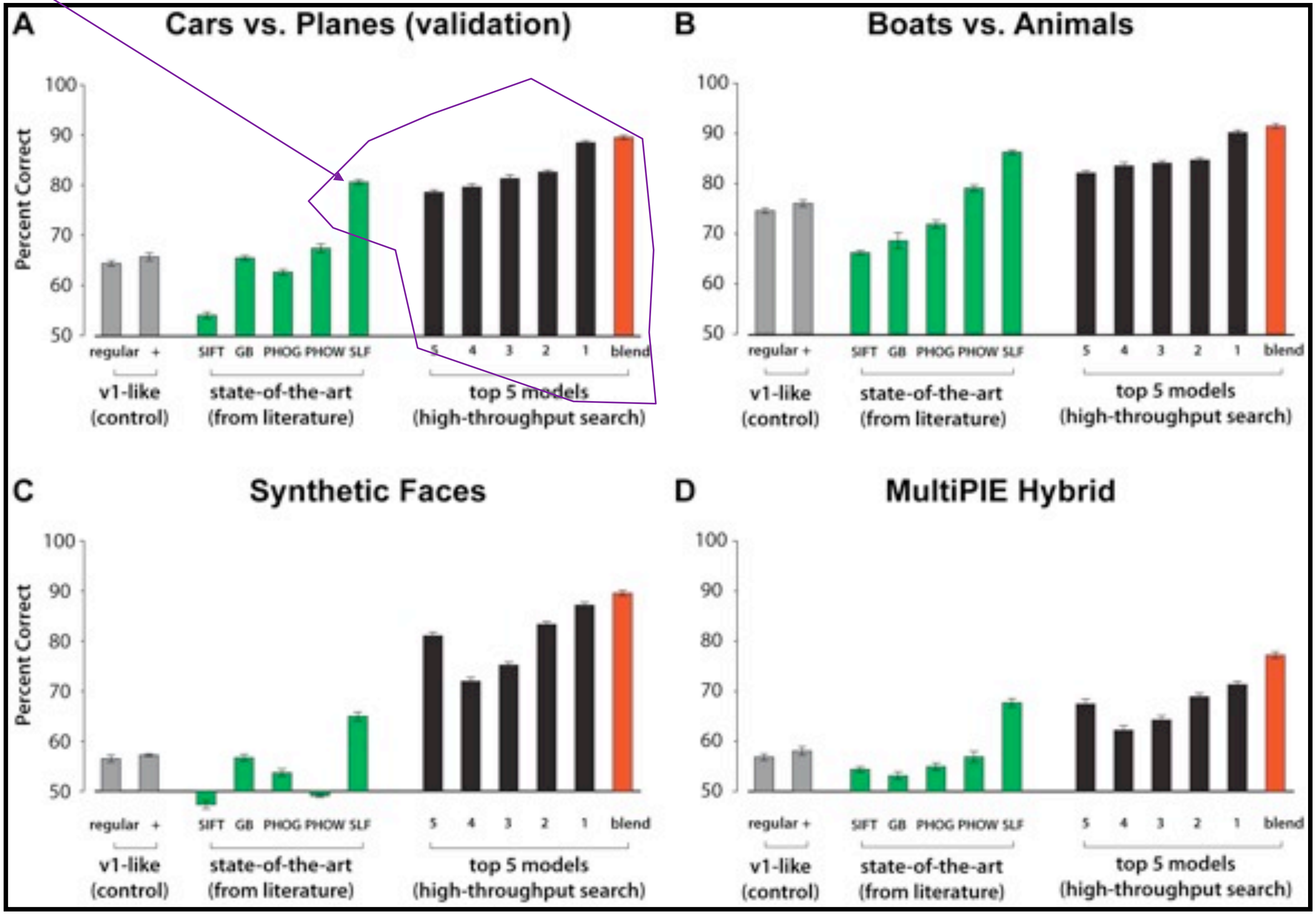
Performance

human agreement	72%
proposed system	77%
commercial system	61%
chance	12%

Models of cortex lead to better systems for action recognition in videos: automatic phenotyping of mice

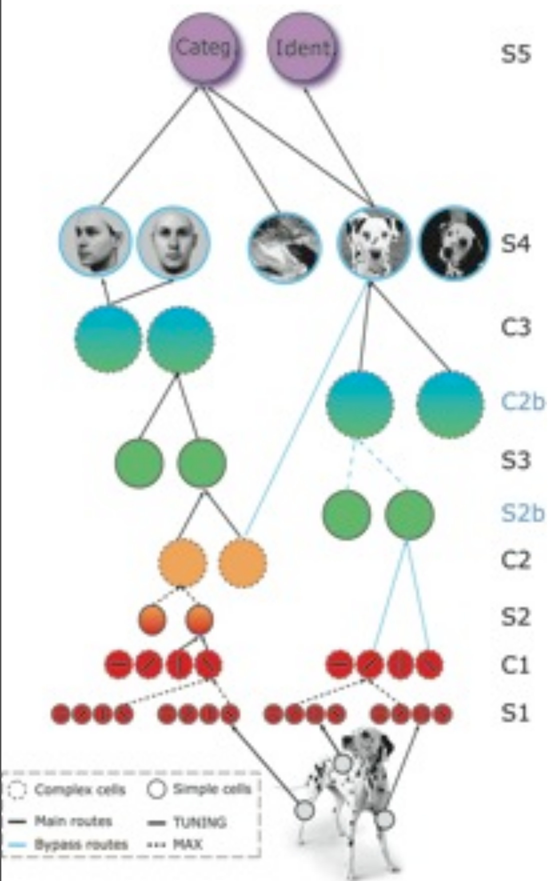
Jhuang , Garrote, Yu, Khilnani, Poggio, Mutch Steele, Serre, Nature Communicatons, 2010

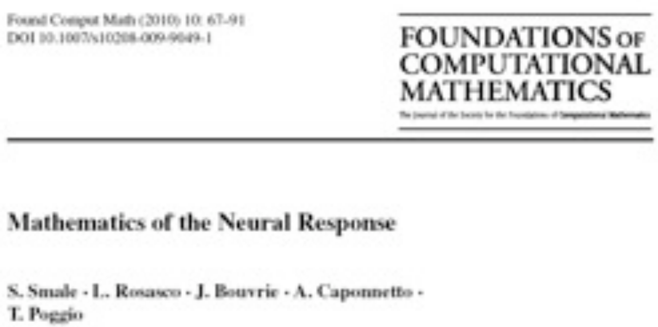
Recognition in Visual Cortex: testing computational performance



Nicholas Pinto, PhD thesis, 2010

Recognition in Visual Cortex: computation and mathematical theory





Found Comput Math (2010) 10: 67–91
DOI 10.1007/s10208-009-9049-1

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS

Mathematics of the Neural Response

S. Smale · L. Rosasco · J. Bouvrie · A. Caponnetto · T. Poggio

For 10years+...

I did not manage to understand how model works....

we need theories -- not only models!

What do hierarchical architectures compute? How? How do they develop?

THE COMPUTATIONAL MAGIC OF THE VENTRAL STREAM: TOWARDS A THEORY

Tomaso Poggio[⋆,†] (section 4 with Jim Mutch[⋆]; appendix 7.2 with Joel Leibo[⋆] and appendix 7.9 with Lorenzo Rosasco†)
⋆ *CBCL, McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA, USA*
† Istituto Italiano di Tecnologia, Genova, Italy

More on models of the dorsal stream: action recognition and applications

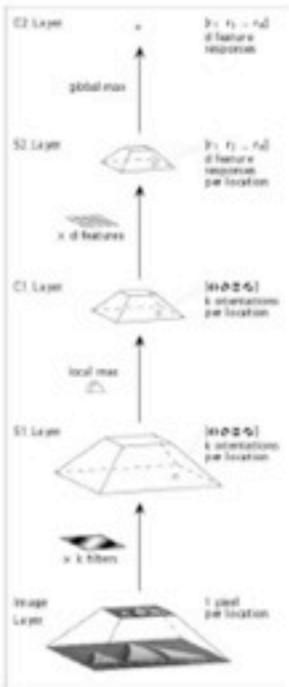
Hueihan Jhuang

HLMs:

a mathematical framework for hierarchical learning machines

Lorenzo Rosasco: Class 22

Efficient software implementation: a GPU-based framework for simulating cortically-organized networks (*CNS: available on our Web site*)

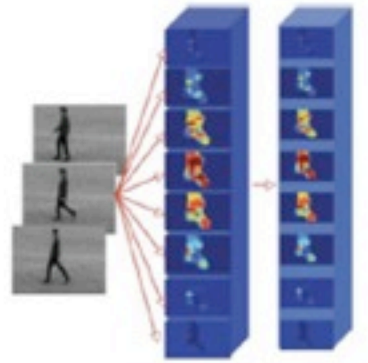


Feedforward object recognition (static CBCL model):

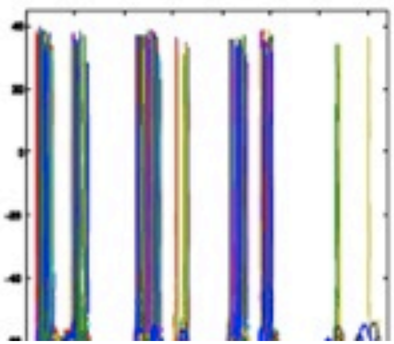
- 256x256 input, 12 orientations, 4,075 "S2" features.
- Best CPU-based implementation: 28.2 sec/image.
- CNS (on NVIDIA GTX 295): 0.291 sec/image (97x speedup).

Action recognition in streaming video:

- 8 9x9x9 spatiotemporal filters, 300 S2 features.
- Best CPU-based implementation: 0.55 fps.
- CNS: 32 fps (58x speedup).



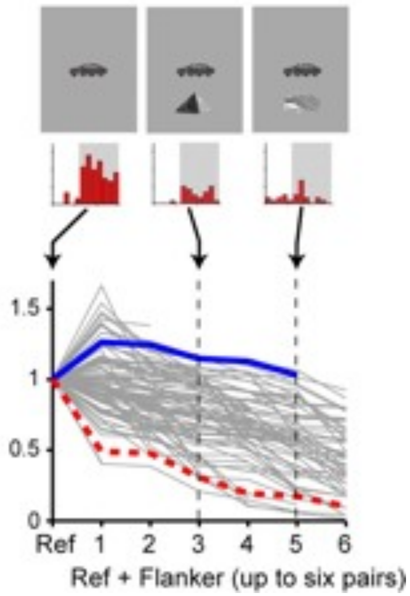
Jhuang et al. 2007



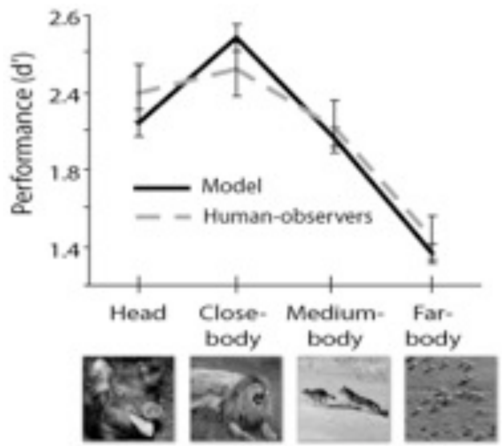
Spiking neuron simulation (dynamic model):

- 9,808 Hodgkin-Huxley neurons and 330,295 synapses.
- 310,000 simulated time steps required 57 seconds.

Extension to attention: dealing with clutter



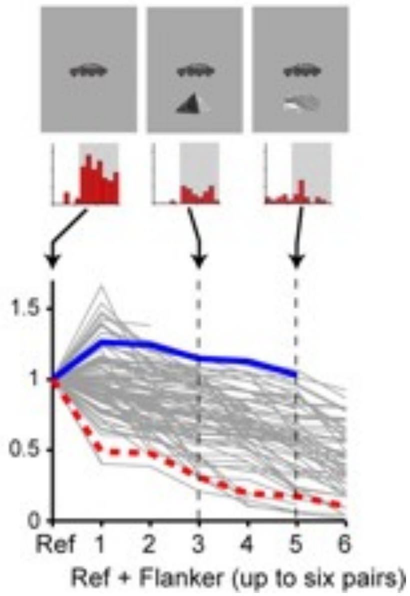
Zoccolan Kouh Poggio DiCarlo 2007



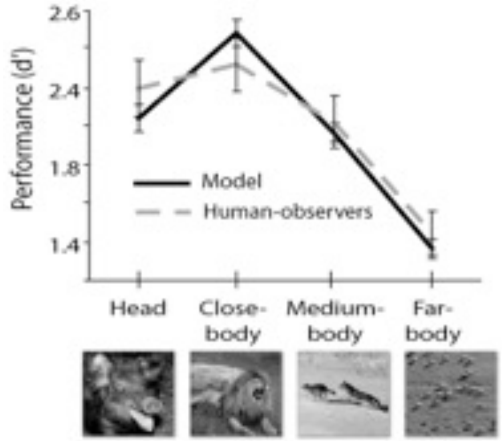
Serre Oliva Poggio 2007

see also Broadbent 1952 1954; Treisman 1960; Treisman & Gelade 1980; Duncan & Desimone 1995; Wolfe, 1997; Tsotsos and many others

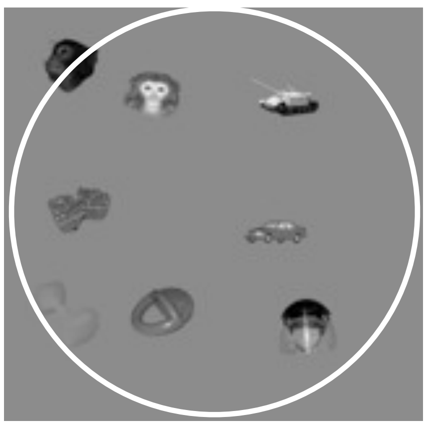
Extension to attention: dealing with clutter

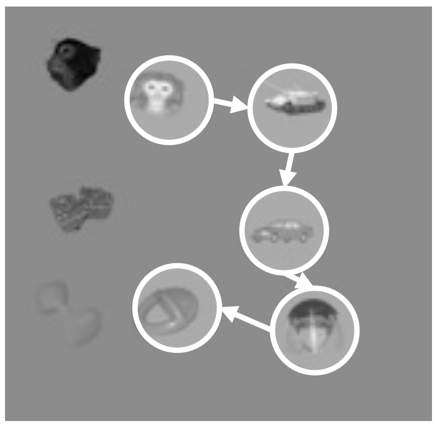


Zoccolan Kouh Poggio DiCarlo 2007



Serre Oliva Poggio 2007

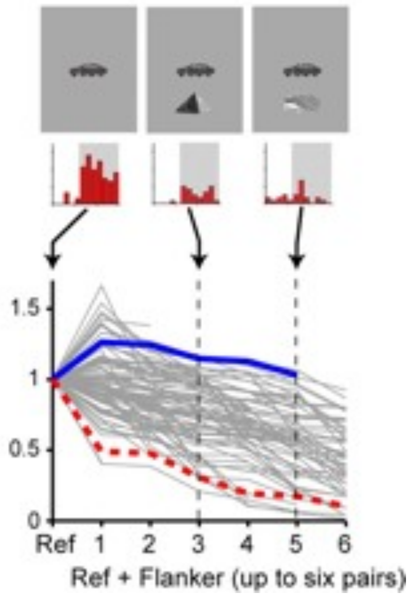




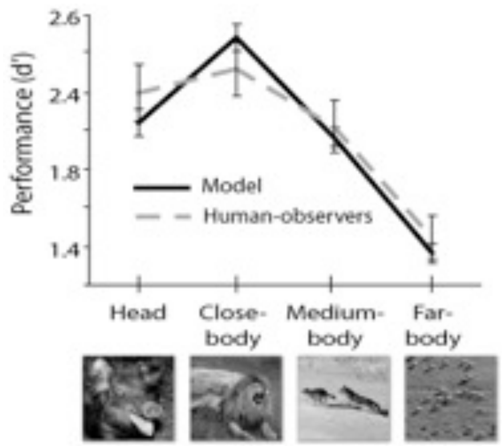
Parallel processing (No attention)

see also Broadbent 1952 1954; Treisman 1960; Treisman & Gelade 1980; Duncan & Desimone 1995; Wolfe, 1997; Tsotsos and many others

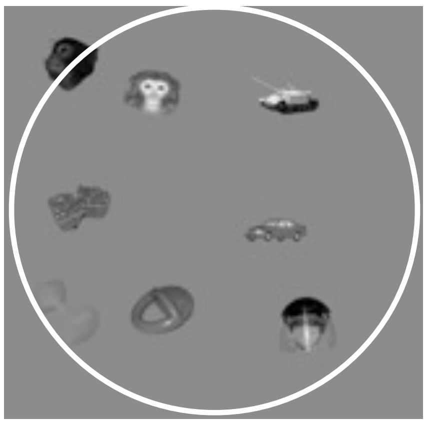
Extension to attention: dealing with clutter



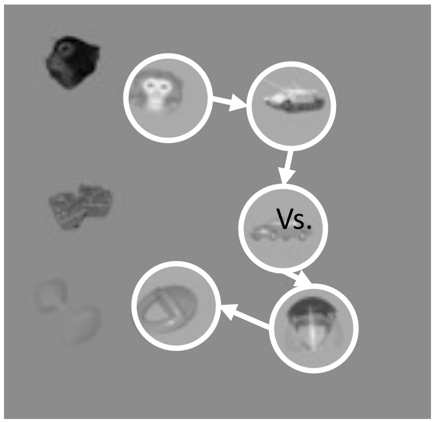
Zoccolan Kouh Poggio DiCarlo 2007



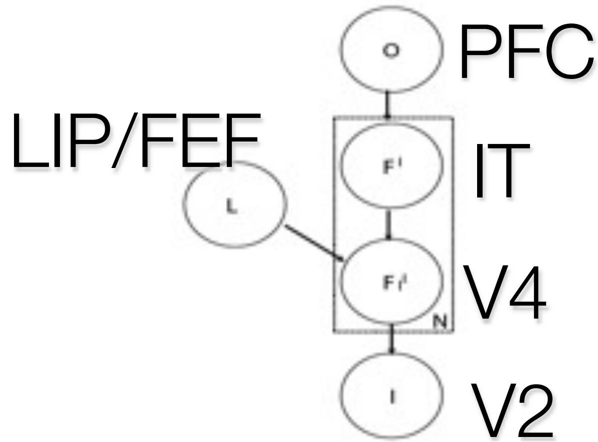
Serre Oliva Poggio 2007



Parallel processing (No attention)



Serial processing (With attention)



see also Broadbent 1952 1954; Treisman 1960; Treisman & Gelade 1980; Duncan & Desimone 1995; Wolfe, 1997; Tsotsos and many others

Readings on the work with many relevant references

A detailed description of much of the work is in the "supermemo" at

http://cbcl.mit.edu/projects/cbcl/publications/ai-publications/2005/AIM-2005-036.pdf

Other recent publications and references can be found at

http://cbcl.mit.edu/publications/index-pubs.html

Collaborators in recent work

F. Anselmi, G. Spigler, J. Mutch, L. Rosasco,
H. Jhuang, C. Tan, J. Leibo, N. Edelman,
E. Meyers, S. Ullman, B. Desimone, S. Smale,

Also: T. Serre, S. Chikkerur, A. Wibisono, J. Bouvrie, M. Kouh, M. Riesenhuber, J. DiCarlo, E. Miller, A. Oliva, C. Koch, A. Caponnetto ,D. Walther, C. Cadieu, U. Knoblich, T. Masquelier, S. Bileschi, L. Wolf, E. Connor, D. Ferster, I. Lampl, S. Chikkerur, G. Kreiman, N. Logothetis