Vision and visual neuroscience II

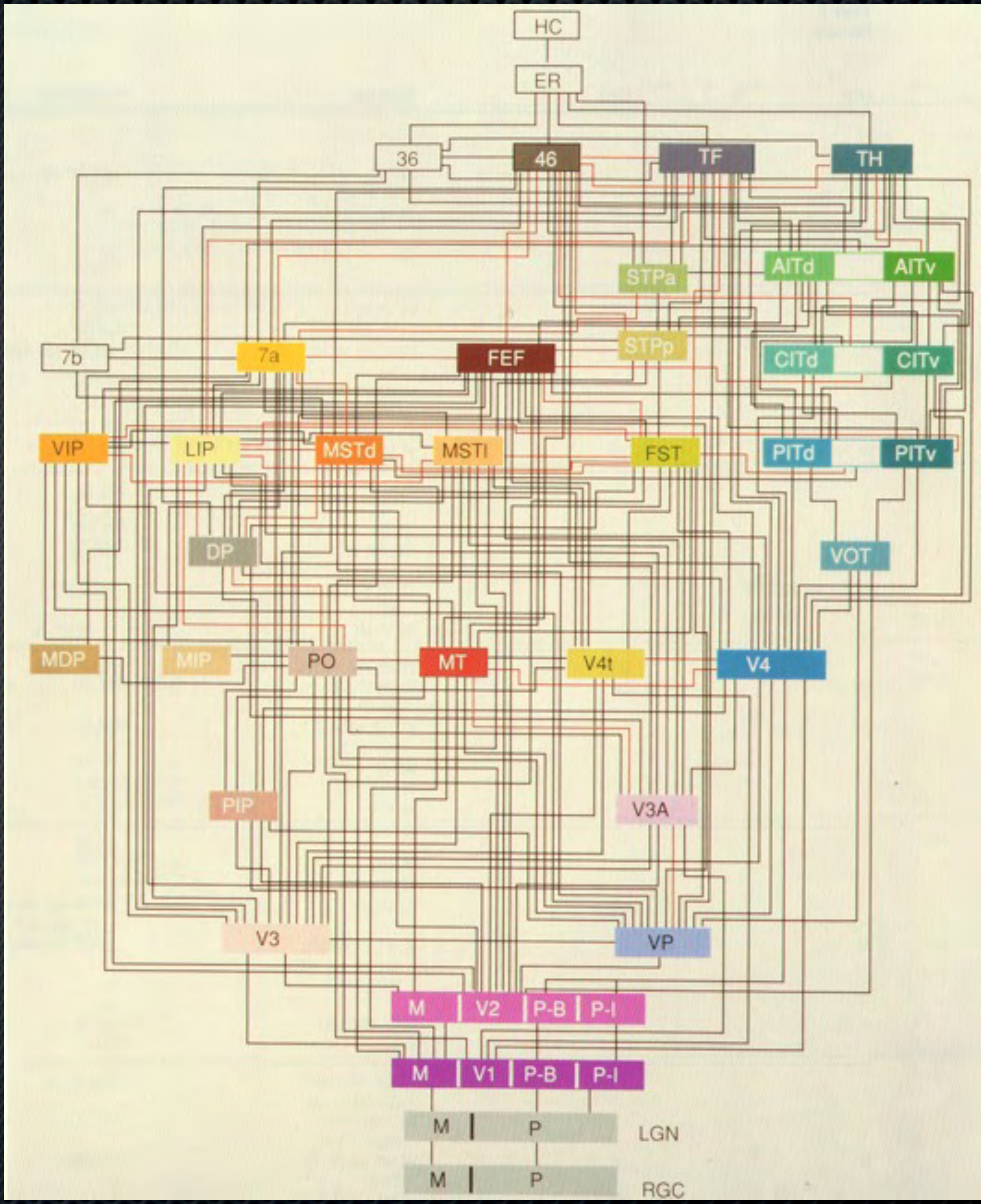
Thomas Serre & Tomaso Poggio

McGovern Institute for Brain Research
Department of Brain & Cognitive Sciences
Massachusetts Institute of Technology

Past lecture

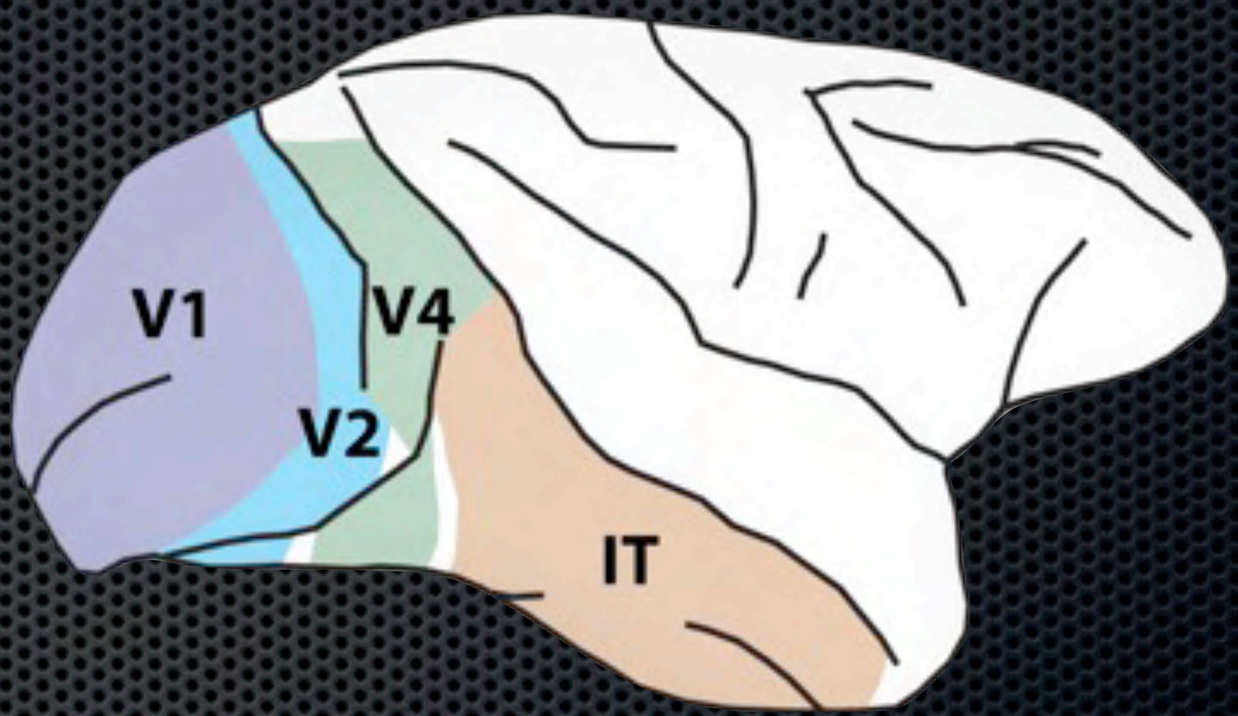
- Problem of visual recognition and visual cortex
- Historical background
- Neurons and areas in the visual system
- Feedforward hierarchical models

Hierarchical anatomical organization



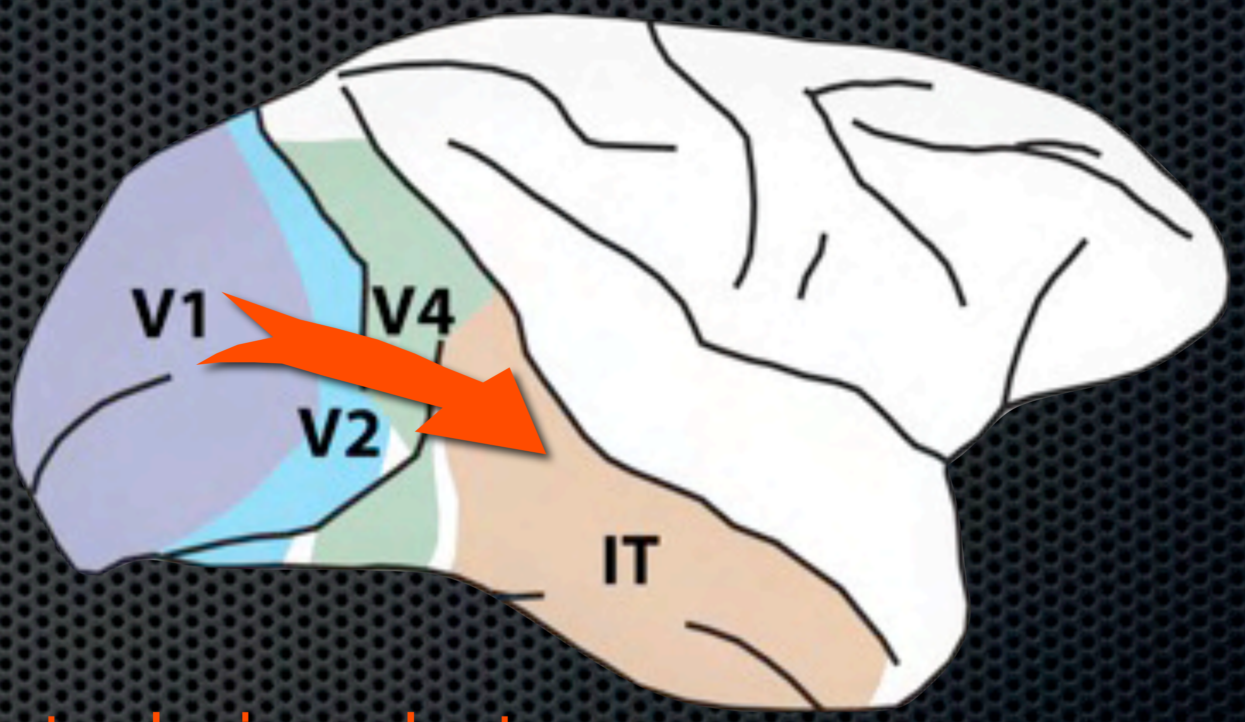
Felleman & van Essen 1991

Object recognition in the visual cortex



source: Jim DiCarlo

Object recognition in the visual cortex

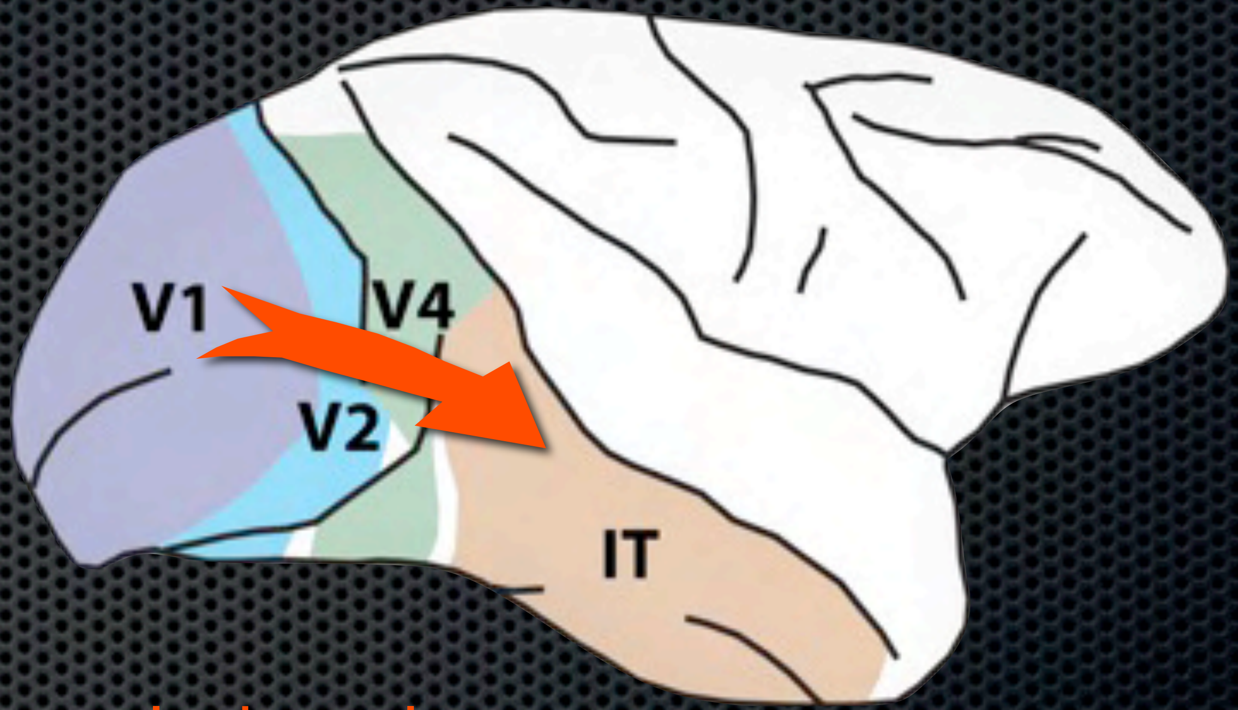


Ventral visual stream

source: Jim DiCarlo

Object recognition in the visual cortex

- Hierarchical architecture:

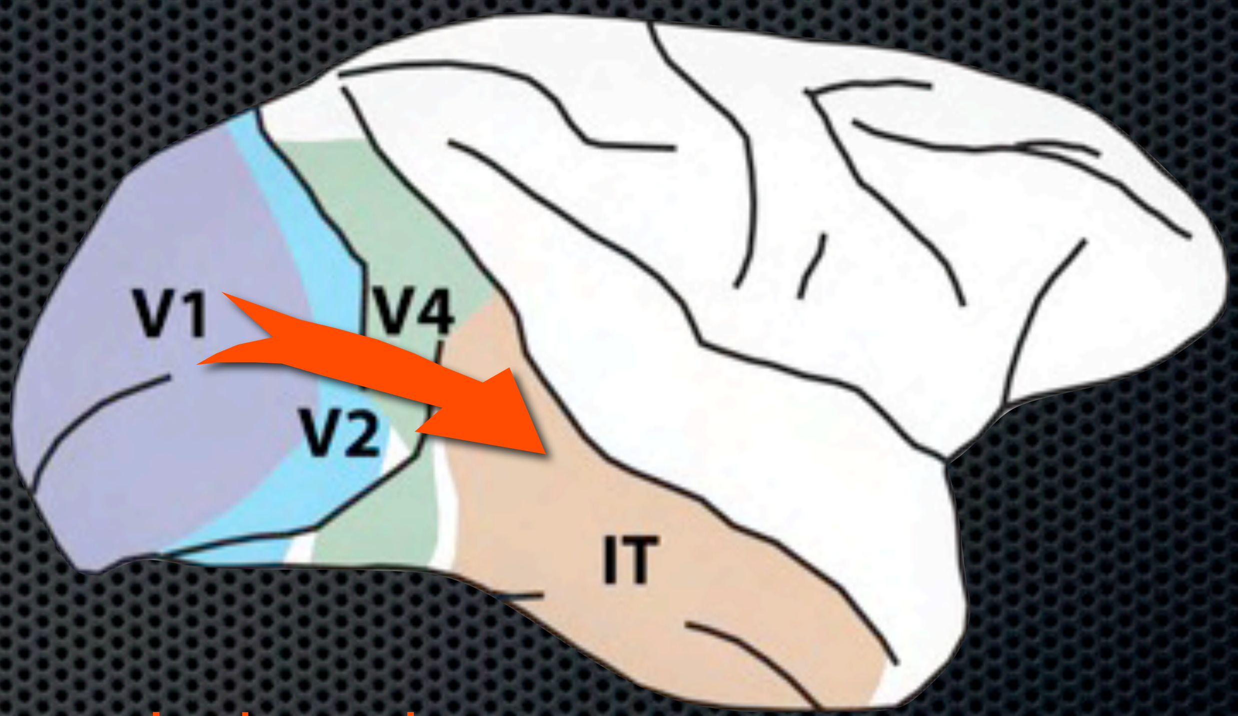


Ventral visual stream

source: Jim DiCarlo

Object recognition in the visual cortex

- Hierarchical architecture:
  - Latencies

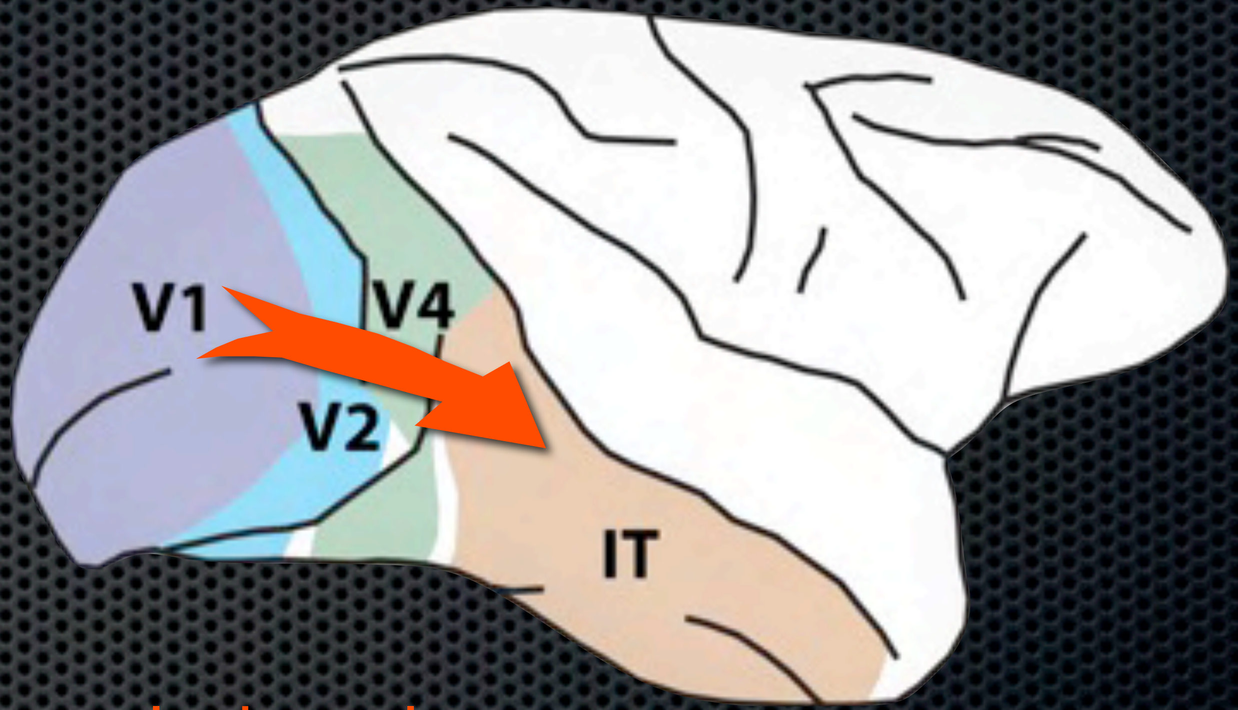


Ventral visual stream

source: Jim DiCarlo

Object recognition in the visual cortex

- Hierarchical architecture:
  - Latencies
  - Anatomy

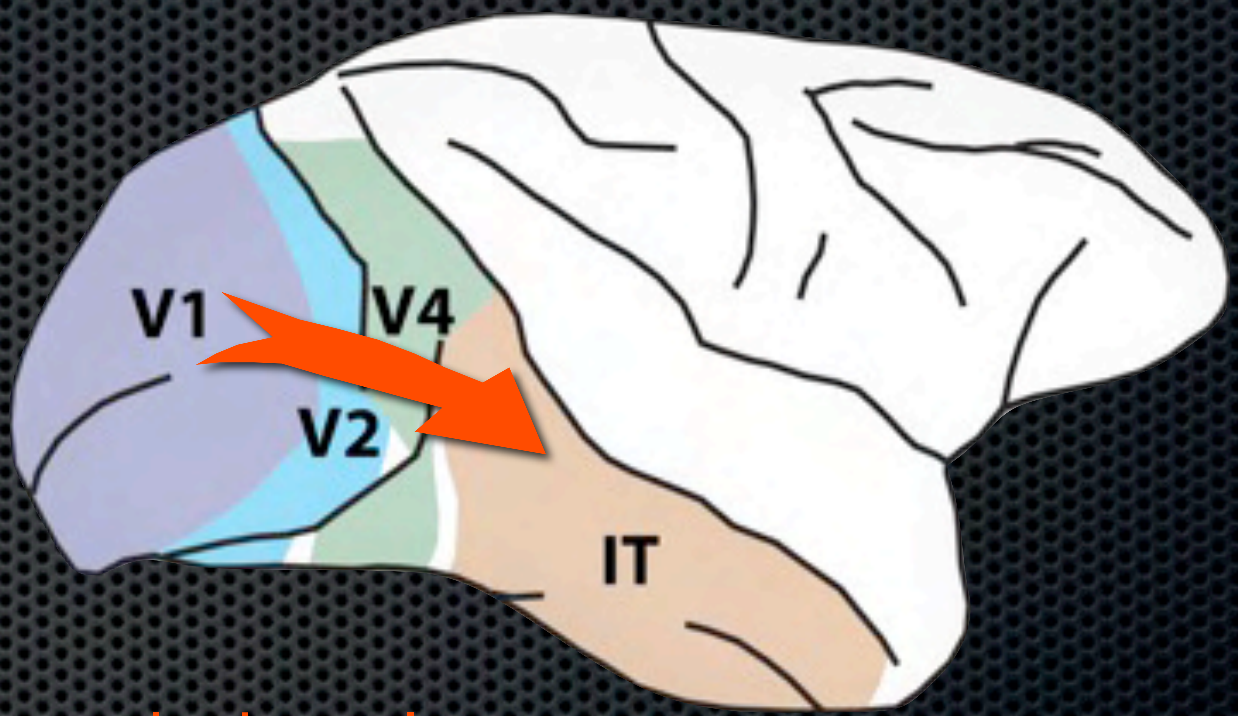


Ventral visual stream

source: Jim DiCarlo

Object recognition in the visual cortex

- Hierarchical architecture:
  - Latencies
  - Anatomy
  - Function

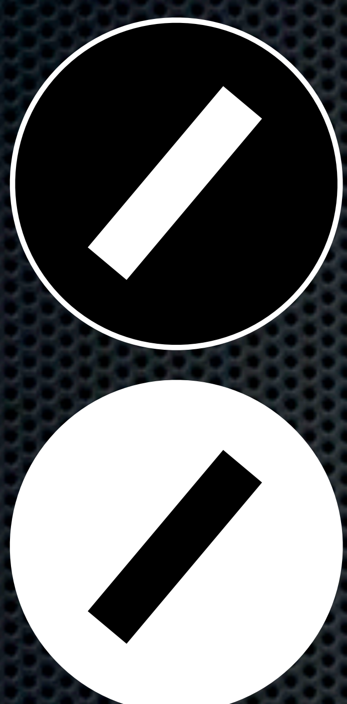


Ventral visual stream

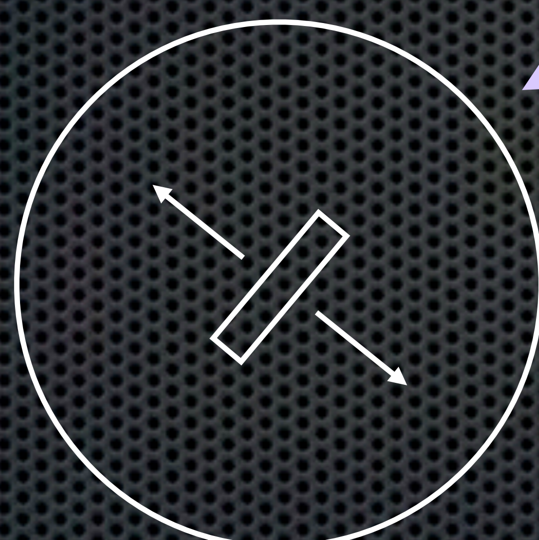
source: Jim DiCarlo

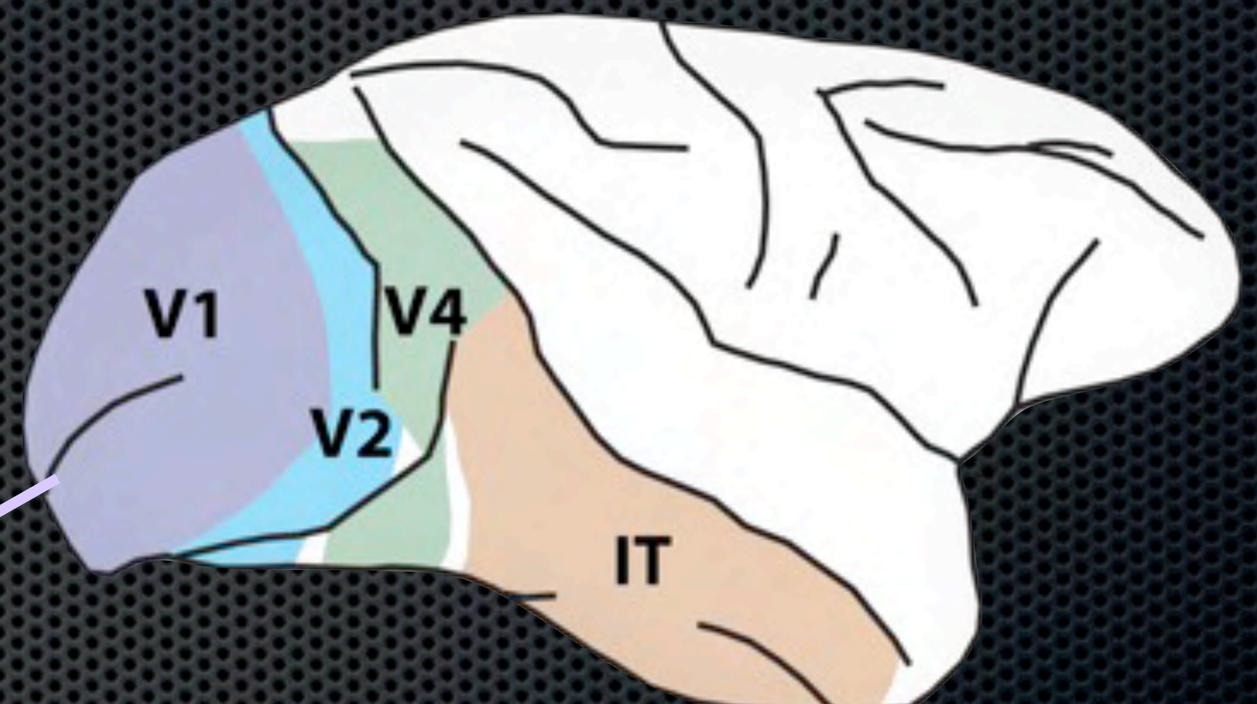
Object recognition in the visual cortex

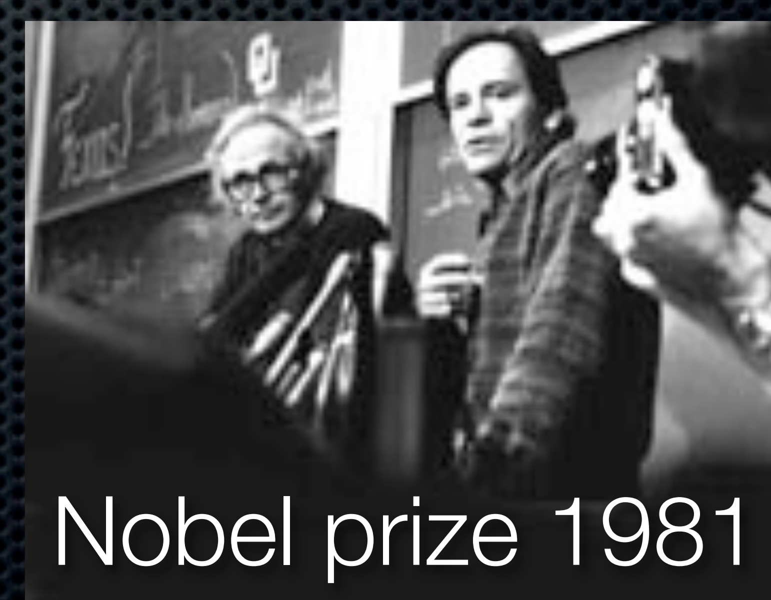
simple cells



complex cells

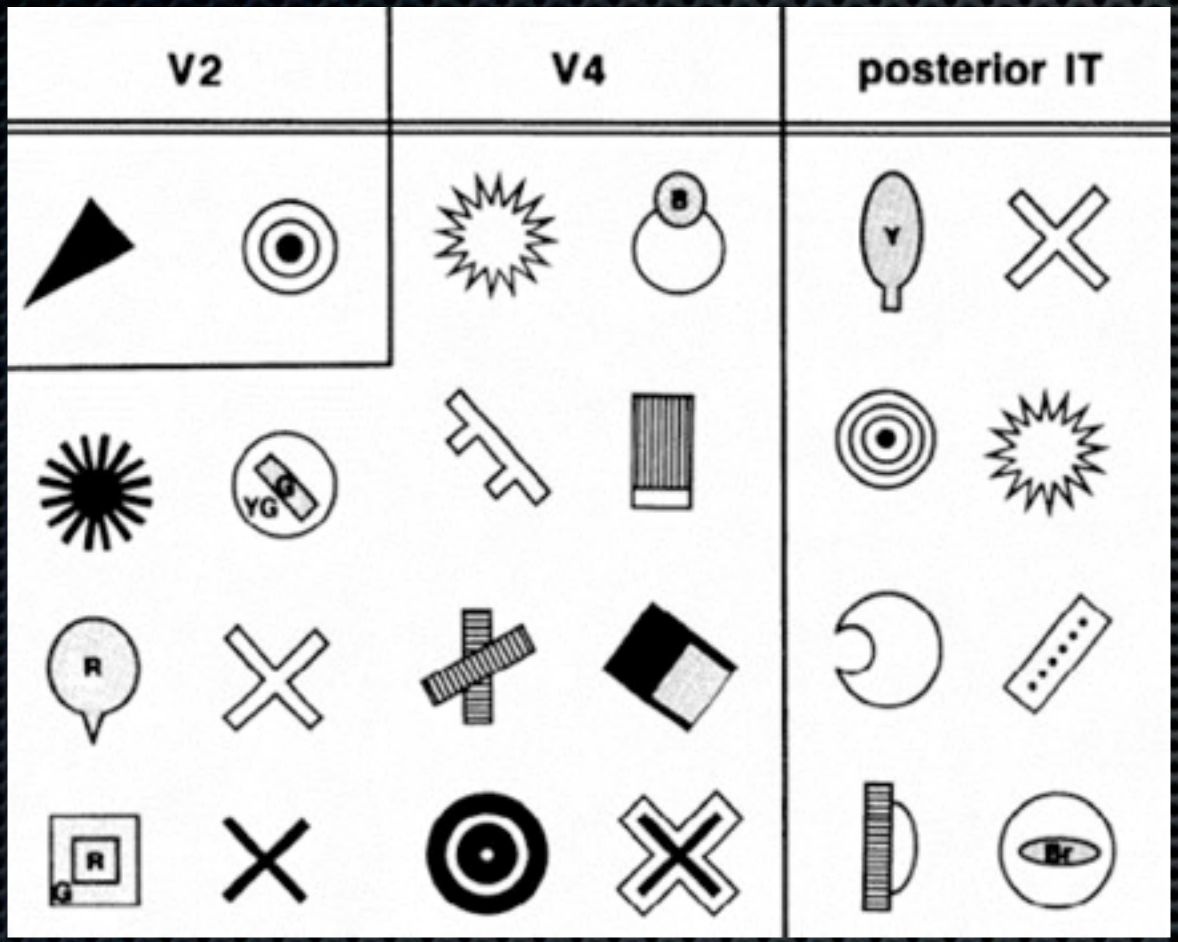




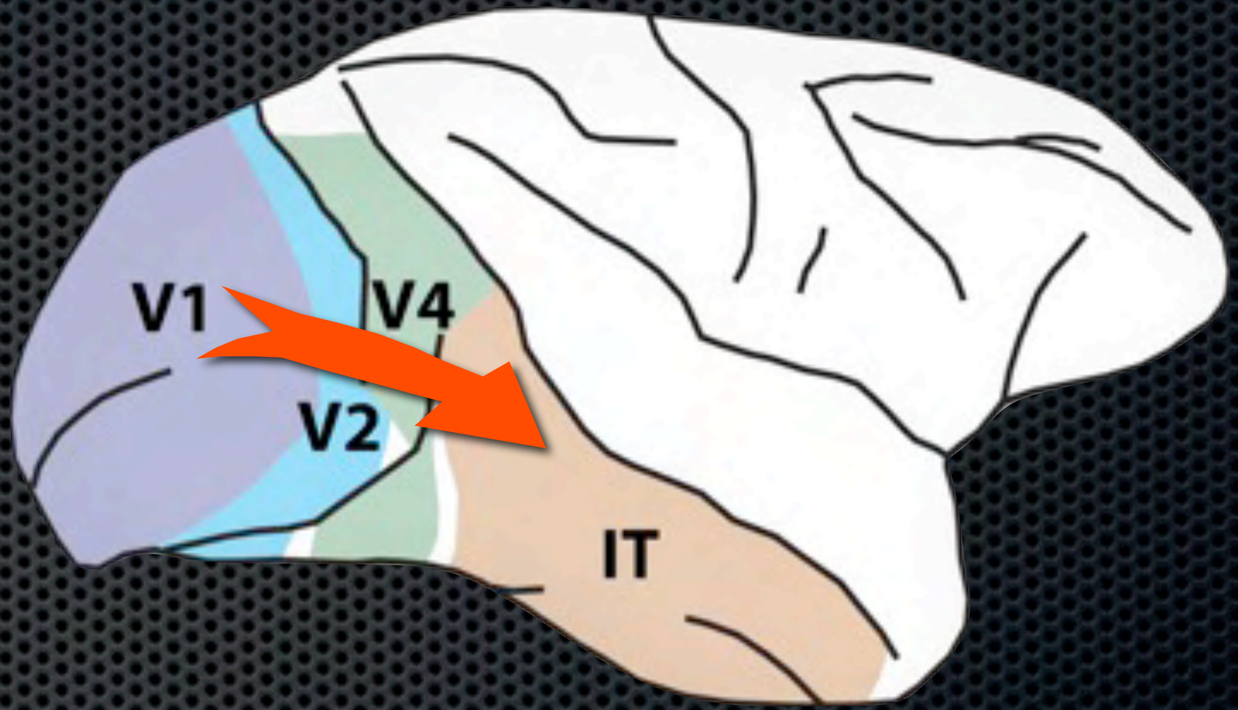


Hubel & Wiesel 1959, 1962, 1965, 1968

Object recognition in the visual cortex



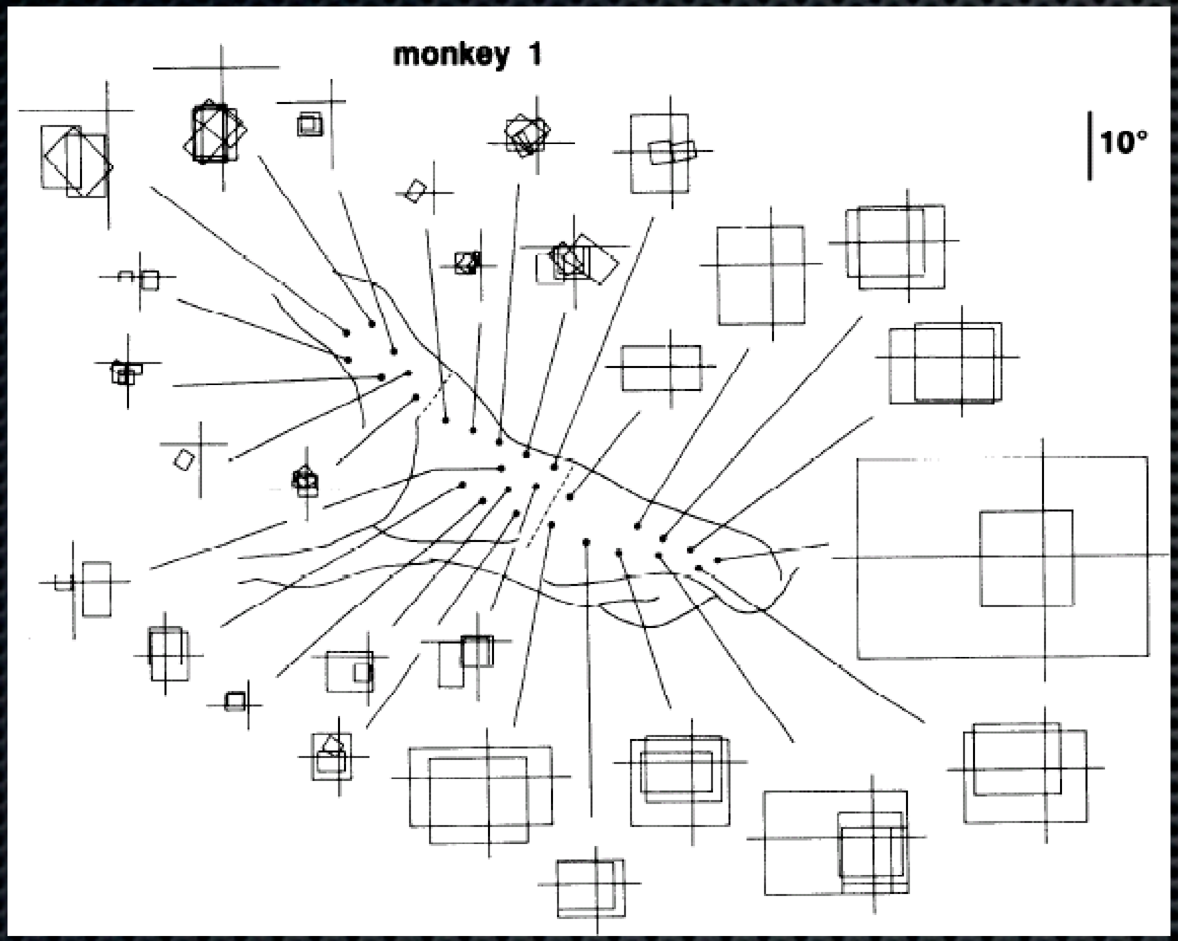
Kobatake & Tanaka 1994



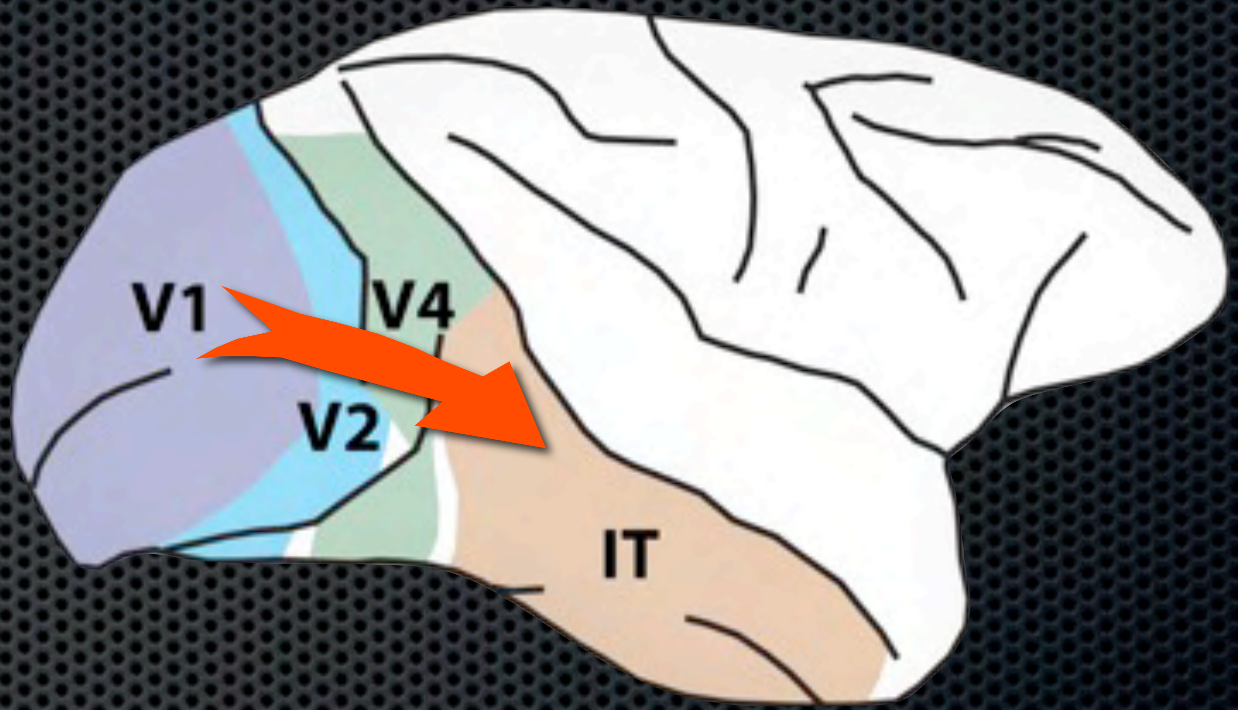
gradual increase in complexity of preferred stimulus

see also Oram & Perrett 1993; Sheinberg & Logothetis 1996; Gallant et al 1996; Riesenhuber & Poggio 1999

Object recognition in the visual cortex



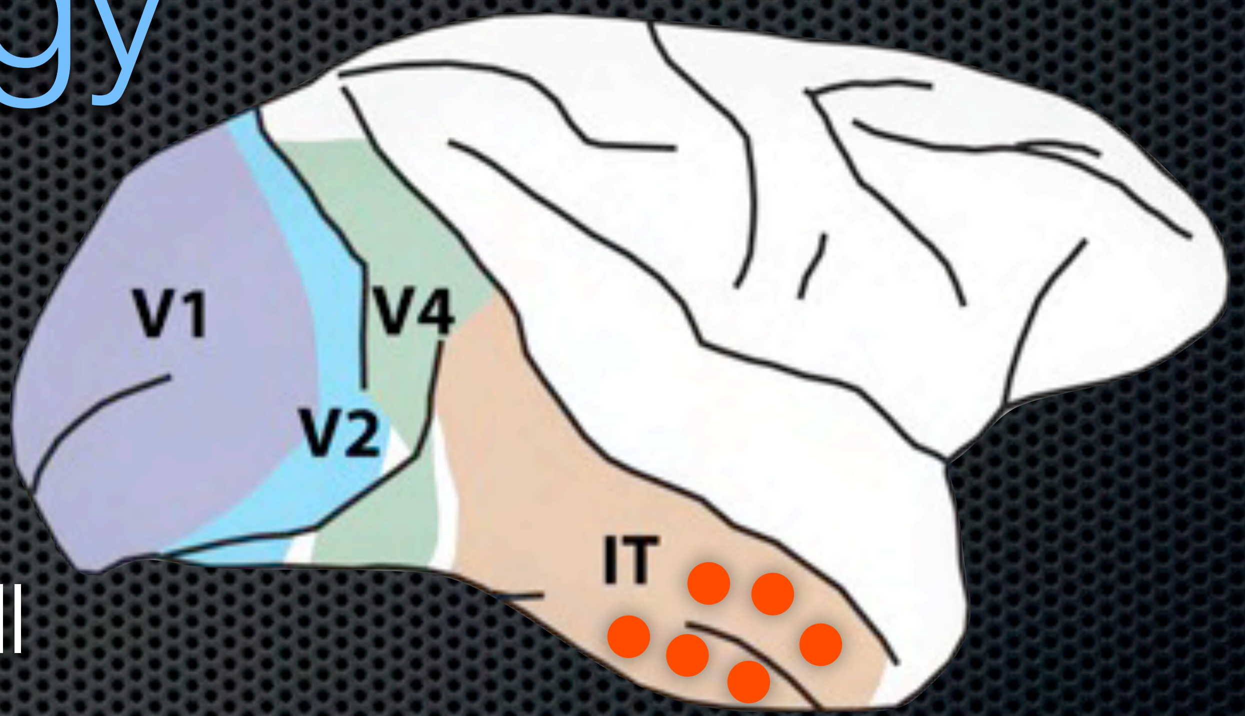
Kobatake & Tanaka 1994



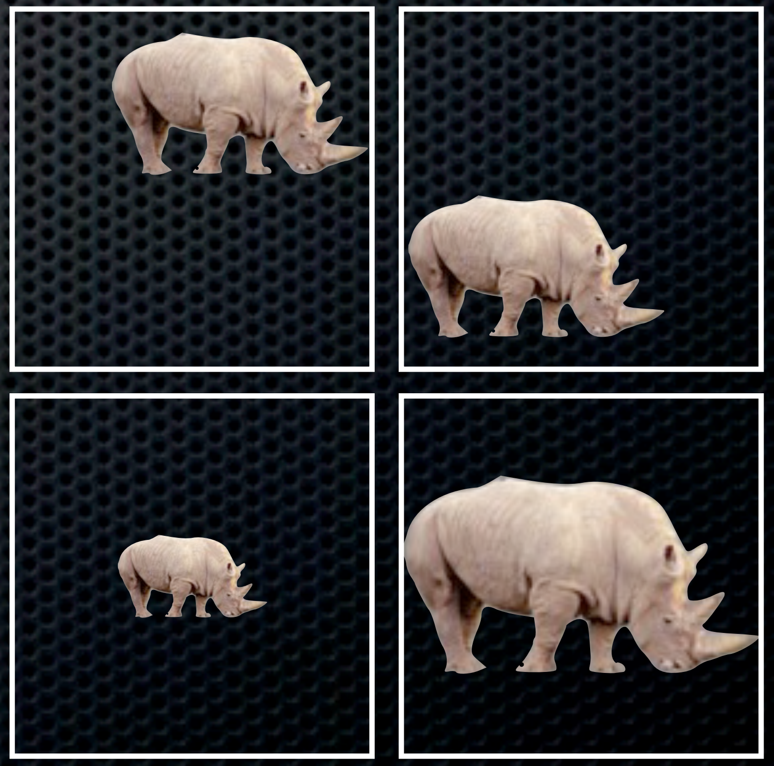
Parallel increase in invariance properties (position and scale) of neurons

see also Oram & Perrett 1993; Sheinberg & Logothetis 1996; Gallant et al 1996; Riesenhuber & Poggio 1999

Rapid recognition: monkey electrophysiology



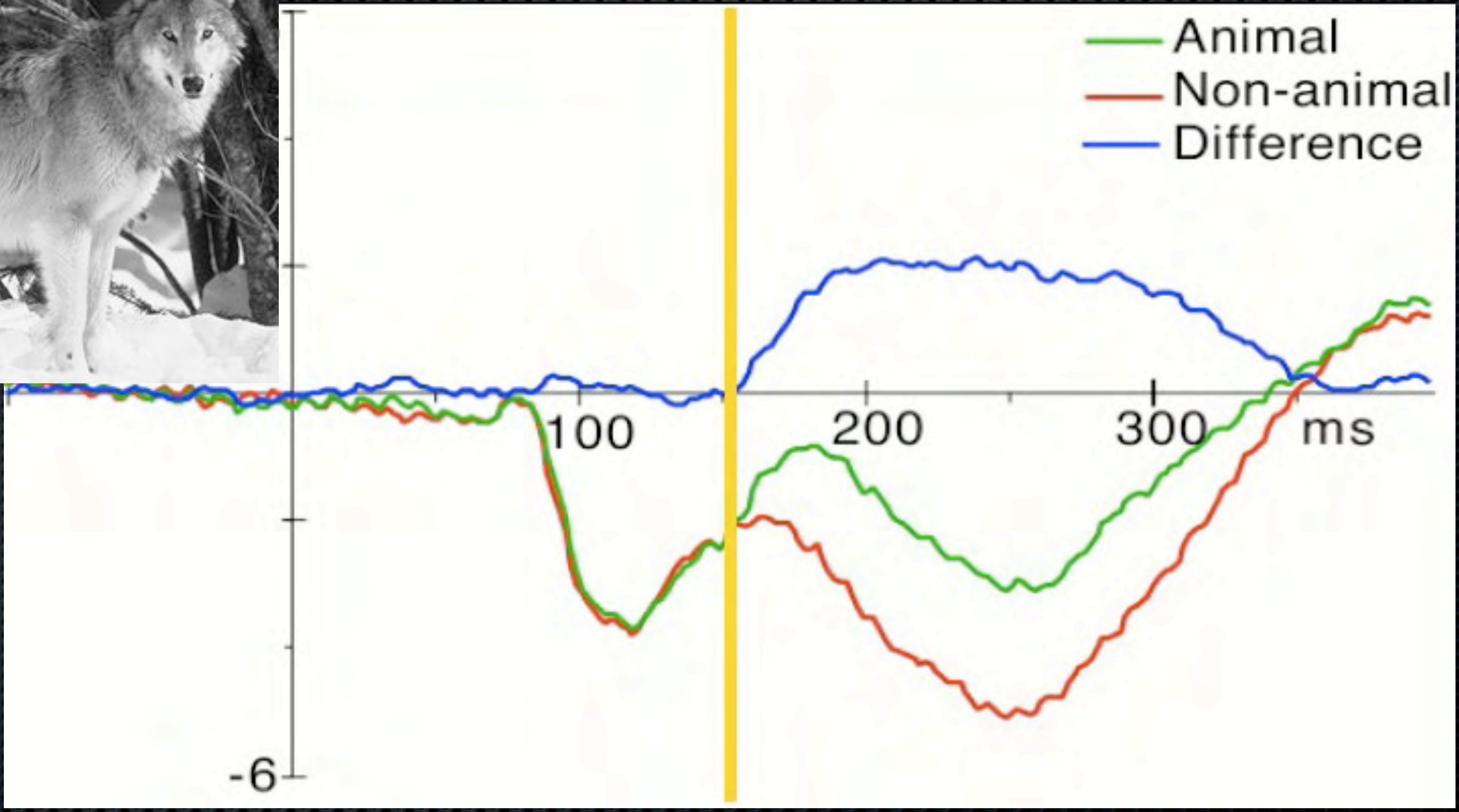
- Robust invariant readout of category information from small population of neurons
- Single spikes after response onset carry most of the information



Hung* Kreiman* Poggio & DiCarlo 2005

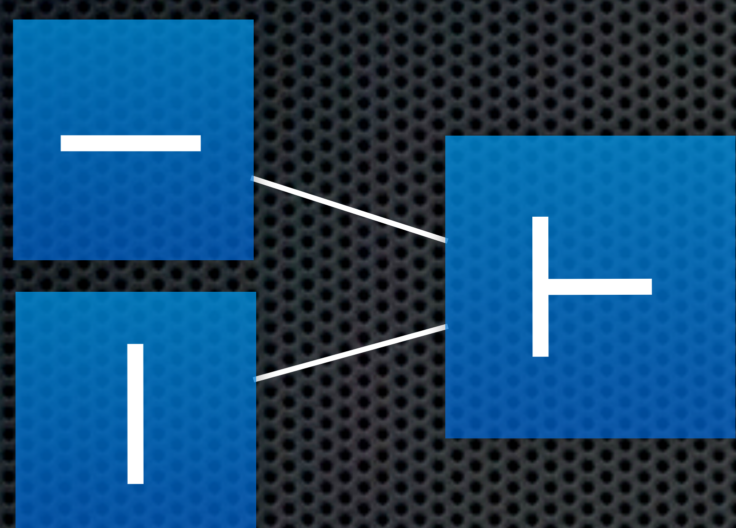
Rapid recognition: human behavior

Thorpe et al '96



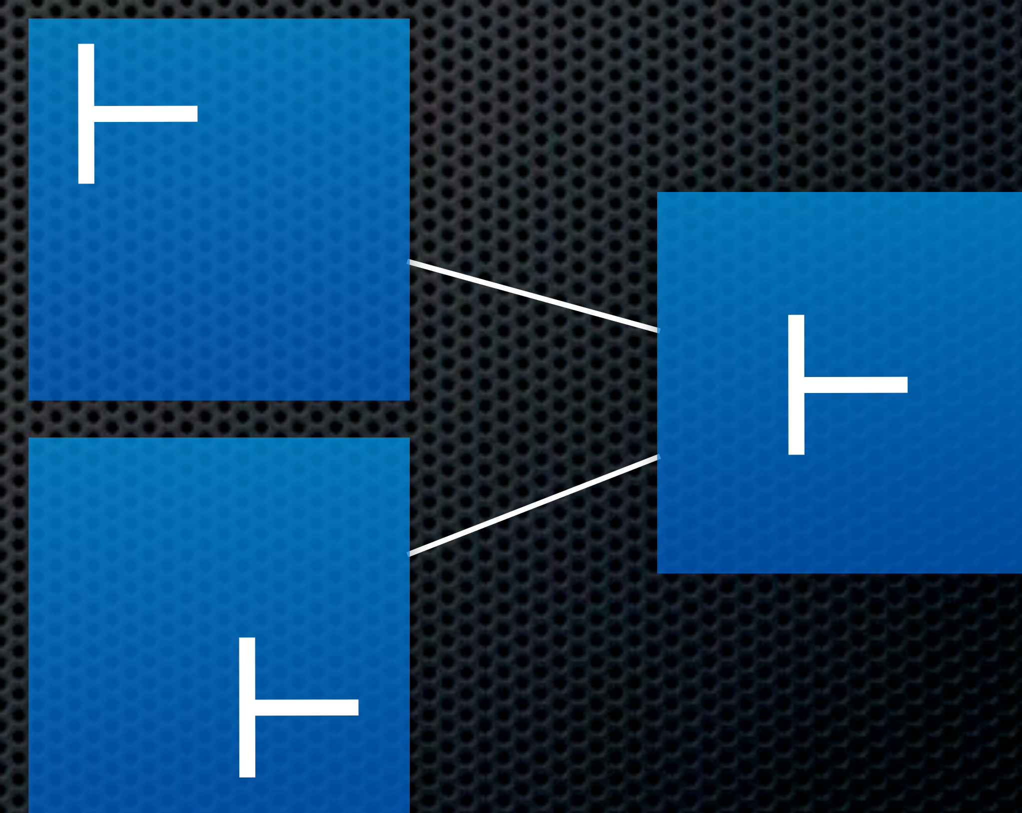
Computational considerations

Simple units



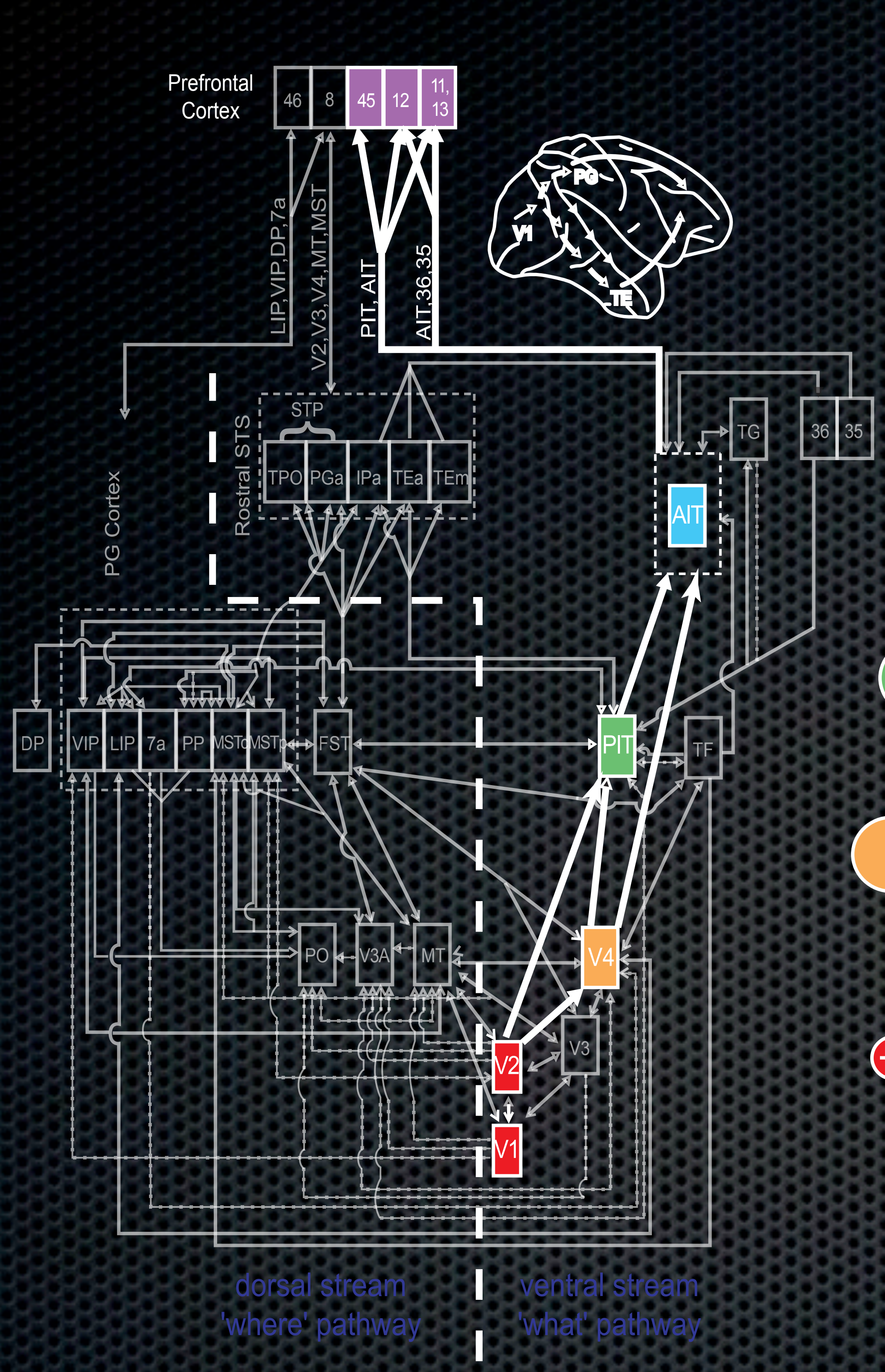
Template matching
Gaussian-like tuning
~ “AND”

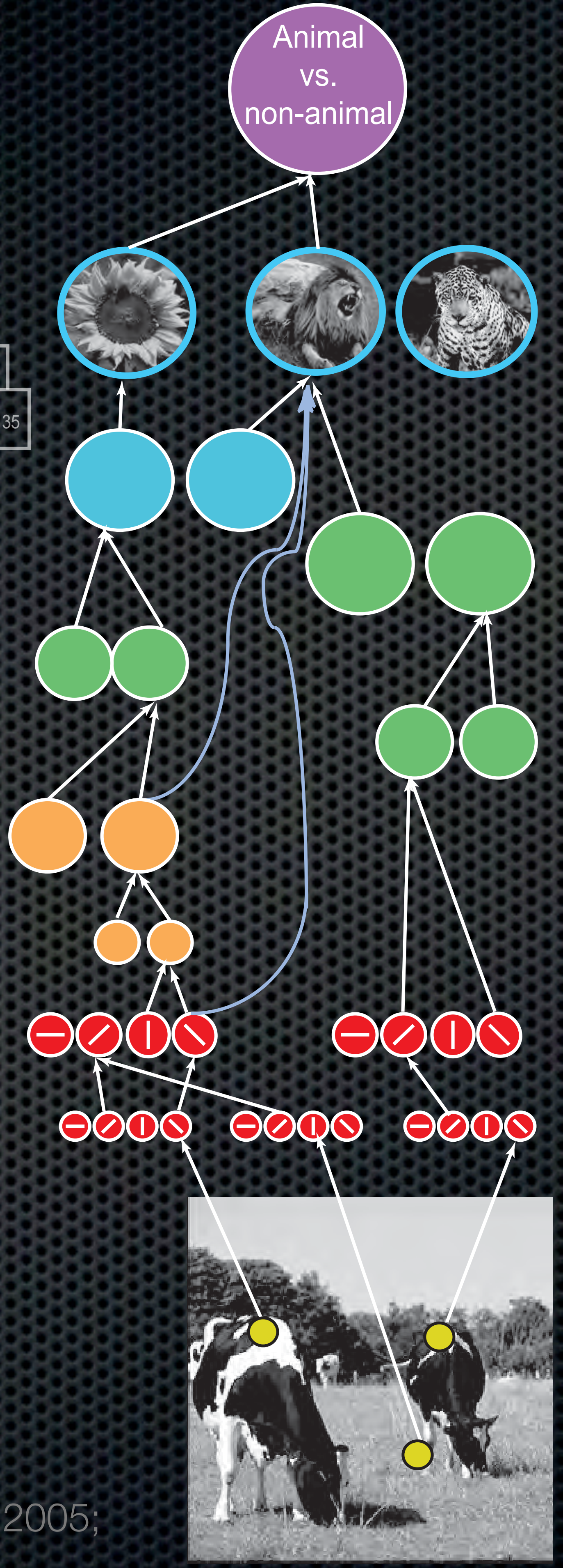
Complex units



Invariance
max-like operation
~”OR”

Riesenhuber & Poggio 1999 (building on Fukushima 1980 and Hubel & Wiesel 1962)





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

✦V1:

- Simple and complex cells tuning properties (Schiller et al 1976; Hubel & Wiesel 1965; Devalois et al 1982)
- MAX operation in subset of complex cells (Lampl et al 2004)

✦V4:

- Tuning for two-bar stimuli (Reynolds Chelazzi & Desimone 1999)
- MAX operation (Gawne et al 2002)
- Two-spot interaction (Freiwald et al 2005)
- Tuning for boundary conformation (Pasupathy & Connor 2001)
- Tuning for Cartesian and non-Cartesian gratings (Gallant et al 1996)

✦IT:

- Tuning and invariance properties (Logothetis et al 1995)
- Differential role of IT and PFC in categorization (Freedman et al 2001 2002 2003)
- Read out data (Hung Kreiman Poggio & DiCarlo 2005)
- Average effect in IT (Zoccolan Cox & DiCarlo 2005; Zoccolan Kouh Poggio & DiCarlo in press)

✦Human behavior:

- Rapid animal categorization (Serre Oliva Poggio 2007)

This lecture

This lecture

1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world

This lecture

1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world
2. Rapid recognition and feedforward processing:
   - Predicting human performance
   - “Clutter problem”

This lecture

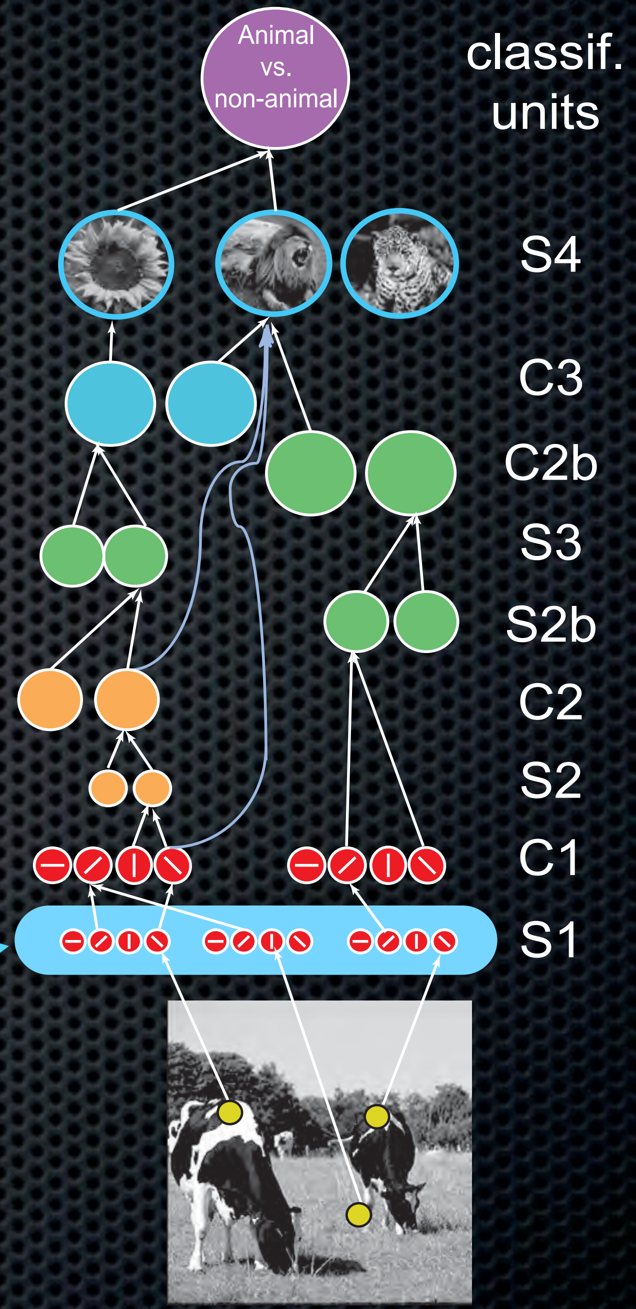
1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world
2. Rapid recognition and feedforward processing:
   - Predicting human performance
   - “Clutter problem”
3. Beyond feedforward processing:
   - Top-down cortical feedback and attention to solve the “clutter problem”
   - Predicting human eye movements

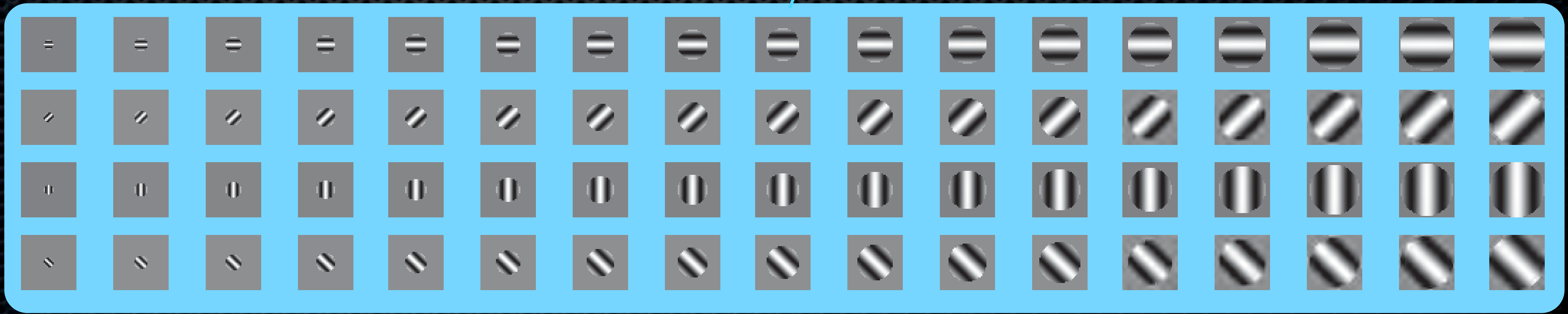
This lecture

1. Learning a loose hierarchy of image fragments
    - The algorithm
    - Recognition in the real-world
2. Rapid recognition and feedforward processing:
    - Predicting human performance
    - “Clutter problem”
3. Beyond feedforward processing:
    - Top-down cortical feedback and attention to solve the “clutter problem”
    - Predicting human eye movements

S1 units

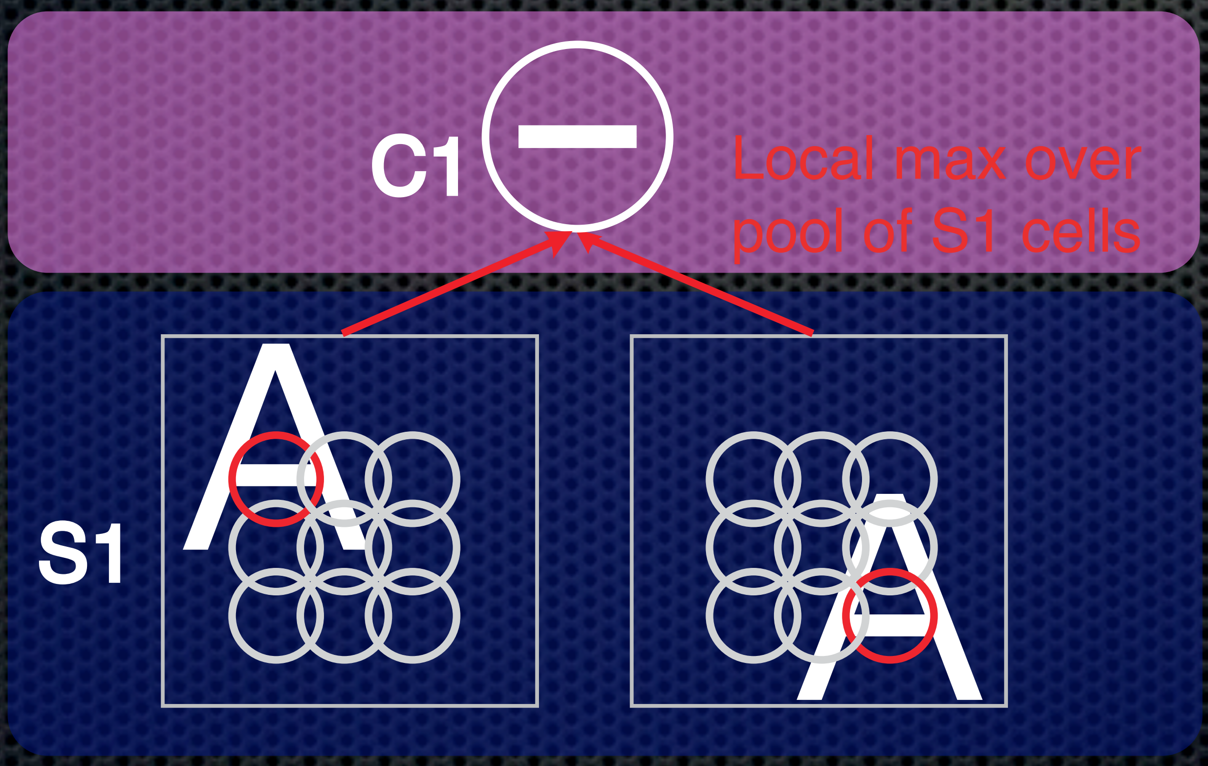
- Gabor filters
- Parameters fit to V1 data (Serre & Riesenhuber 2004)
  - 17 spatial frequencies (=scales)
  - 4 orientations

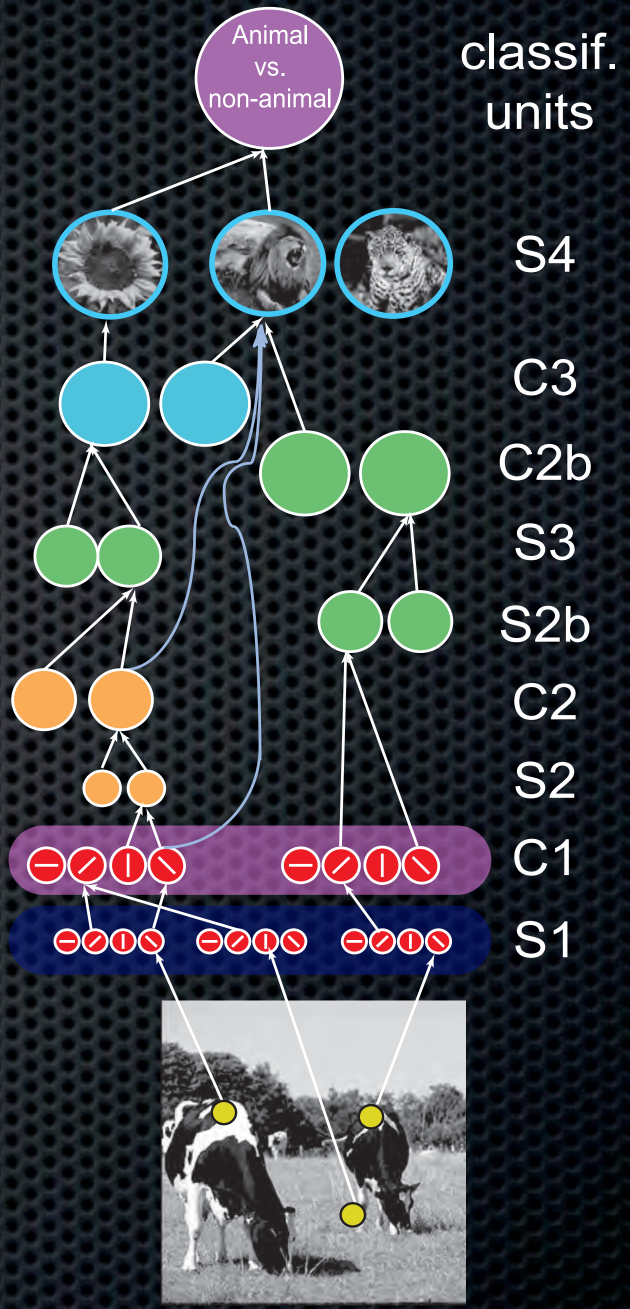




C1 units

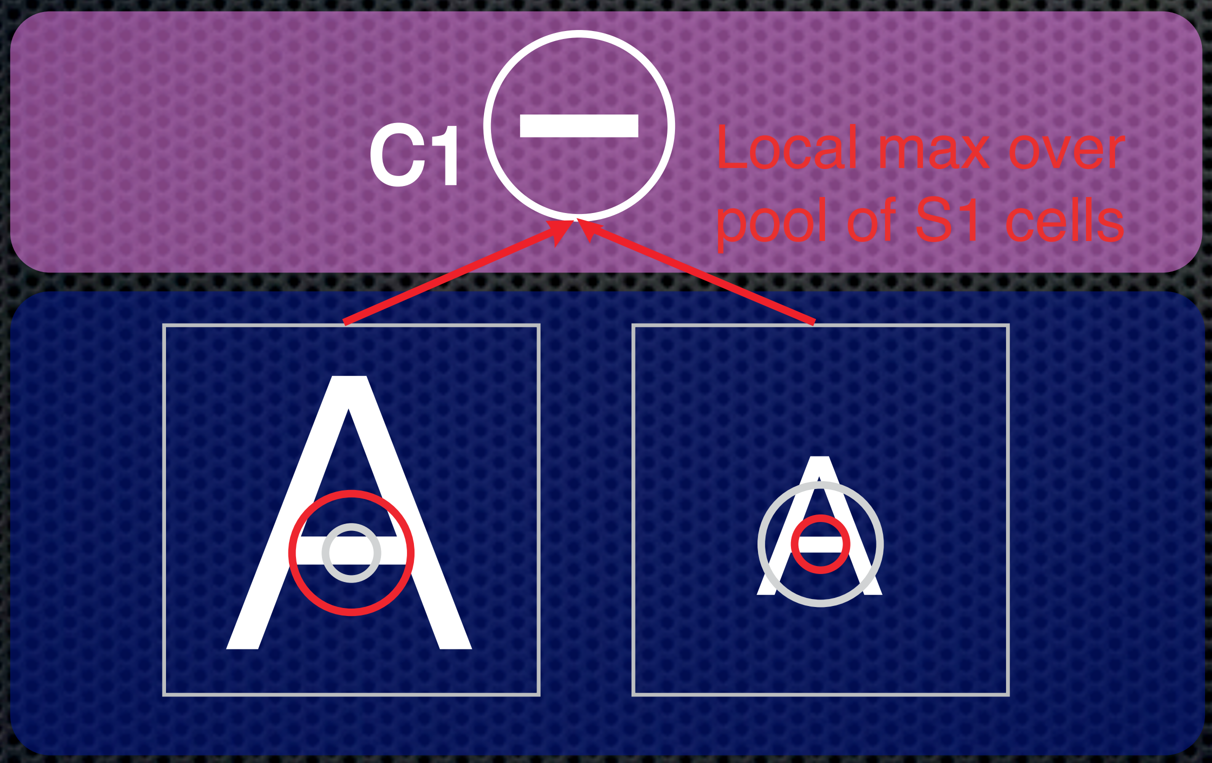
Increase in tolerance to **position** (and in RF size)

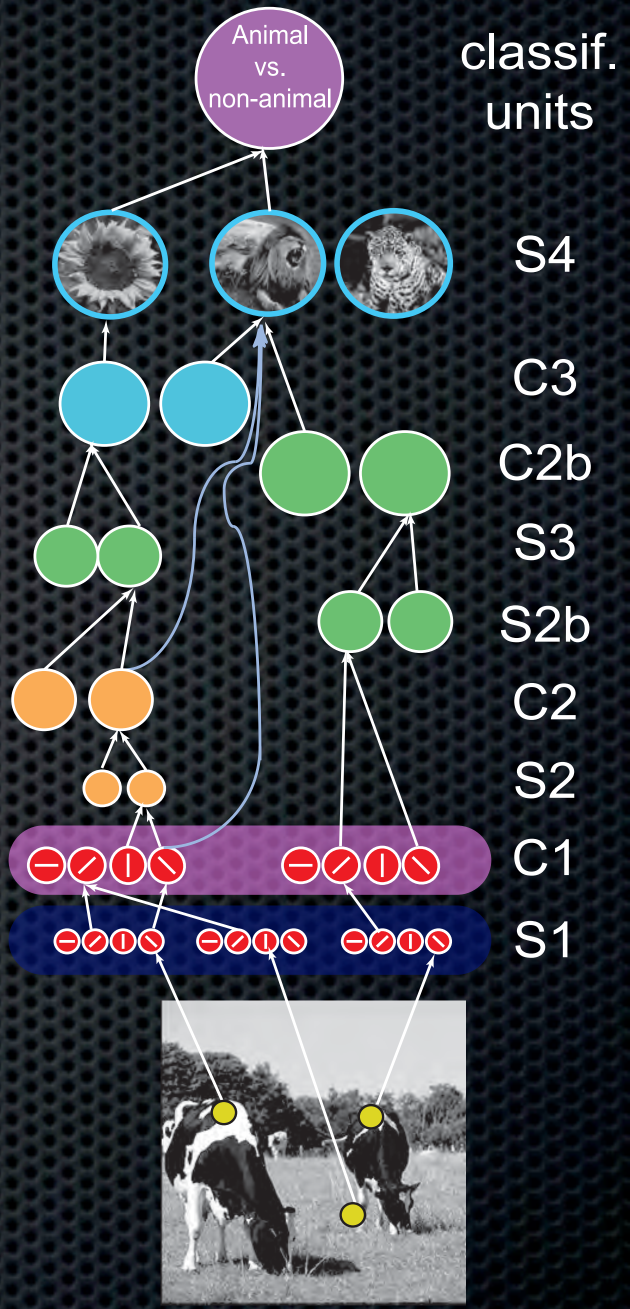




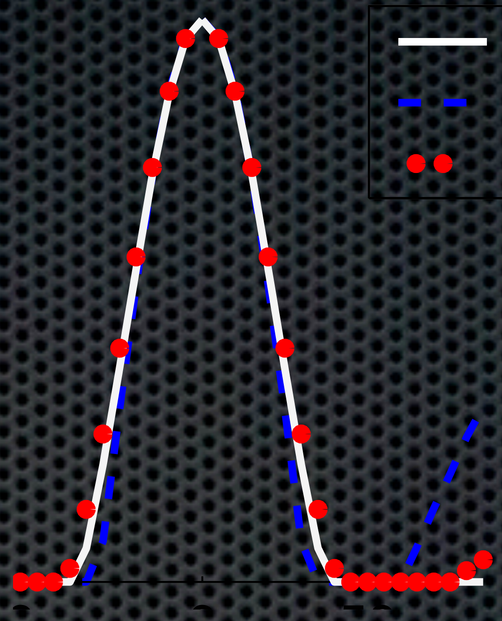
C1 units

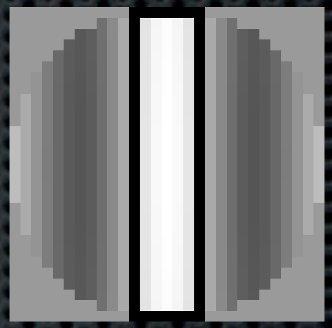
Increase in tolerance to
scale





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

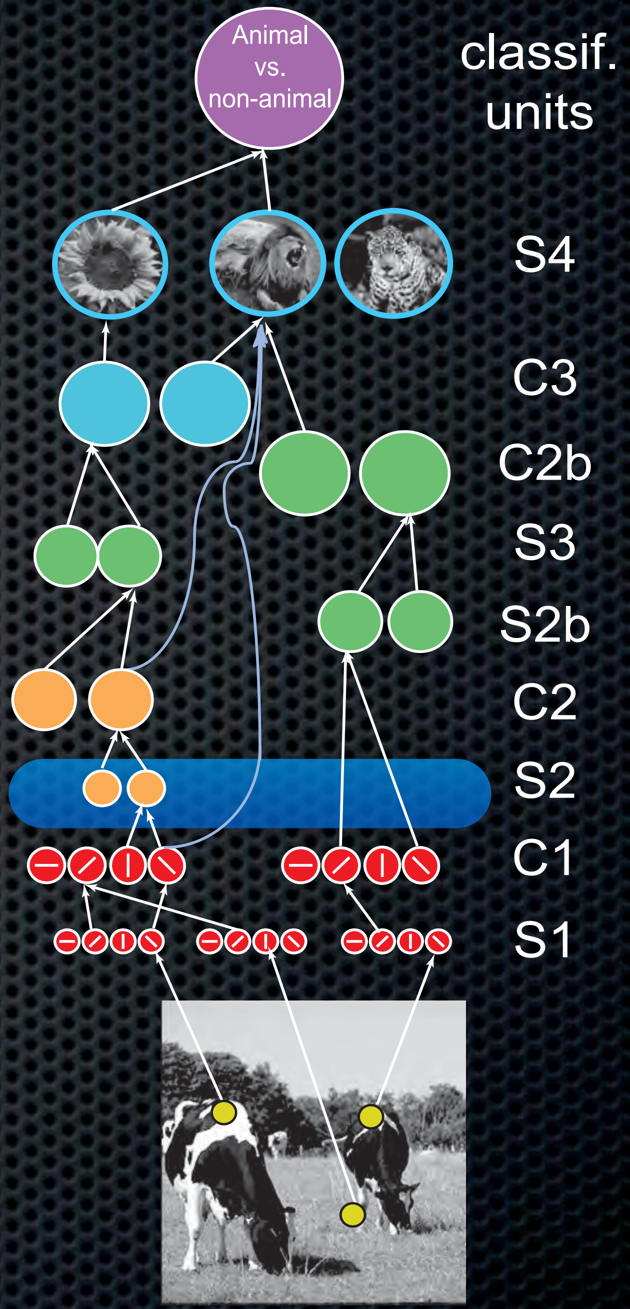




Serre & Riesenhuber 2004

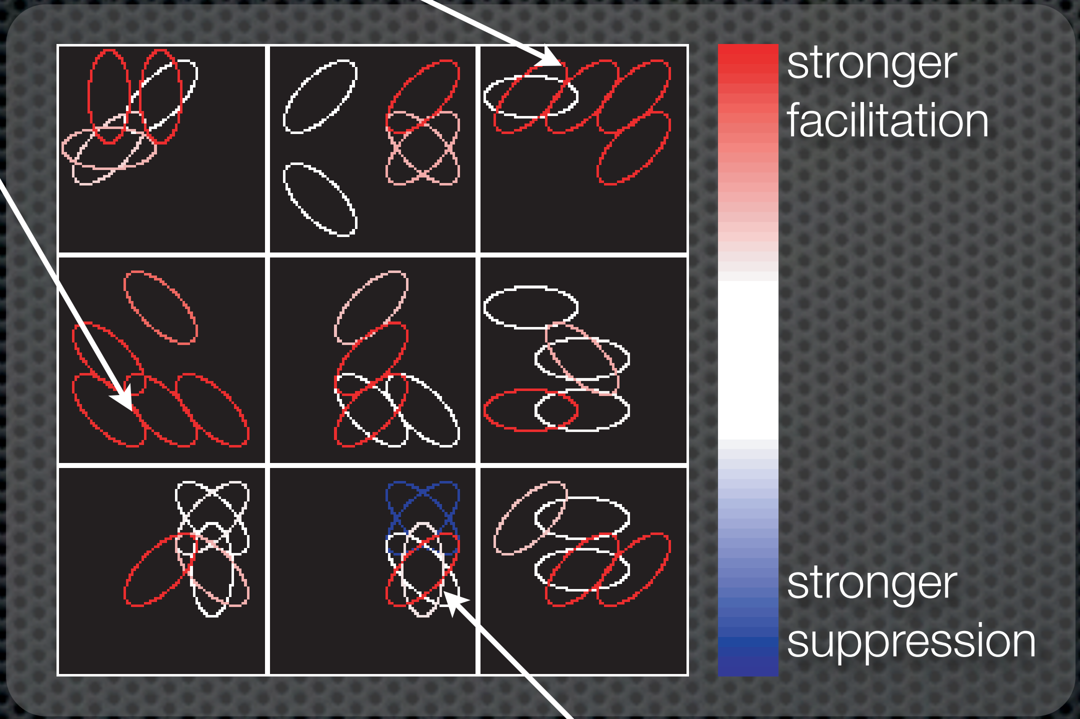
S2 units

- 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)

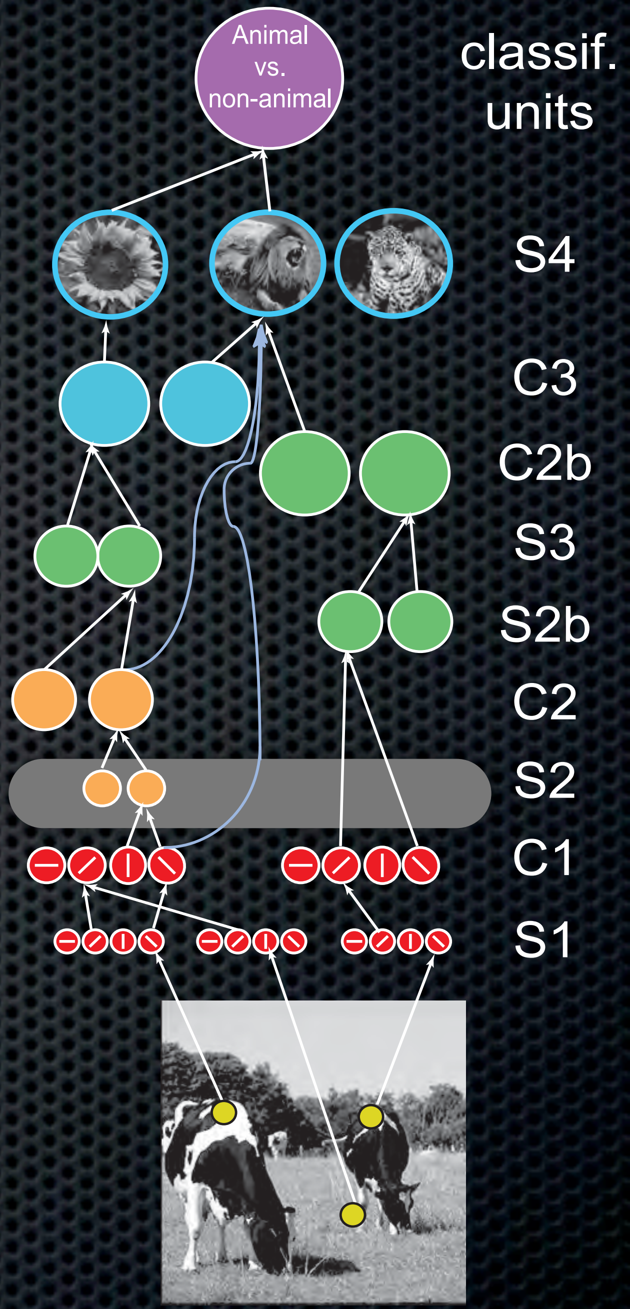


S2 units

homogenous fields



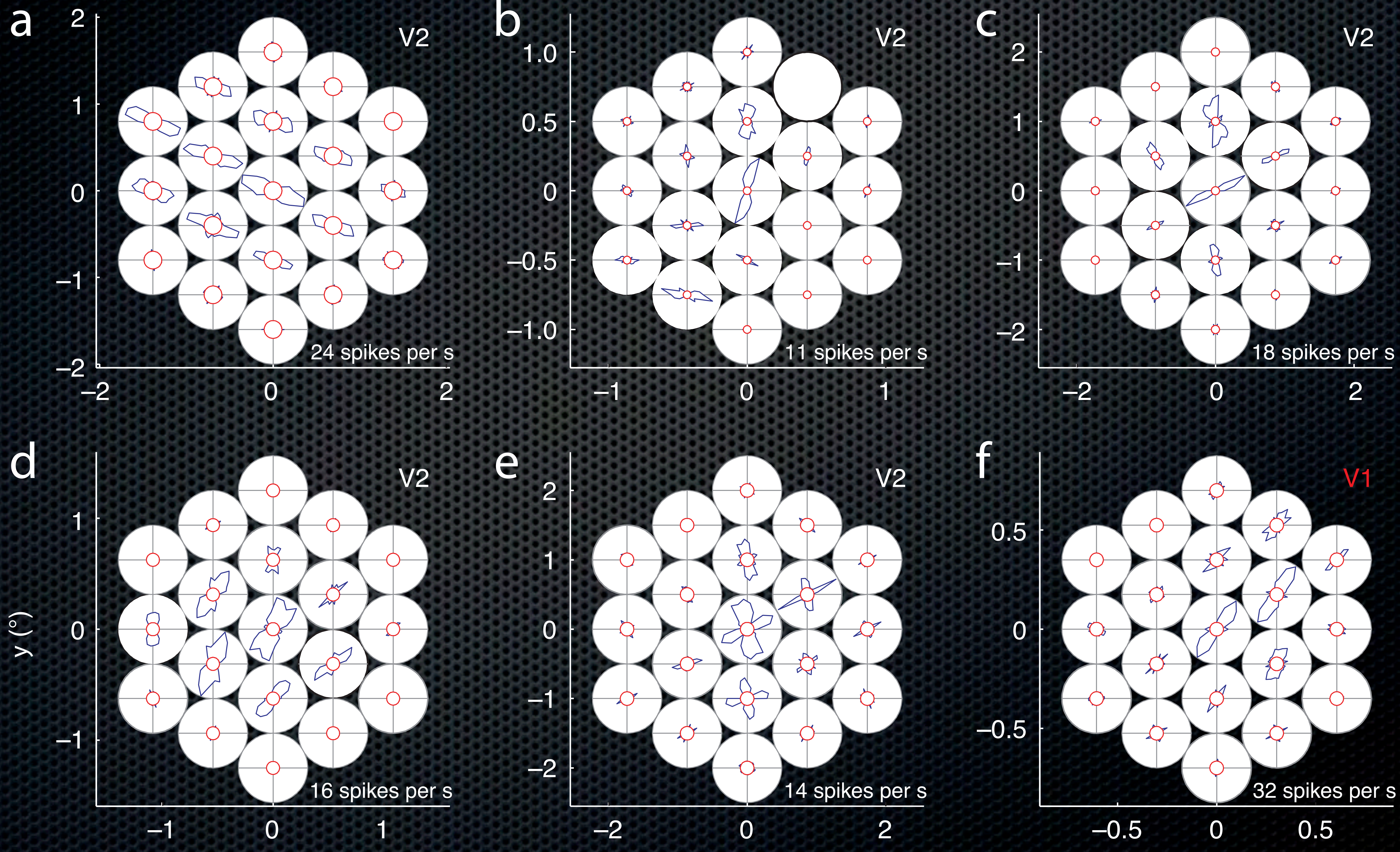
cross-orientation fields



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

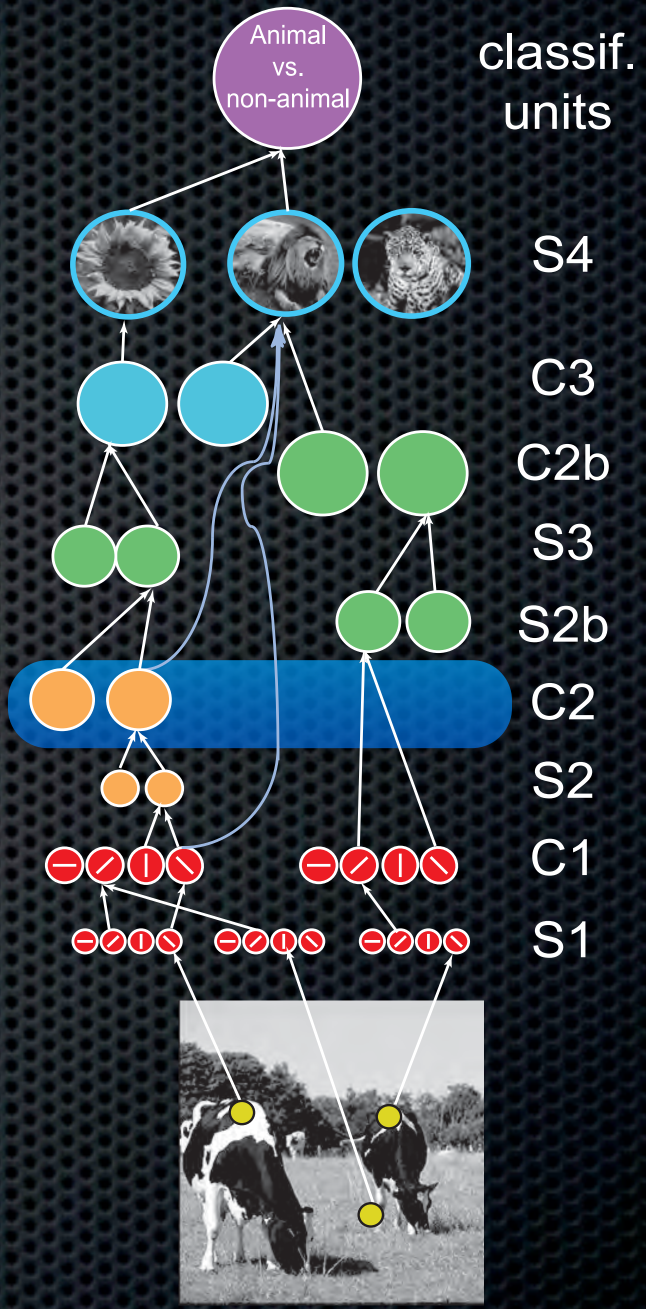
Neurons in monkey visual area V2 encode combinations of orientations

Akiyuki Anzai, Xinmiao Peng & David C Van Essen



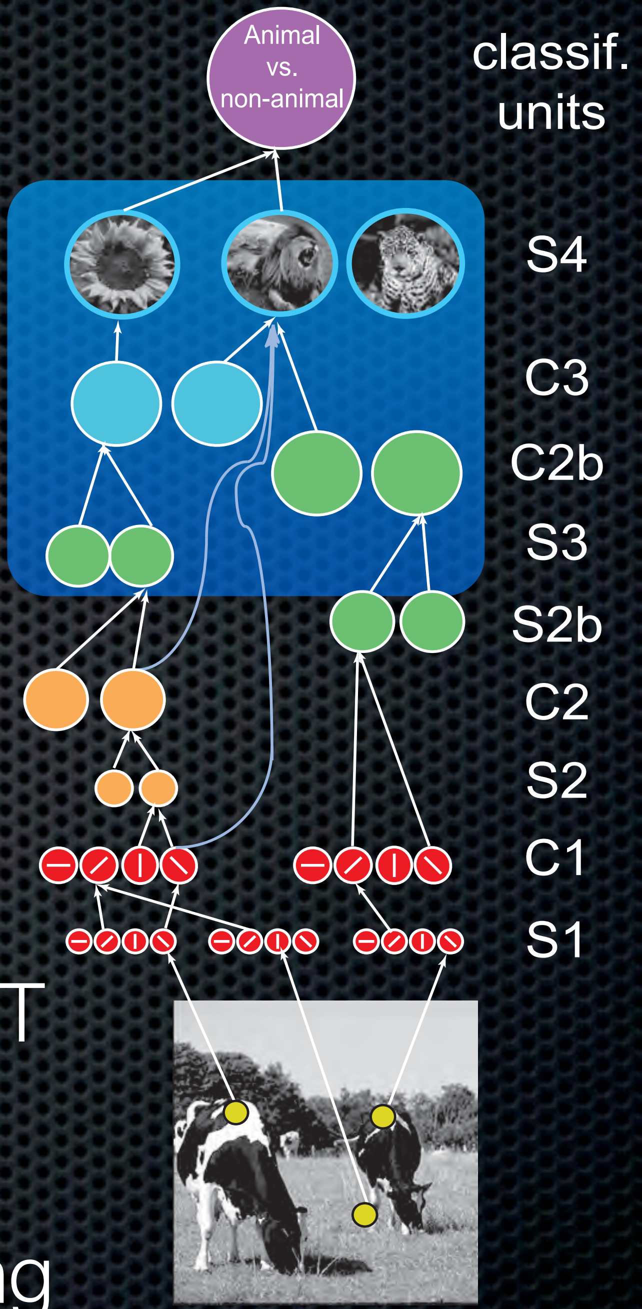
C2 units

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



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)
- S4 units:
  - View-tuned units (imprinted with part of the training set, e.g. animal and non-animal images but still unsupervised)
  - Tuning and invariance properties agrees with IT data (Logothetis Pauls & Poggio 1995)



So why hierarchies?

- Idea 1: Built-in invariance to 2D transformations (rotation and scale)
- Idea 2: Generic features shared between multiple categories
- Overall reduce “sample complexity” and reduces number of training examples needed to learn a task

PFC

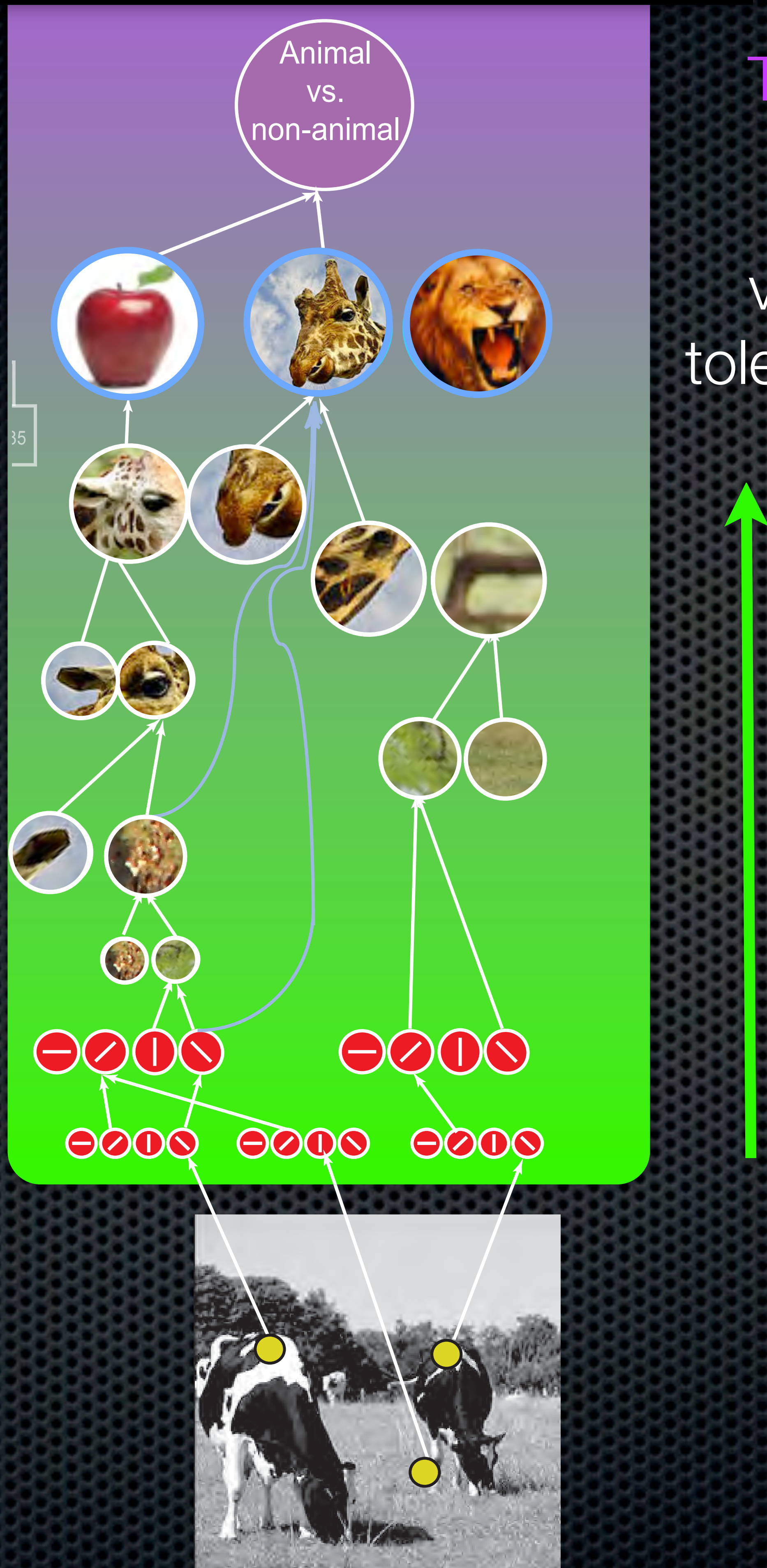
AIT

PIT

V4

V2

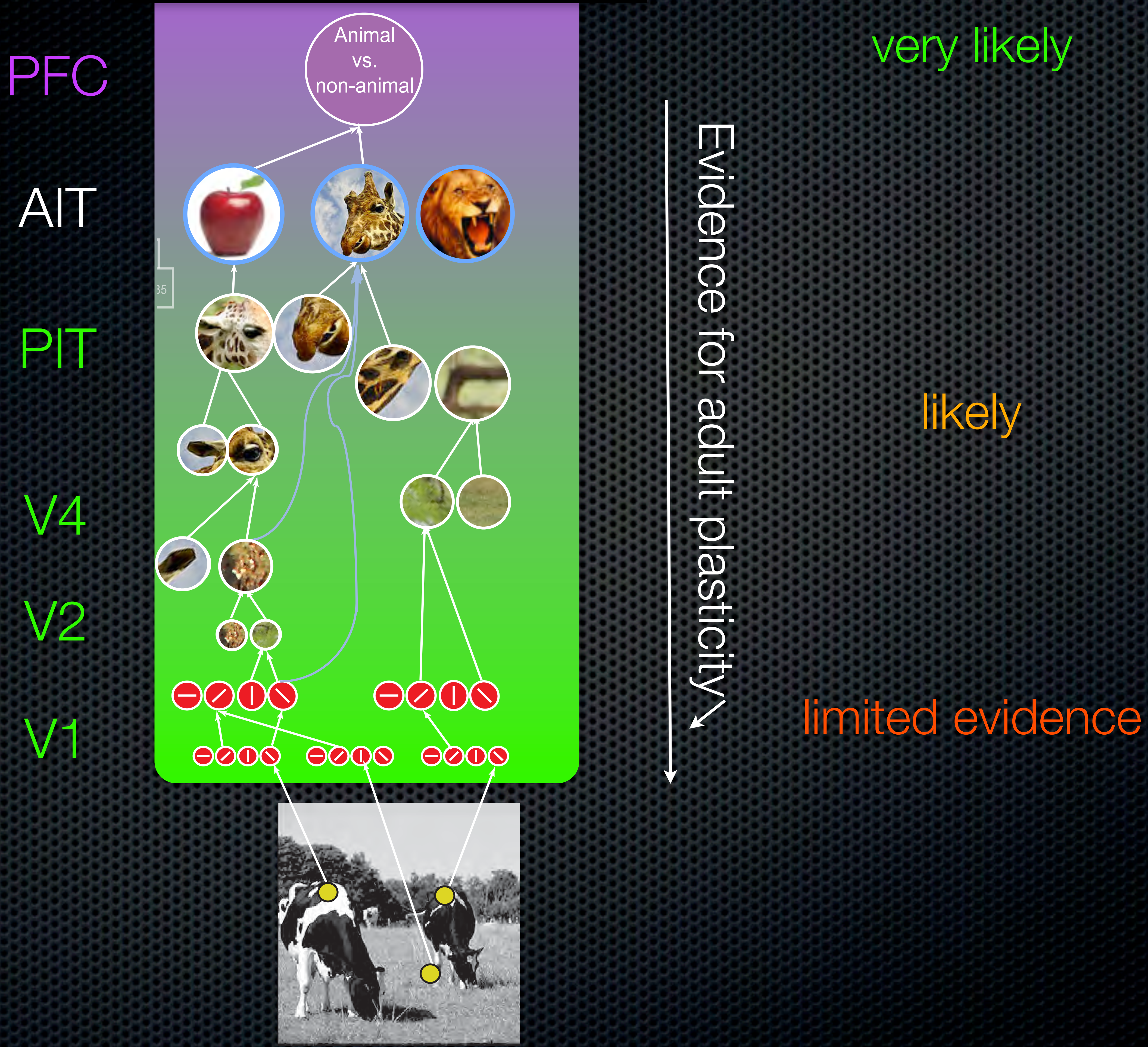
V1



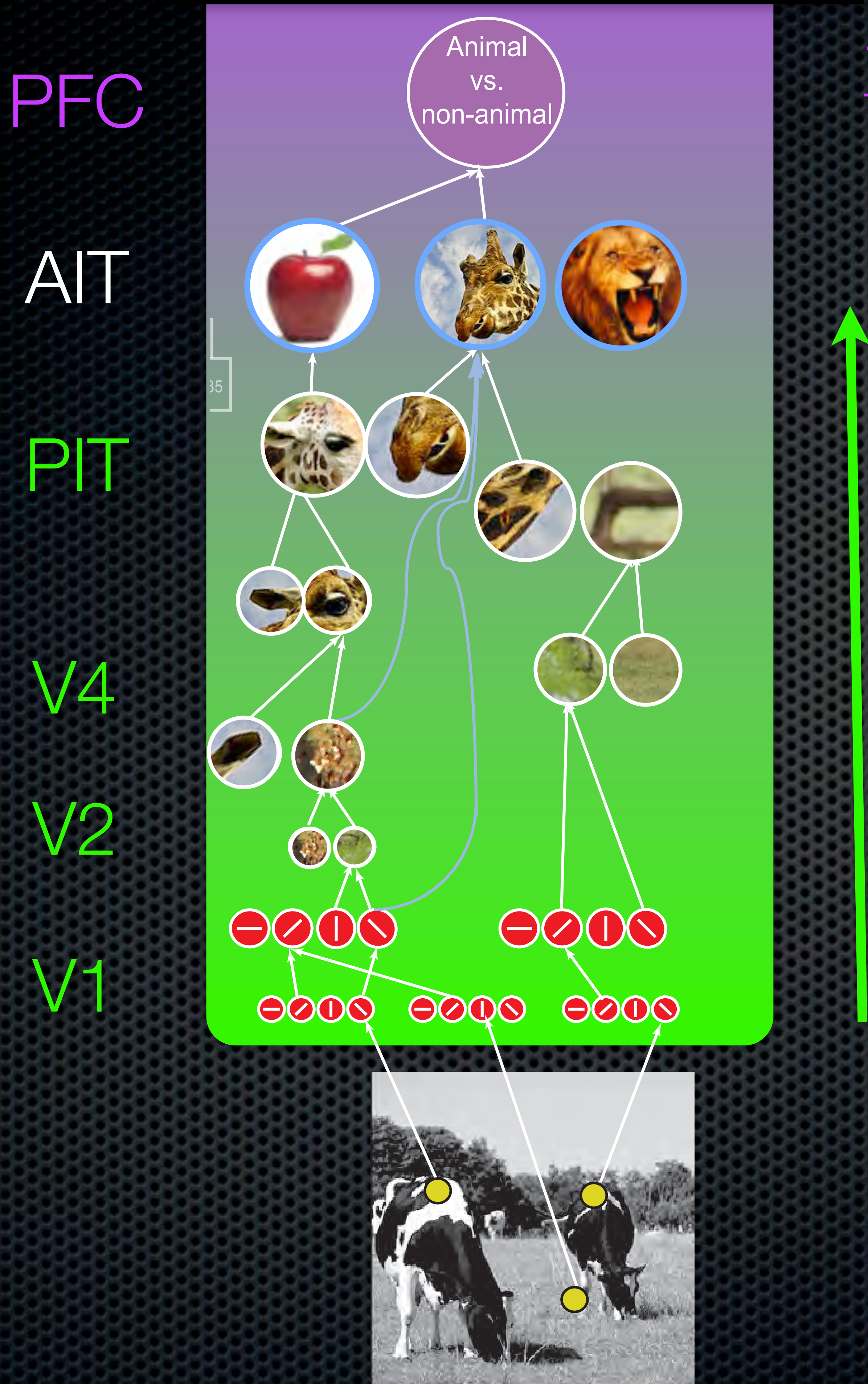
Task-specific = categorization circuits

view-based object representation but tolerant position, scale and small rotations

features of increasing complexity and tolerance to position and scale

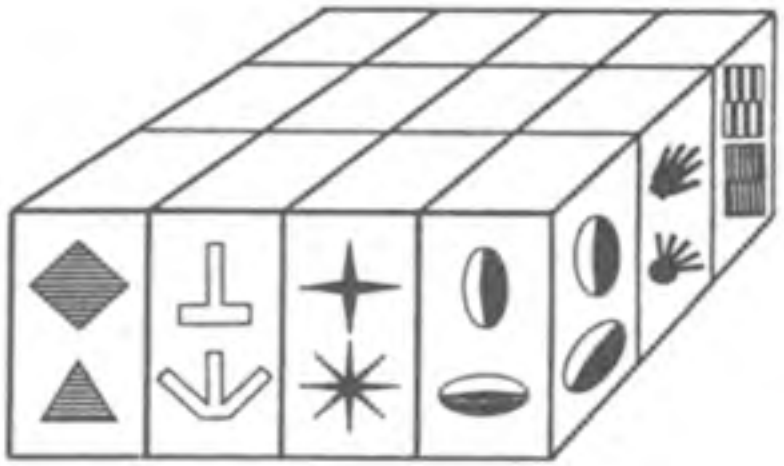

PFC
AIT
PIT
V4
V2
V1
Animal vs. non-animal
Evidence for adult plasticity
very likely
likely
limited evidence

supervised learning from a handful of training examples ~ linear perceptron

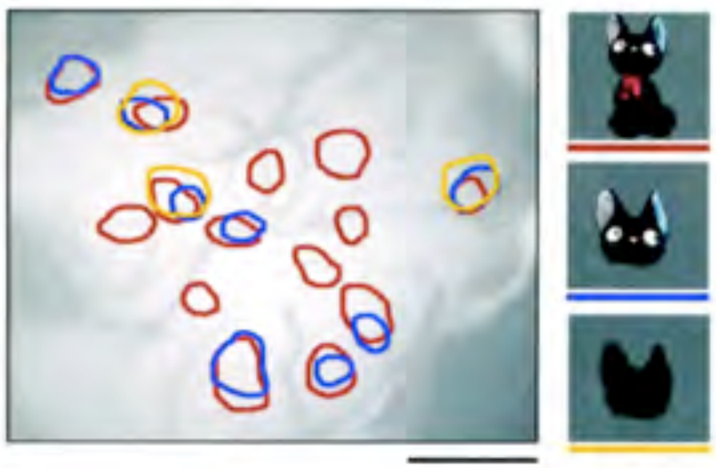


unsupervised developmental-like learning stage

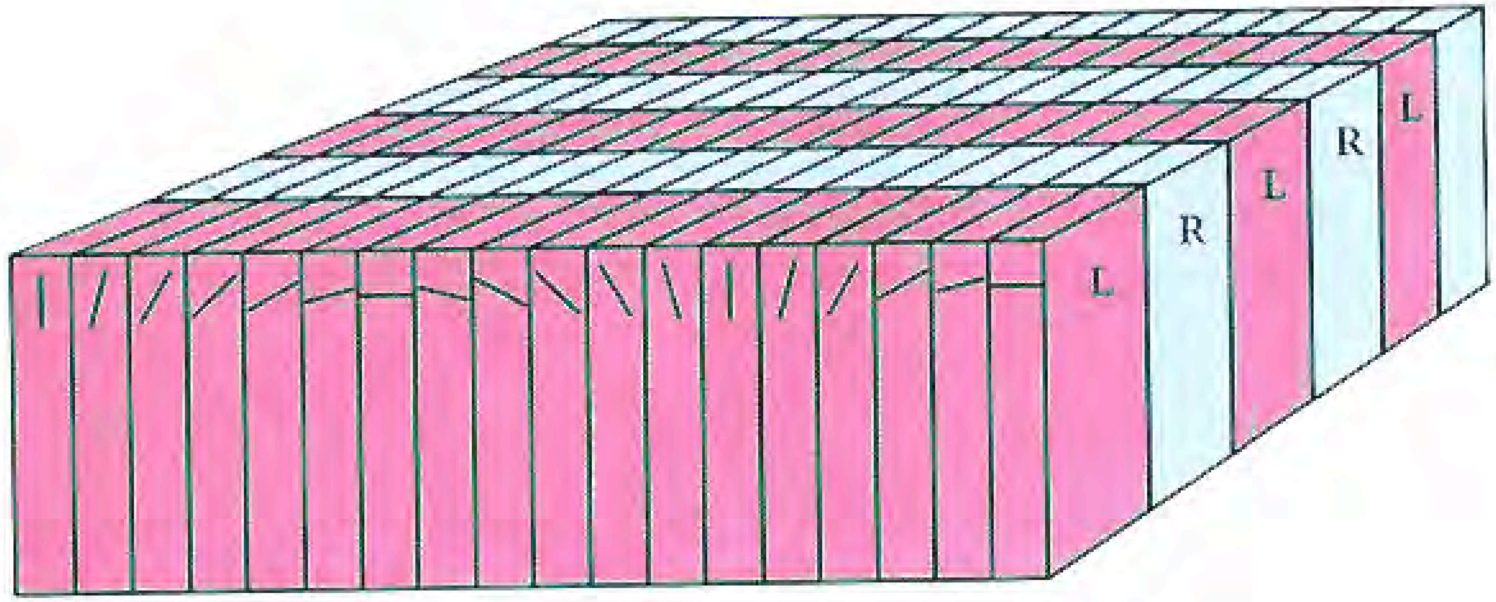
Columns in the cortex



Tanaka et al.



Tsunoda et al.

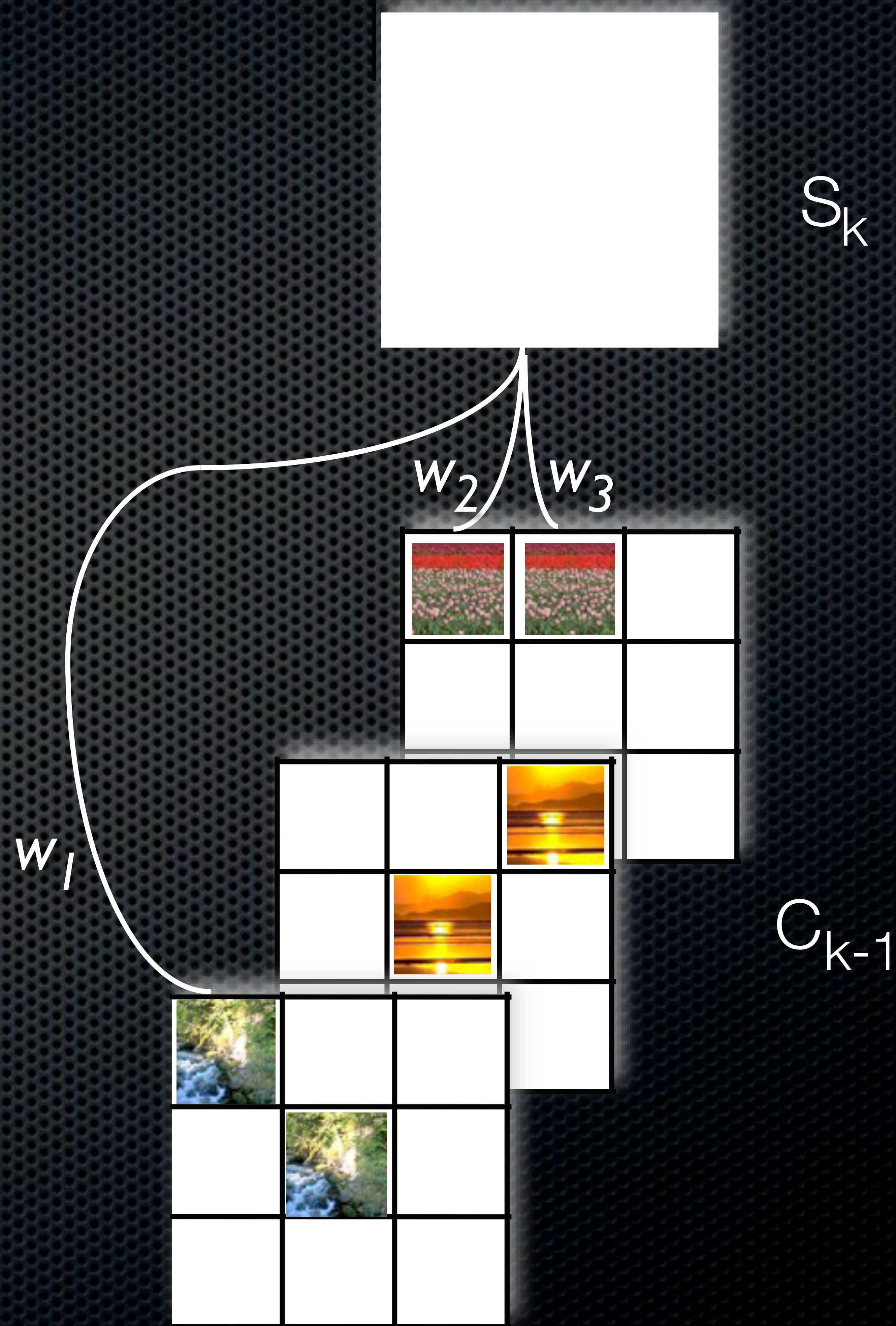


Orientation and ocular dominance columns

Figure 23. The ice-cube model of the cortex. It illustrates how the cortex is divided, at the same time, into two kinds of slabs, one set of ocular dominance (left and right) and one set for orientation. The model should not be taken literally: Neither set is as regular as this, and the orientation slabs especially are far from parallel or straight.

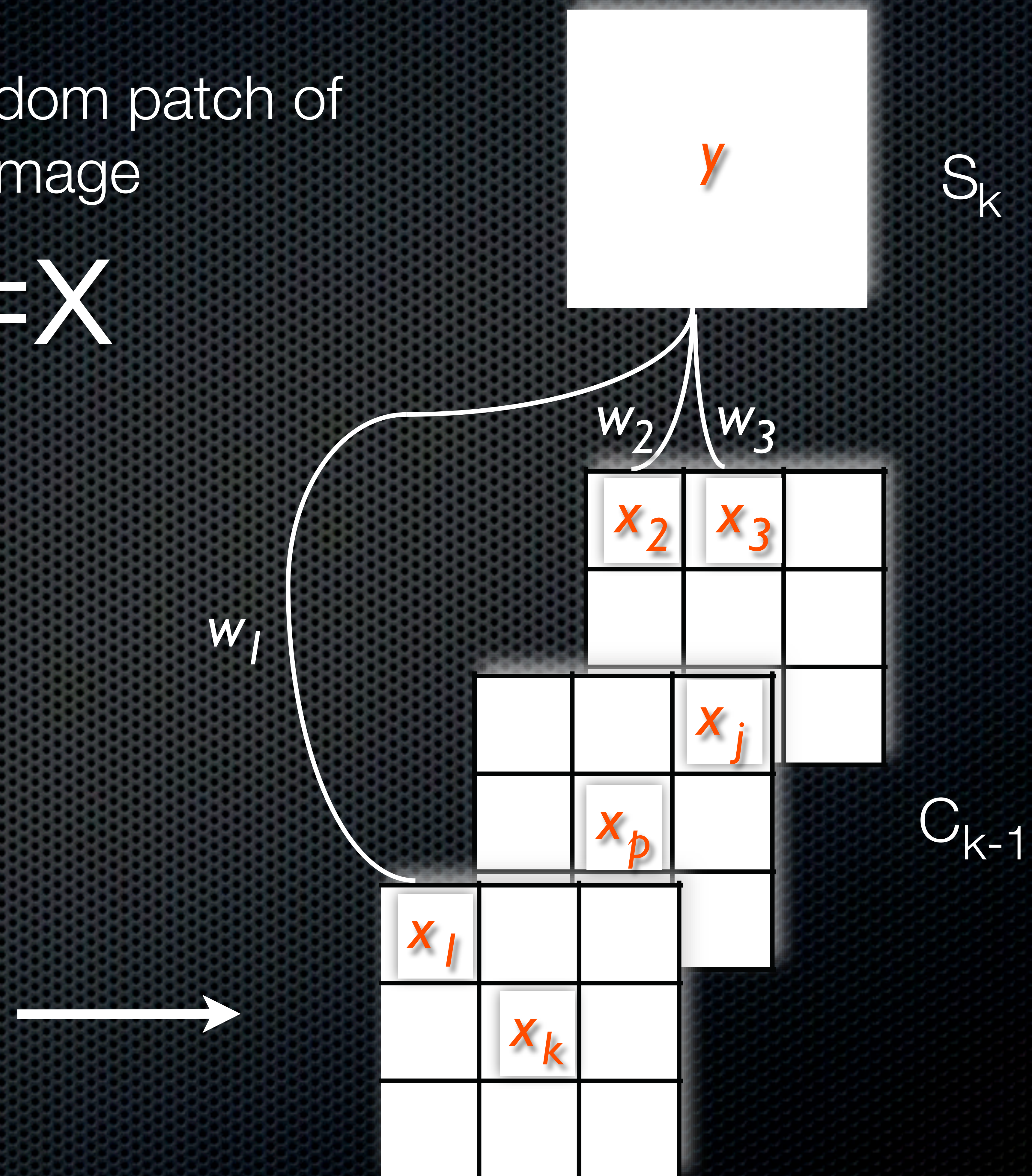
- Layers of the model are organized in columns
- Each model unit is equivalent to ~100 IF (~1 column of cortex)
- Each hypercolumn contains the same basic dictionary of features and is replicated at all positions and scales

- Learning is sequential
- Start with layer S2/C2 then S2b/C2b and S3/C3
- Pick one unit in layer Sk
- Select random set of inputs from retinotopically organized afferents

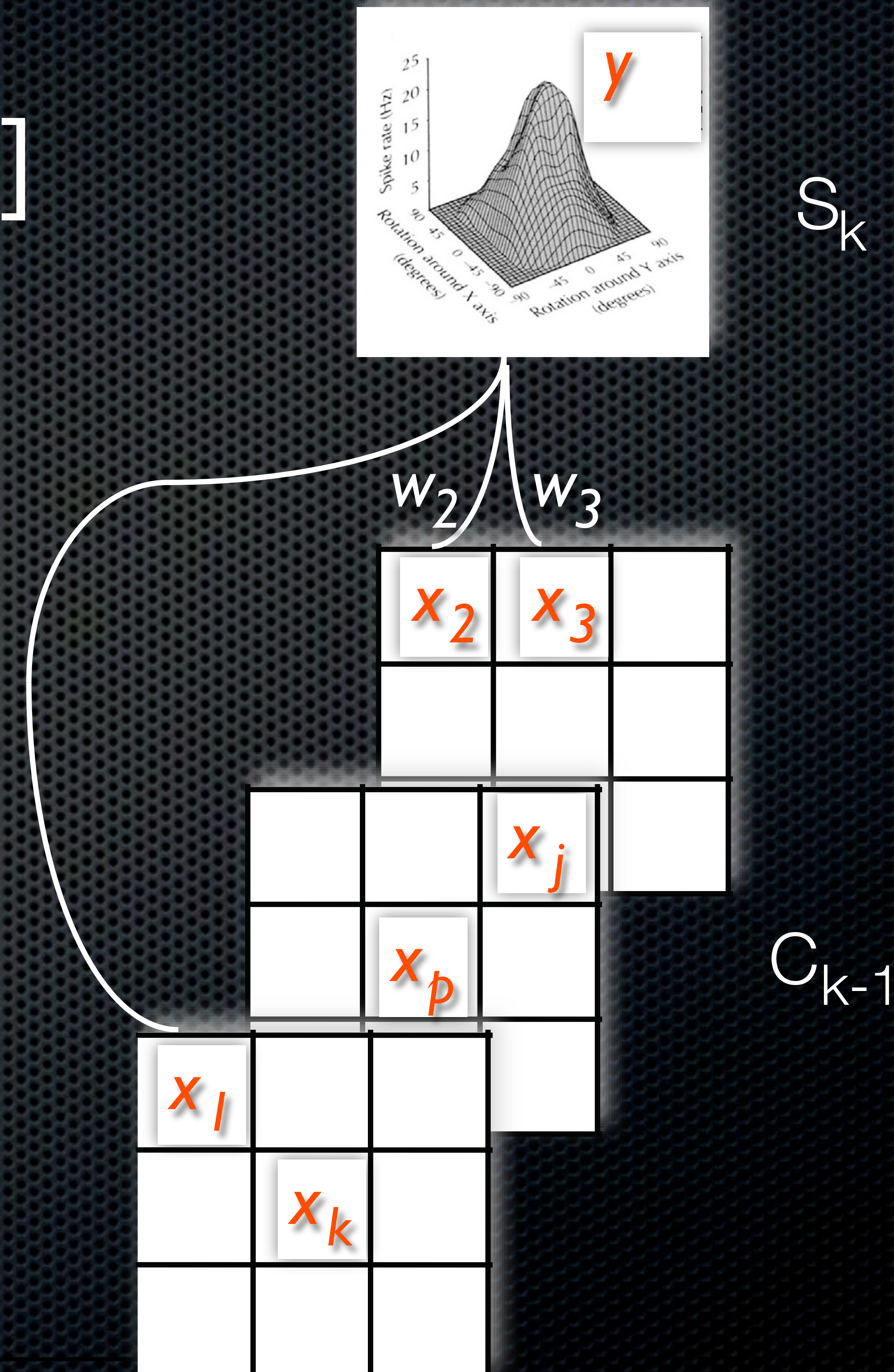


Imprint with random patch of natural image

W=X



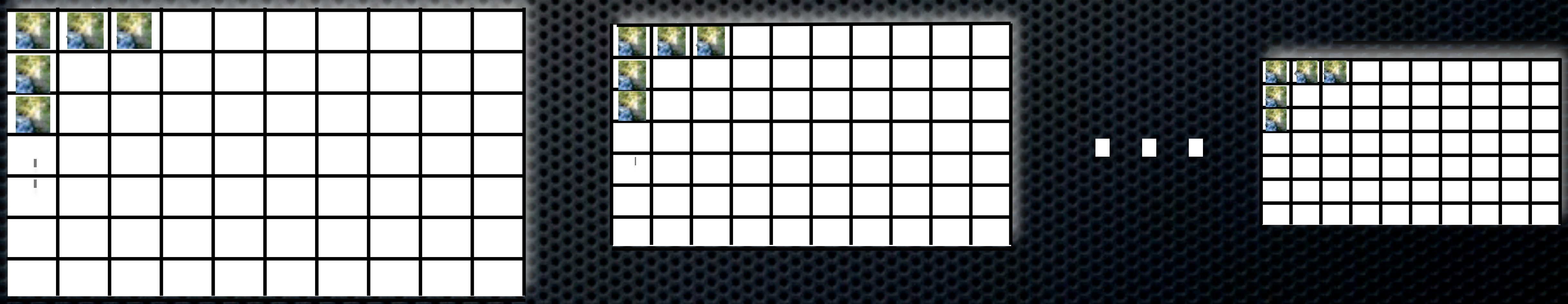
$$y = \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(w_j - x_j)^2\right]$$



✦ We learn ~1,000 units this way and then move to the next layer

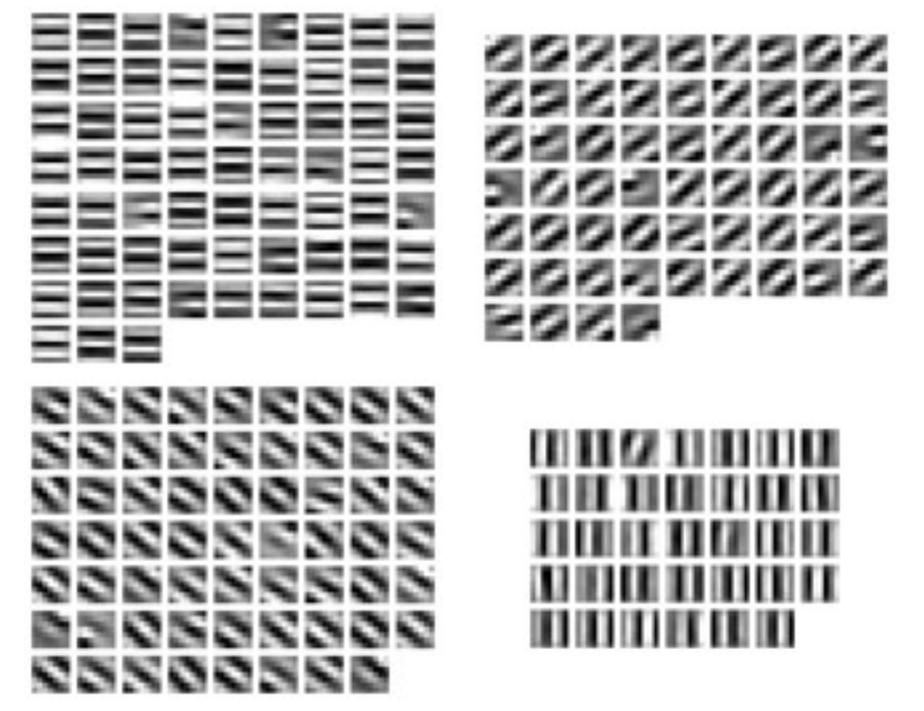
✦ Learning follows a long tradition of researchers who have argued that the visual system may be adapted to the statistics of the natural environment (Attneave 1954; Barlow 1961; Atick 1992; Ruderman 1994; Simoncelli & Olshausen 2001)

✦Here we assume the input image moves (shifting and looming) so that the selectivity of the imprinted units gets replicated at all positions and scales



Learning invariances

w| T. Masquelier & S. Thorpe
(CNRS, France)



- ✦ Simple cells learn correlation in space (at the same time)
- ✦ Complex cells learn correlation in time

movie courtesy of Wolfgang Einhauser

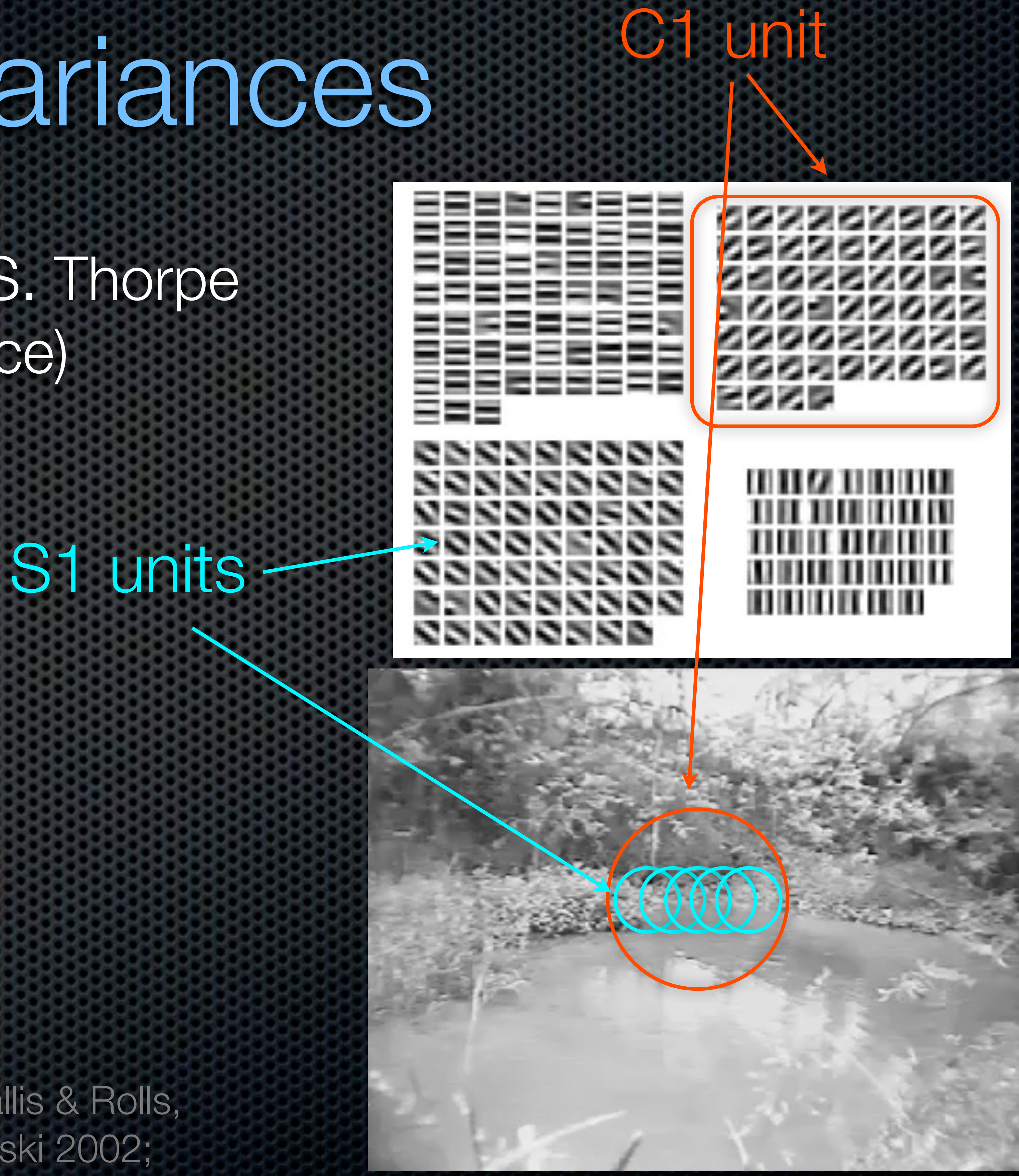
see also (Foldiak 1991; Perrett et al 1984; Wallis & Rolls, 1997; Einhauser et al 2002; Wiskott & Sejnowski 2002; Spratling 2005)

Learning invariances

w| T. Masquelier & S. Thorpe
(CNRS, France)

✦ Simple cells learn correlation in space (at the same time)

✦ Complex cells learn correlation in time



movie courtesy of Wolfgang Einhauser

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

Learning a dictionary of shape-components in the visual cortex

- Learning frequent image features during development
- Object categories share reusable features
- Large redundant vocabulary for implicit geometry

Learning a dictionary of shape-components in the visual cortex

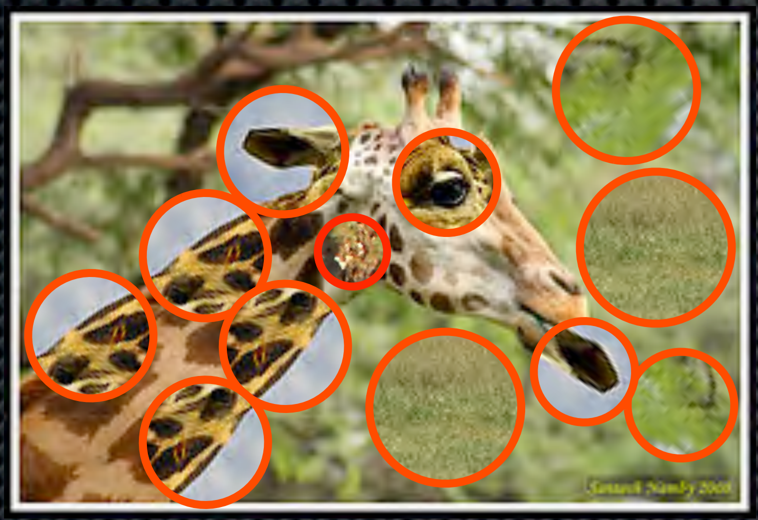
- Learning frequent image features during development
- Object categories share reusable features
- Large redundant vocabulary for implicit geometry

Learning a dictionary of shape-components in the visual cortex

IT

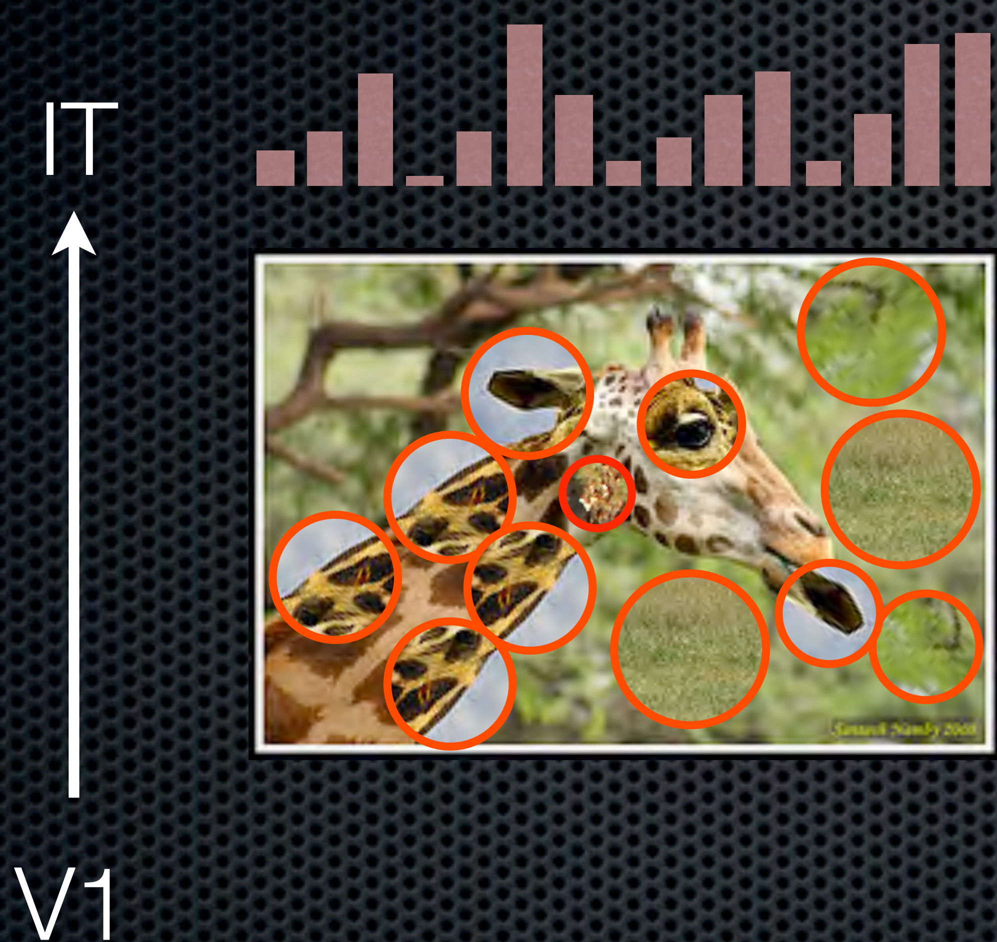
↑

V1



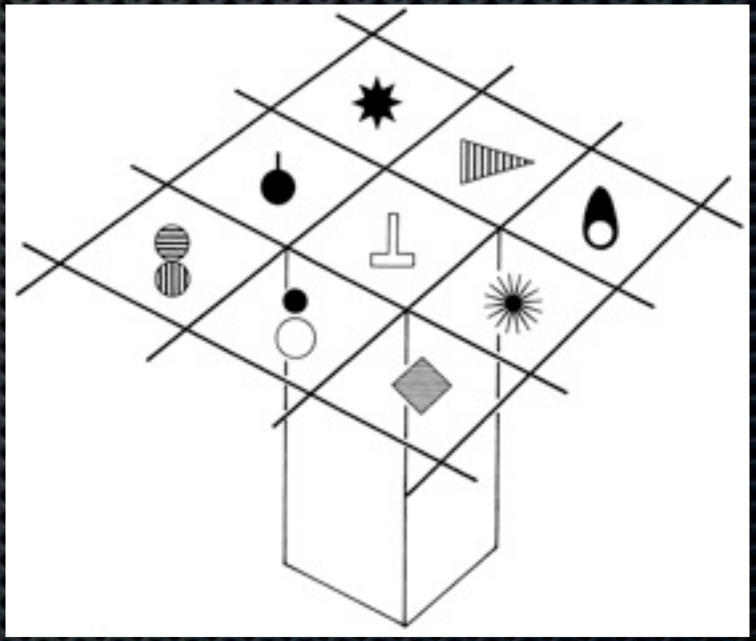
- Learning frequent image features during development
- Object categories share reusable features
- Large redundant vocabulary for implicit geometry

Learning a dictionary of shape-components in the visual cortex



- Learning frequent image features during development
- Object categories share reusable features
- Large redundant vocabulary for implicit geometry

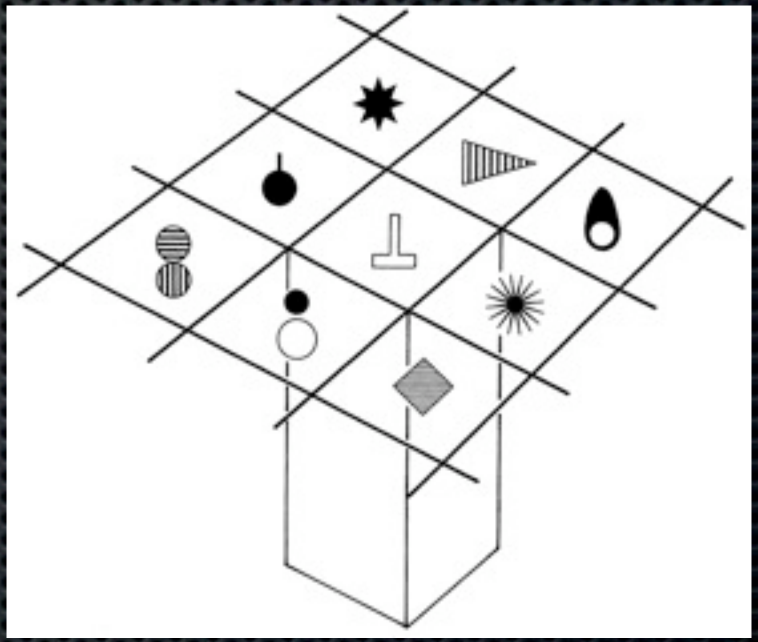
Learning a dictionary of shape-components in visual cortex



“critical” feature columns in IT
(Tanaka, 1996)

Learning a dictionary of shape-components in visual cortex

- Pre-attentive processing:
  - “Loose collection of basic features” (Wolfe & Bennett 1997)
  - “Unbound features” (Treisman et al)

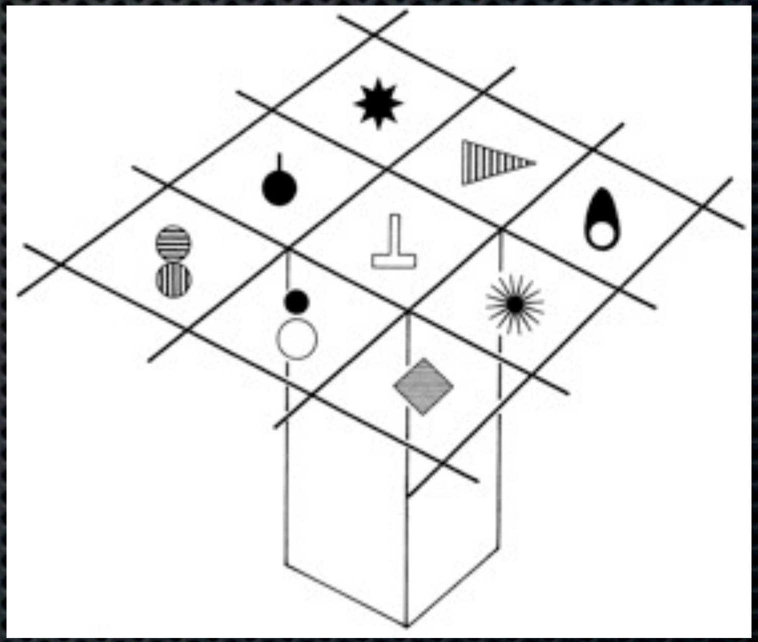


“critical” feature columns in IT
(Tanaka, 1996)

Learning a dictionary of shape-components in visual cortex

✦ Pre-attentive processing:

- "Loose collection of basic features" (Wolfe & Bennett 1997)
- "Unbound features" (Treisman et al)

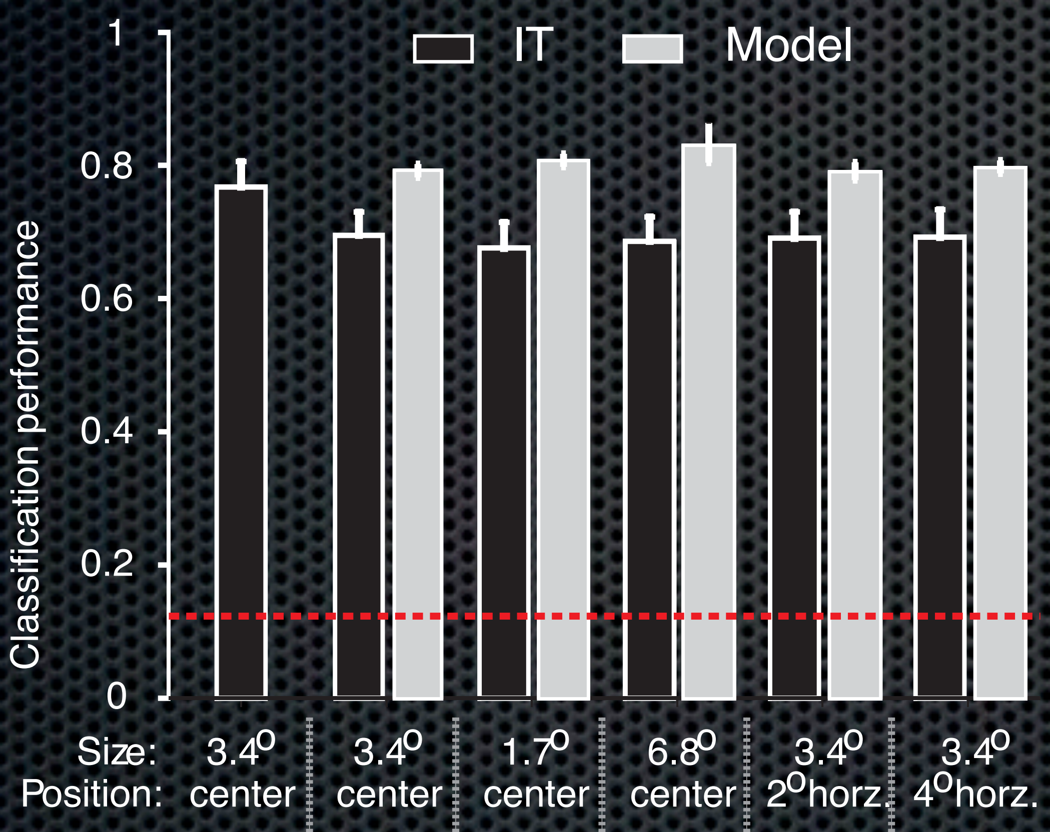


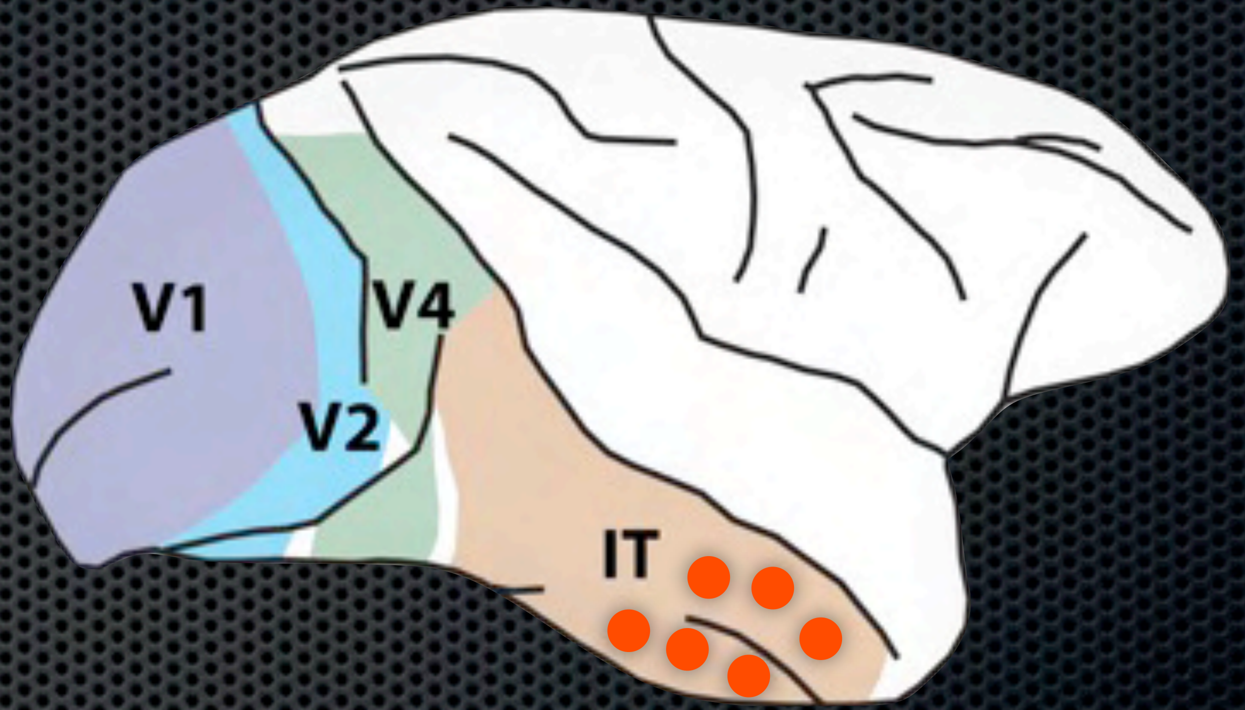
"critical" feature columns in IT
(Tanaka, 1996)

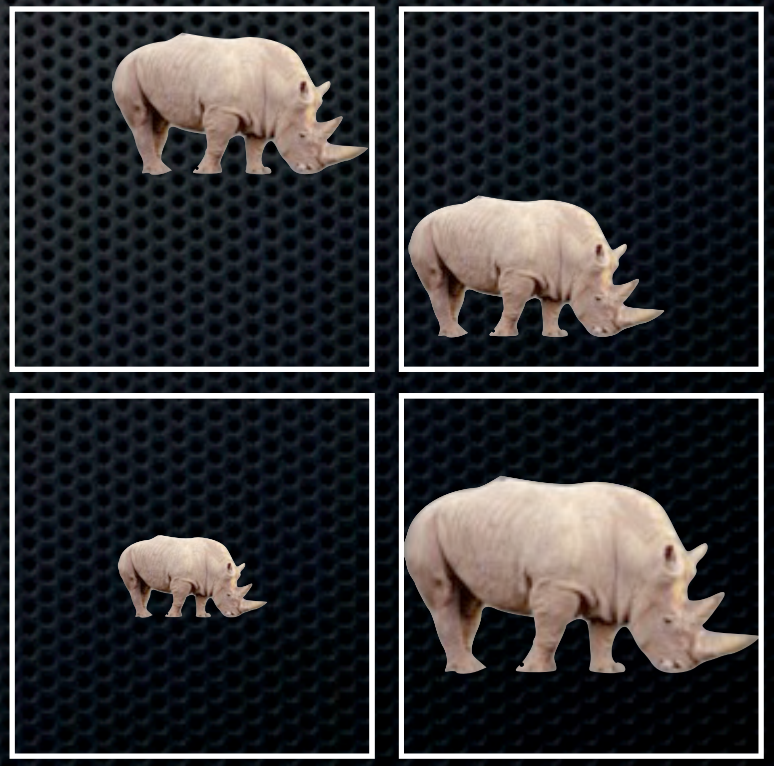
✦ Computer vision:

- Component-based > holistic representation (Perona et al 1995, 1996, 2000; Heisele Serre & Poggio 2001, 2002)
- Features of intermediate complexity are optimal (Ullman, 2002)
- Bag of features (Csurka et al 2004; Sivic et al 2005; Sudderth et al 2005)

C2 vs. IT neurons





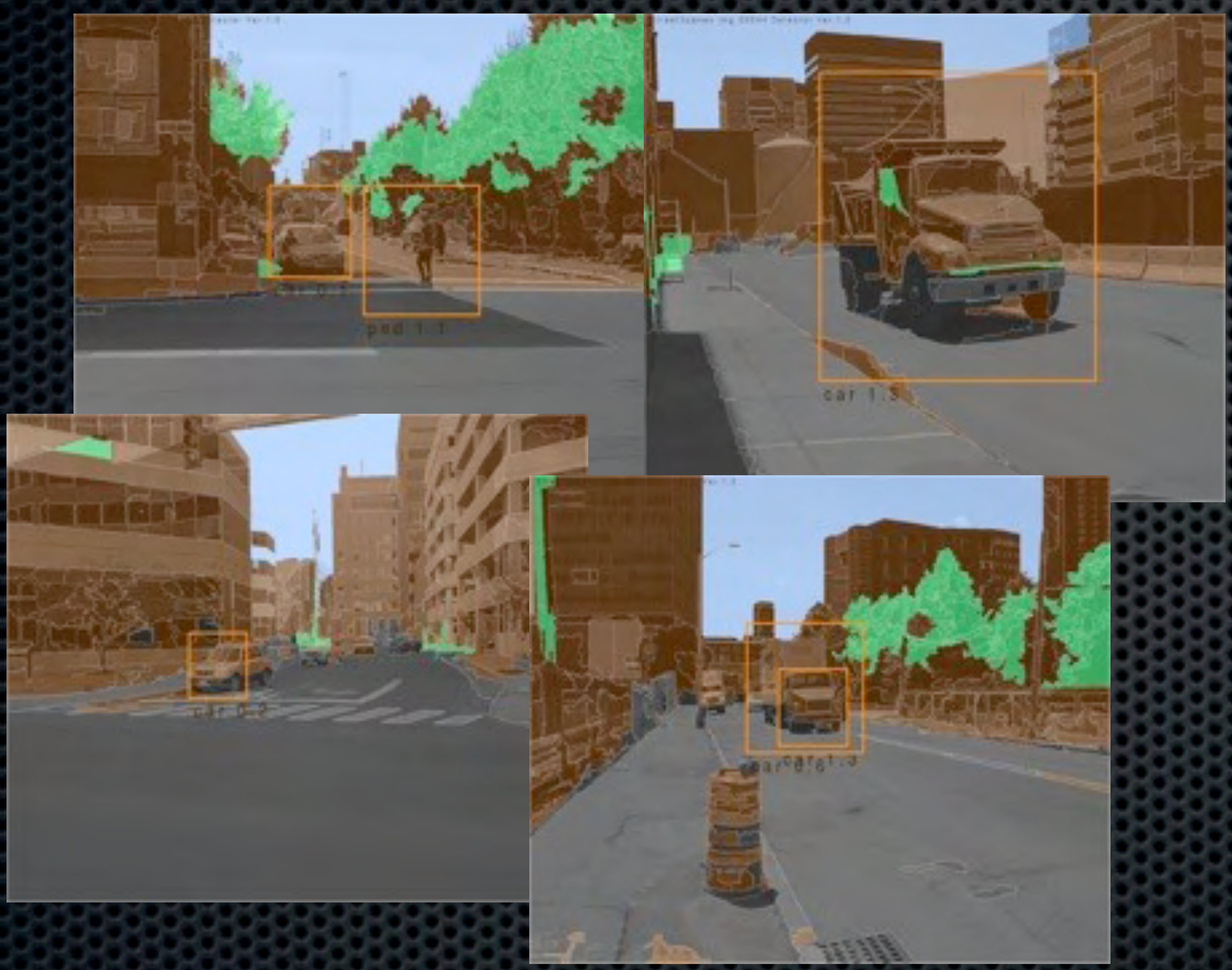


Model data: Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005
Experimental data: Hung* Kreiman* Poggio & DiCarlo 2005

Application to computer vision

Bio-motivated computer vision

Scene parsing and object recognition



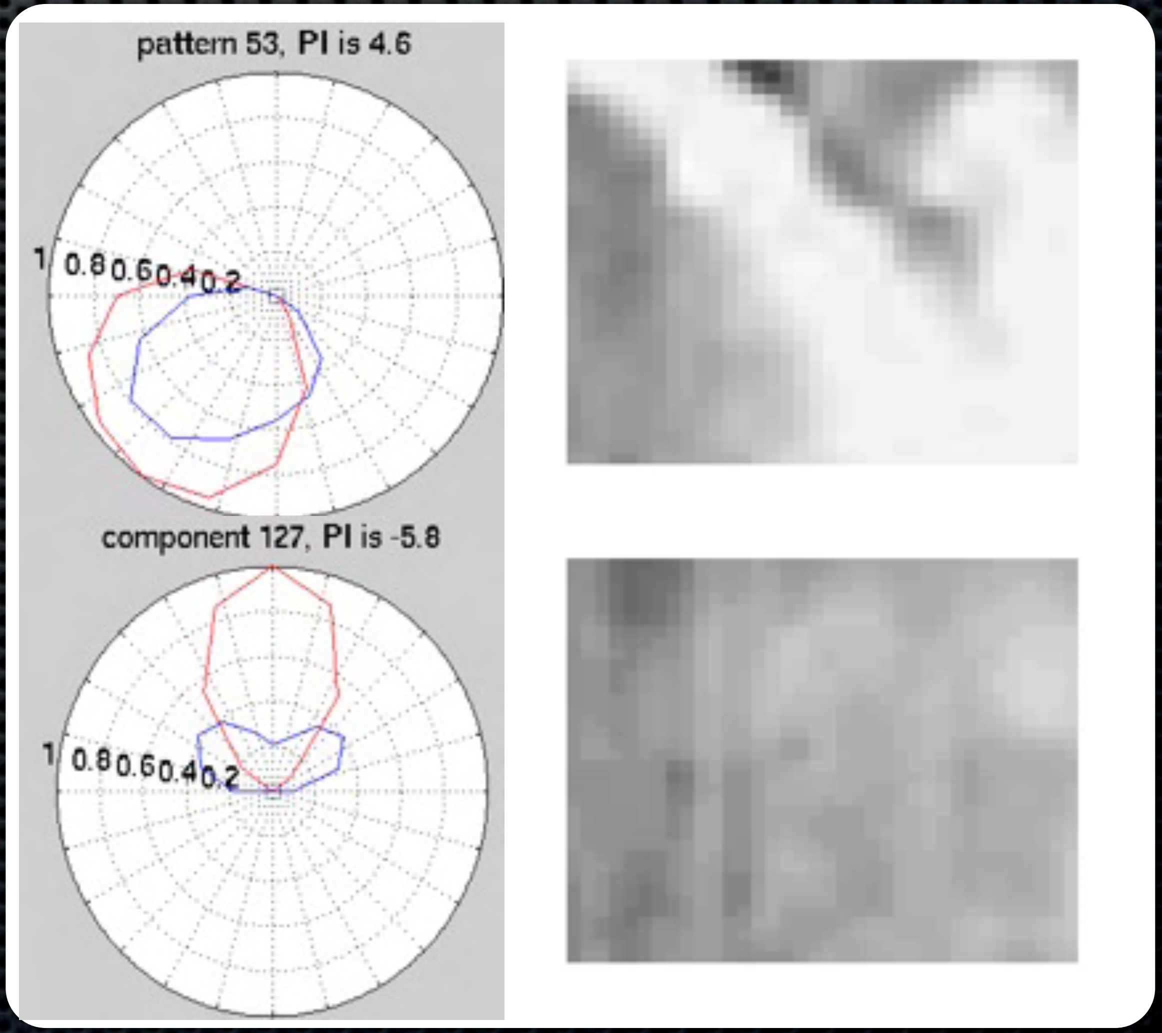
Serre Wolf & Poggio 2005; Wolf & Bileschi 2006;
Serre et al 2007

Computer vision system based on the response properties of neurons in the ventral stream of the visual cortex

Bio-motivated computer vision

Action recognition in video sequences

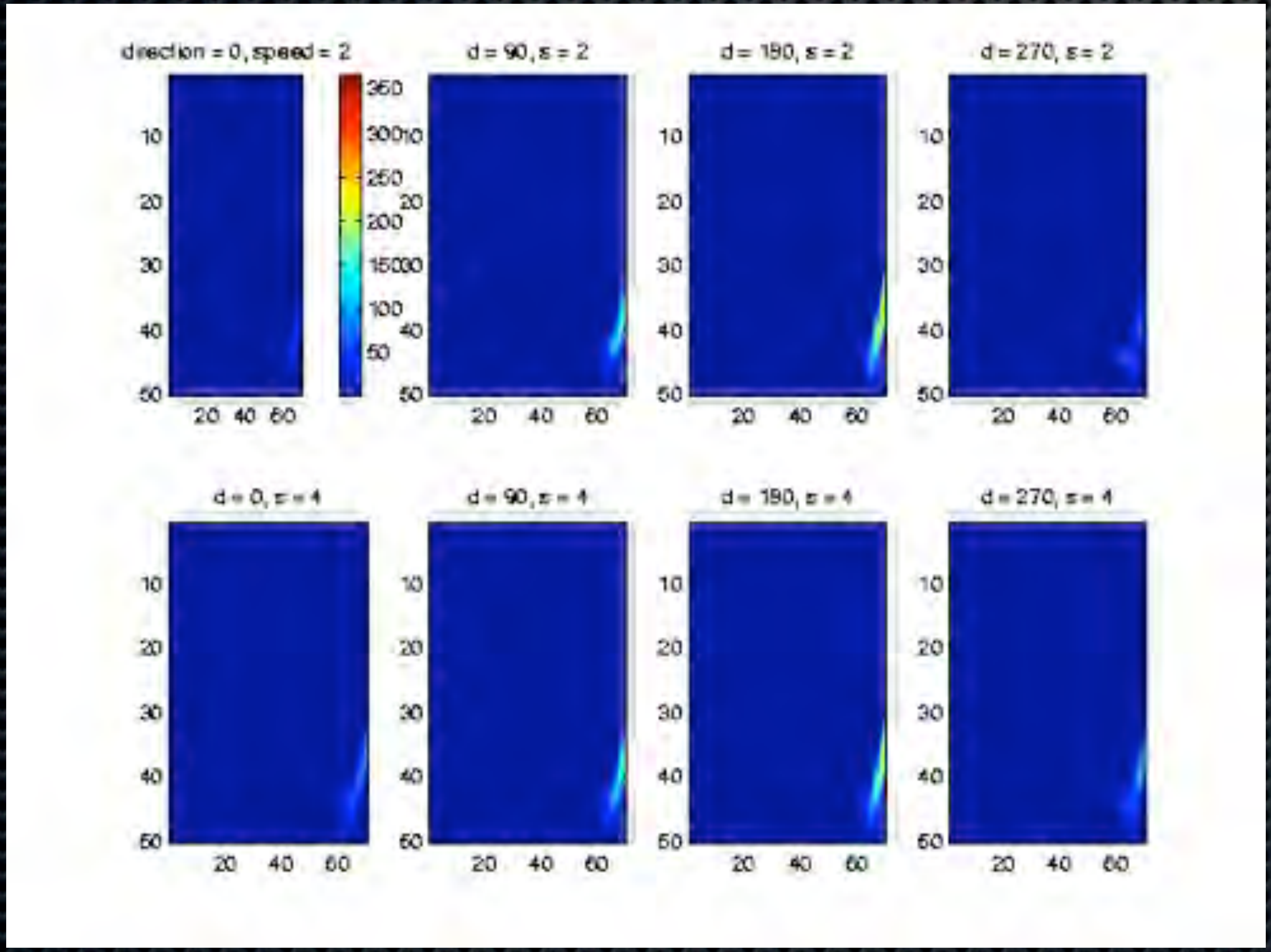
motion-sensitive MT-like units



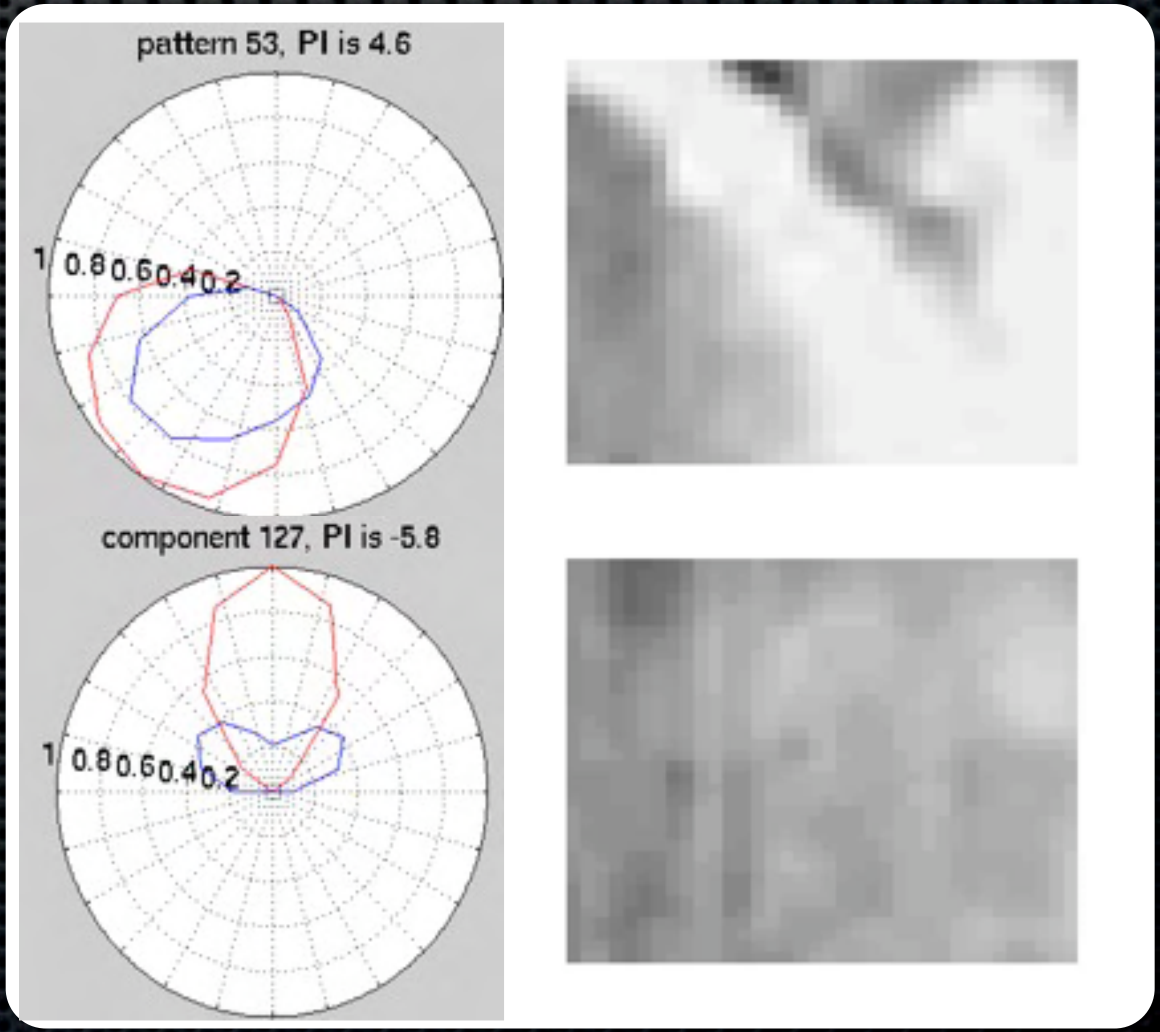
Jhuang Serre Wolf & Poggio 2007

Bio-motivated computer vision

Action recognition in video sequences



motion-sensitive MT-like units



Jhuang Serre Wolf & Poggio 2007

Recognition accuracy

	Dollar et al '05	model	chance
KTH Human	81.3%	**91.6%**	16.7%
Weiz. Human	86.7%	**96.3%**	11.1%
UCSD Mice	75.6%	**79.0%**	20.0%

★ Cross-validation: 2/3 training, 1/3 testing, 10 repeats

Jhuang Serre Wolf & Poggio ICCV'07

Automatic recognition of rodent behavior

Serre Jhuang Garrote Poggio Steele in prep

Automatic recognition of rodent behavior

Performance

human agreement	72%
proposed system	71%
commercial system	56%
chance	12%

Serre Jhuang Garrote Poggio Steele in prep

This lecture

1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world
2. Rapid recognition and feedforward processing:
   - Predicting human performance
   - “Clutter problem”
3. Beyond feedforward processing:
   - Top-down cortical feedback and attention to solve the “clutter problem”
   - Predicting human eye movements

This lecture

1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world
2. Rapid recognition and feedforward processing:
   - Predicting human performance
   - “Clutter problem”
3. Beyond feedforward processing:
   - Top-down cortical feedback and attention to solve the “clutter problem”
   - Predicting human eye movements

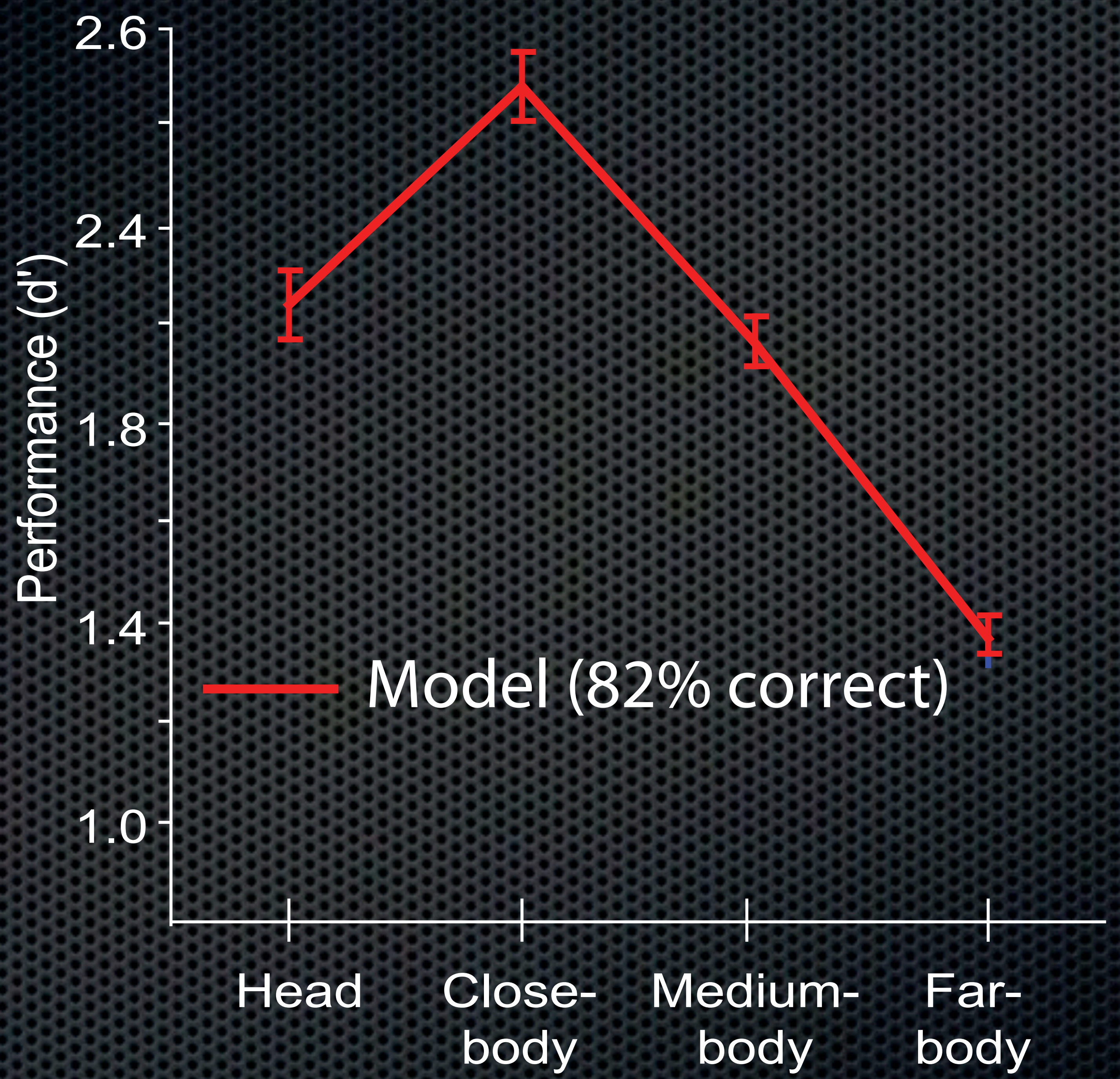
Head | Close-body | Medium-body | Far-body

Animals

Natural distractors

Artificial distractors

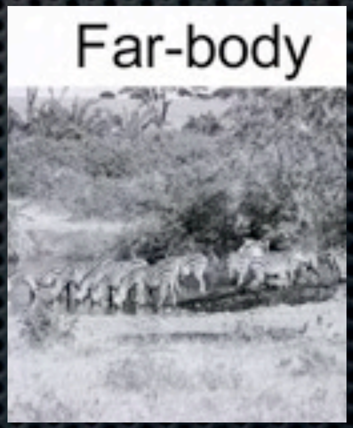
Database collected by Torralba & Oliva (2003)

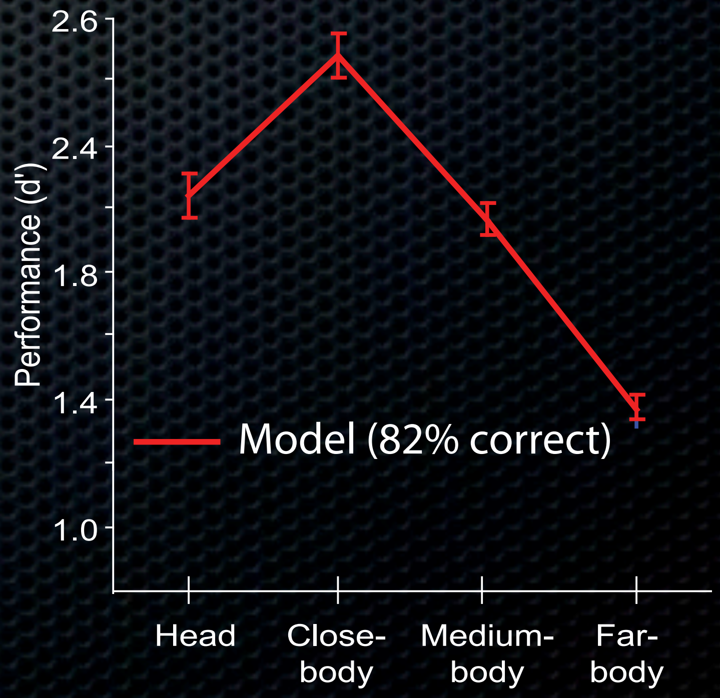


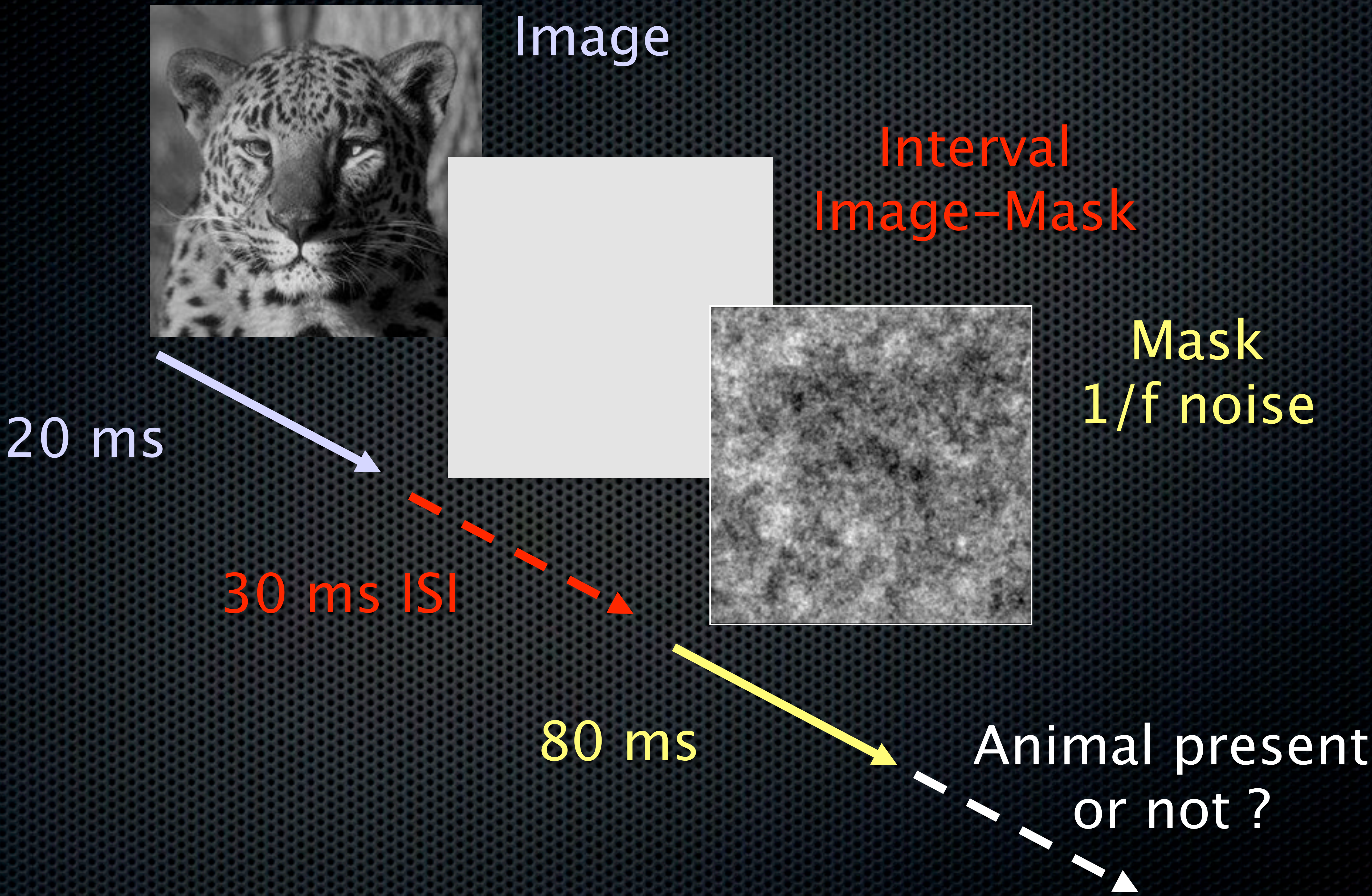
Serre Oliva Poggio 2007

“Clutter effect”

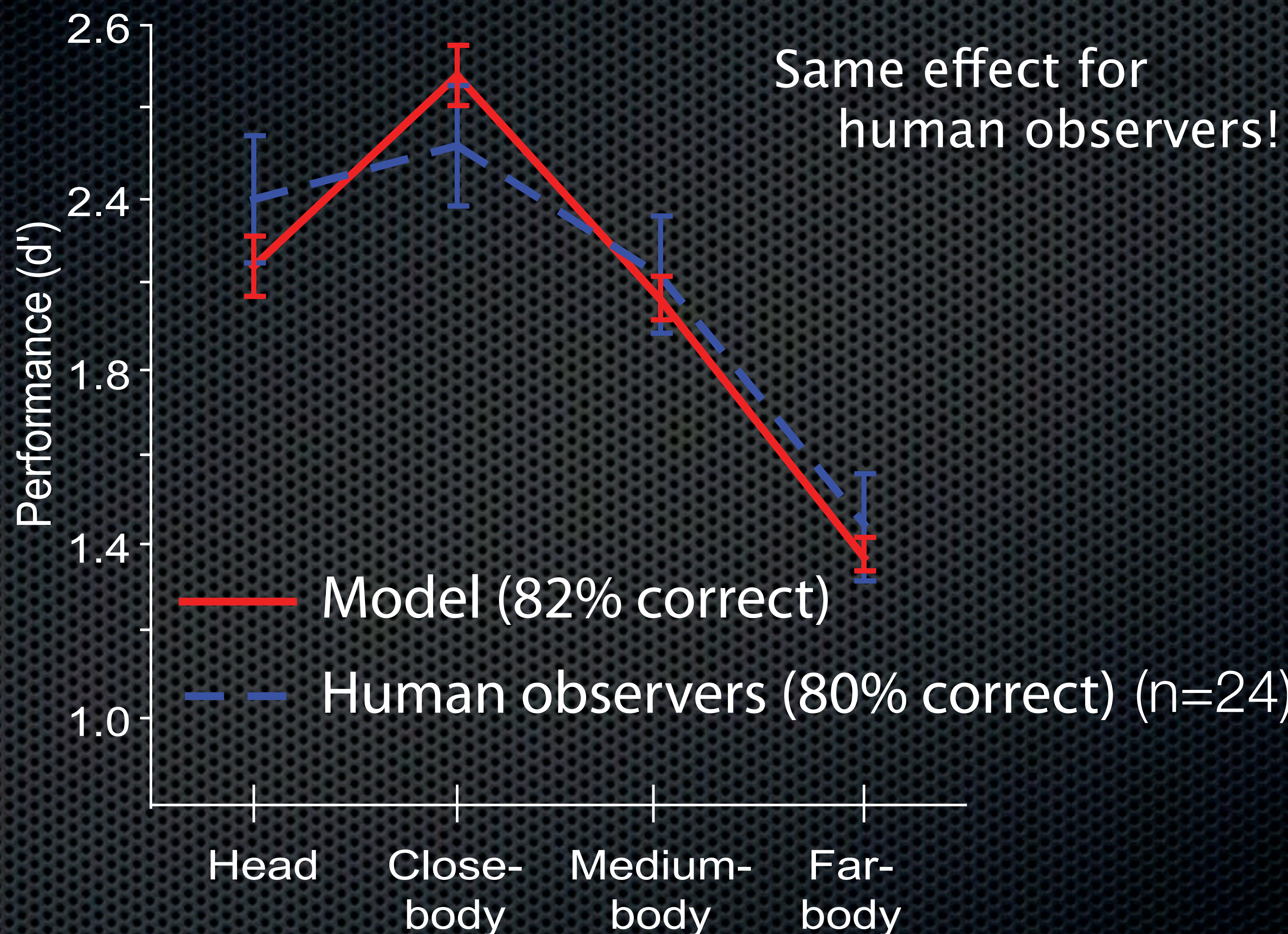
- High performance (~90%) when
  - maximal amount of information present
  - in the absence of clutter
- Performance decreases (~74%) with increasing amount of clutter
- Limitation of feedforward model compatible with decrease in response in V4 (Reynolds Chelazzi & Desimone 1999) and IT in the presence of clutter (Zoccolan, Cox, DiCarlo, 2005; Zoccolan, Kouh, Poggio, DiCarlo, in sub; Rolls, Aggelopoulos, Zheng, 2003)







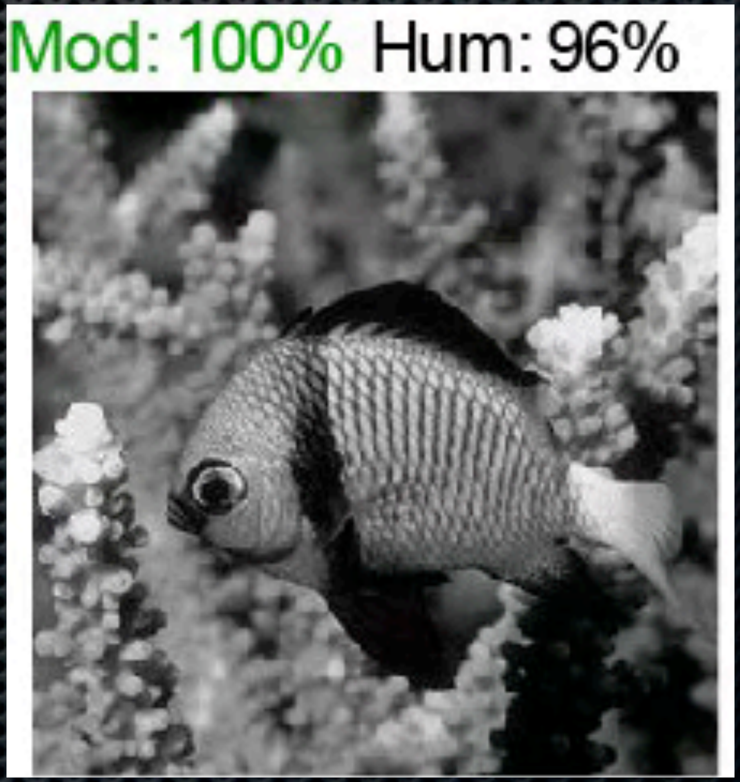
(Thorpe et al 1996; Van Rullen & Koch 2003; Bacon-Mace et al 2005)



Serre Oliva Poggio 2007

Further comparisons

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



- Model predicts level of performance on rotated images (90 deg and inversion)

Serre Oliva Poggio 2007

Show matlab demo

This lecture

1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world
2. Rapid recognition and feedforward processing:
   - Predicting human performance
   - “Clutter problem”
3. Beyond feedforward processing:
   - Top-down cortical feedback and attention to solve the “clutter problem”
   - Predicting human eye movements

This lecture

1. Learning a loose hierarchy of image fragments
   - The algorithm
   - Recognition in the real-world
2. Rapid recognition and feedforward processing:
   - Predicting human performance
   - “Clutter problem”
3. Beyond feedforward processing:
   - Top-down cortical feedback and attention to solve the “clutter problem”
   - Predicting human eye movements

Spatial attention solves the "clutter problem"

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

Problem: How to know where to attend?

Spatial attention solves the "clutter problem"

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

Problem: How to know where to attend?

Spatial attention solves the "clutter problem"

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

Problem: How to know where to attend?

Spatial attention solves the "clutter problem"

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

Problem: How to know where to attend?

Spatial attention solves the "clutter problem"

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

Science 22 April 2005:
Vol. 308. no. 5721, pp. 529 - 534

Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4

Narcisse P. Bichot, Andrew F. Rossi, Robert Desimone

Spatial attention solves the “clutter problem”

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

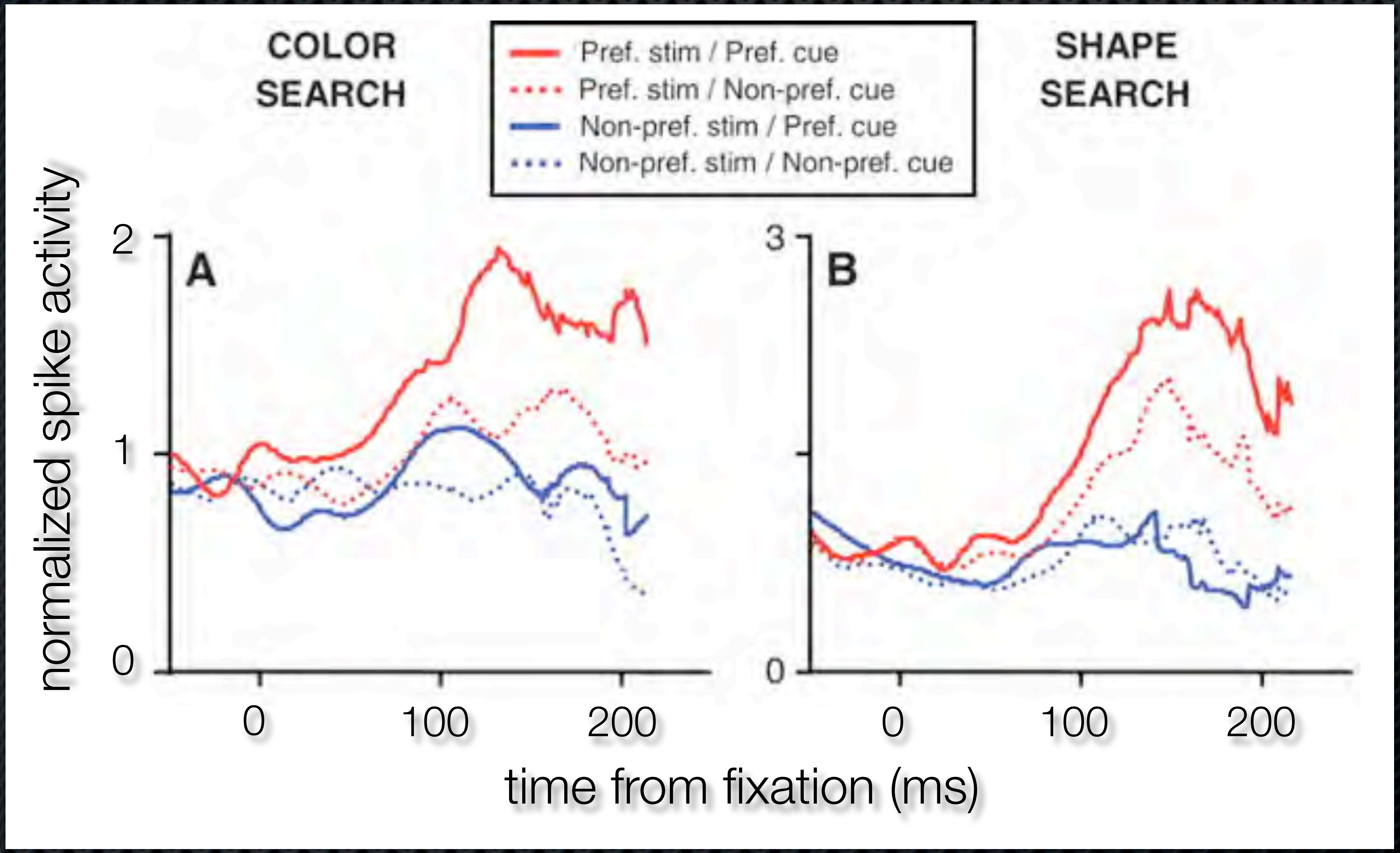
Science 22 April 2005:
Vol. 308. no. 5721, pp. 529 - 534

Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4

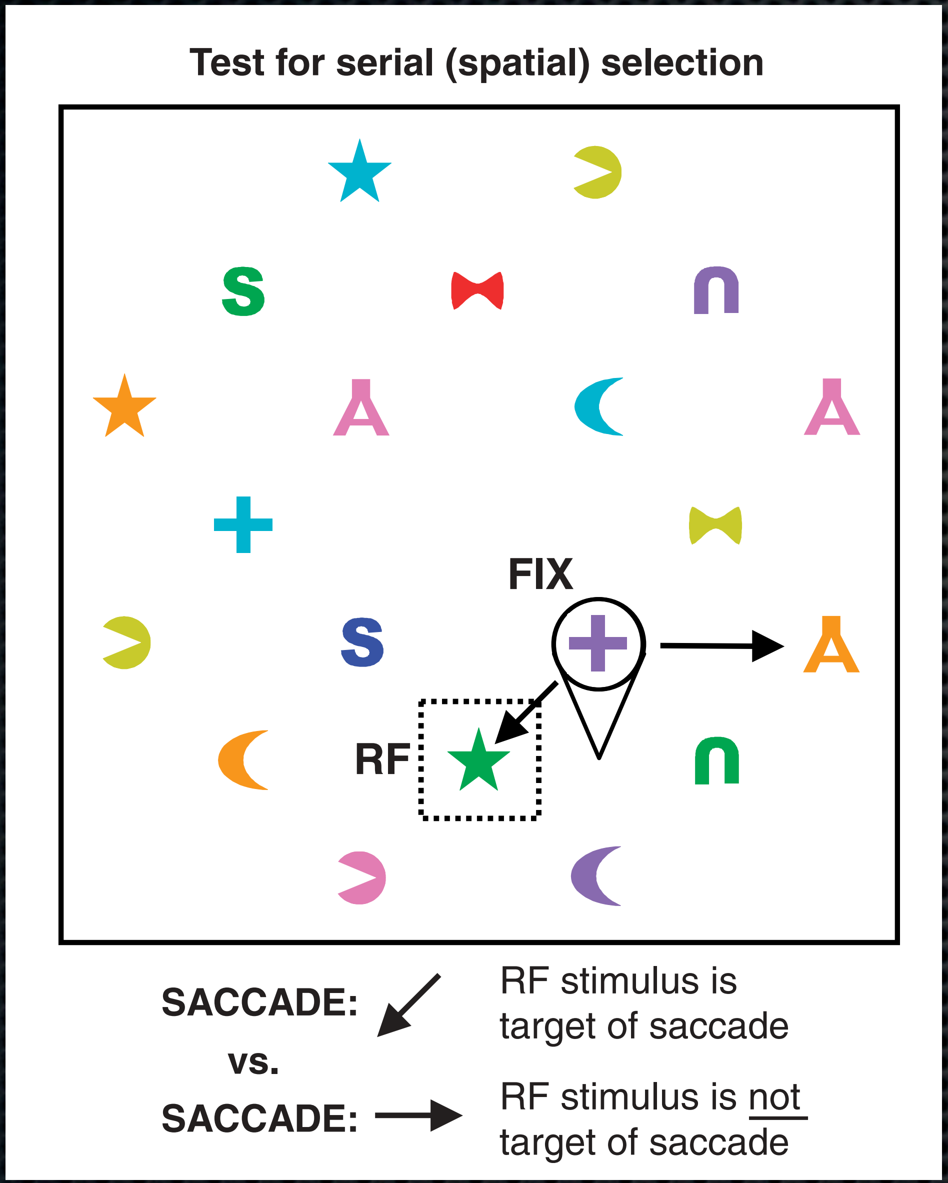
Narcisse P. Bichot, Andrew F. Rossi, Robert Desimone

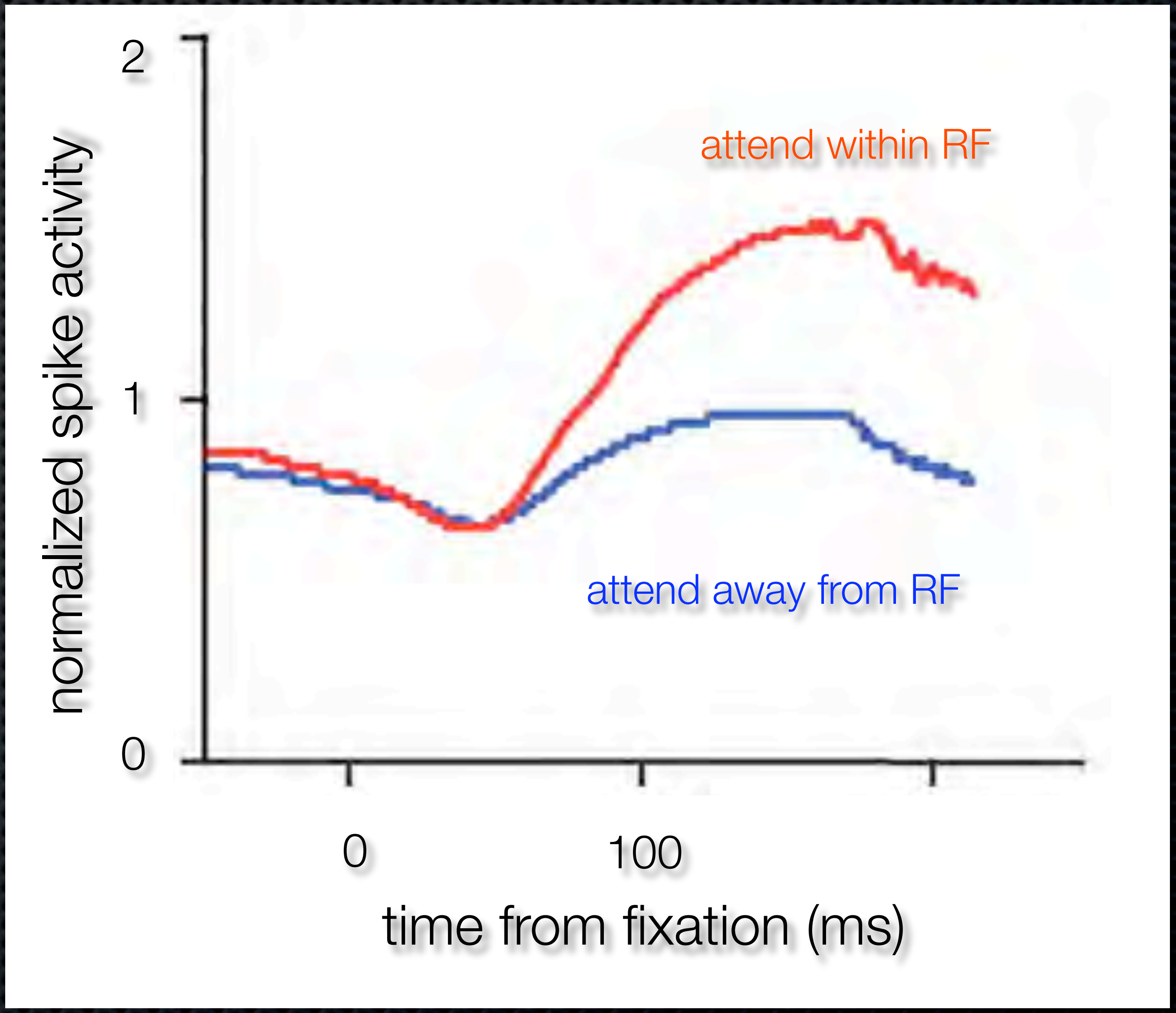
Answer: Parallel feature-based attention

Parallel feature-based attention modulation

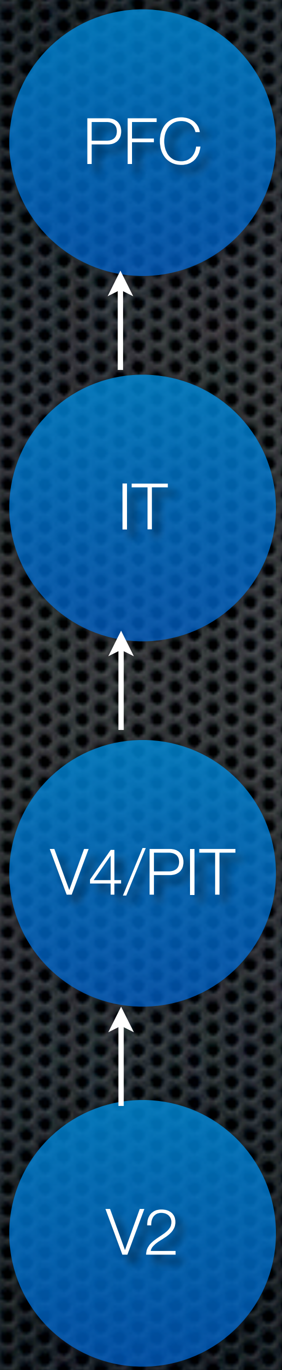


Serial spatial attention modulation





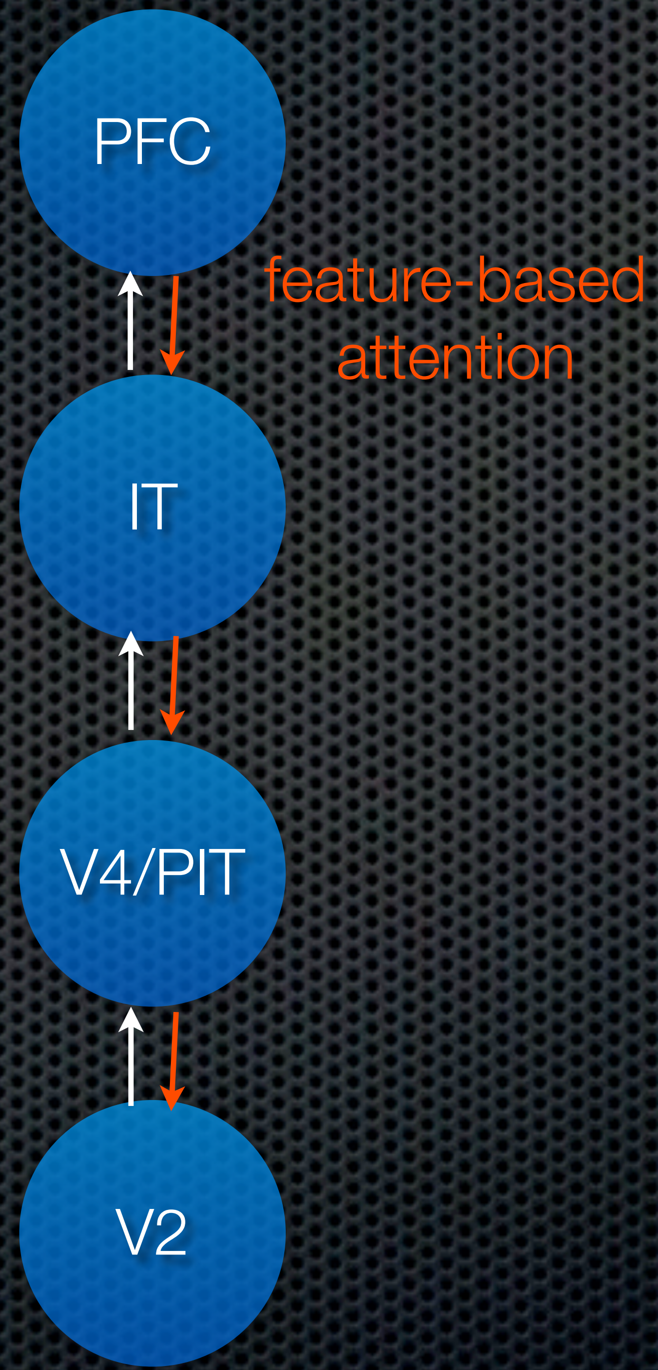
Attention as Bayesian inference



see also Rao 2005; Lee & Mumford 2003

Chikkerur Serre & Poggio in prep

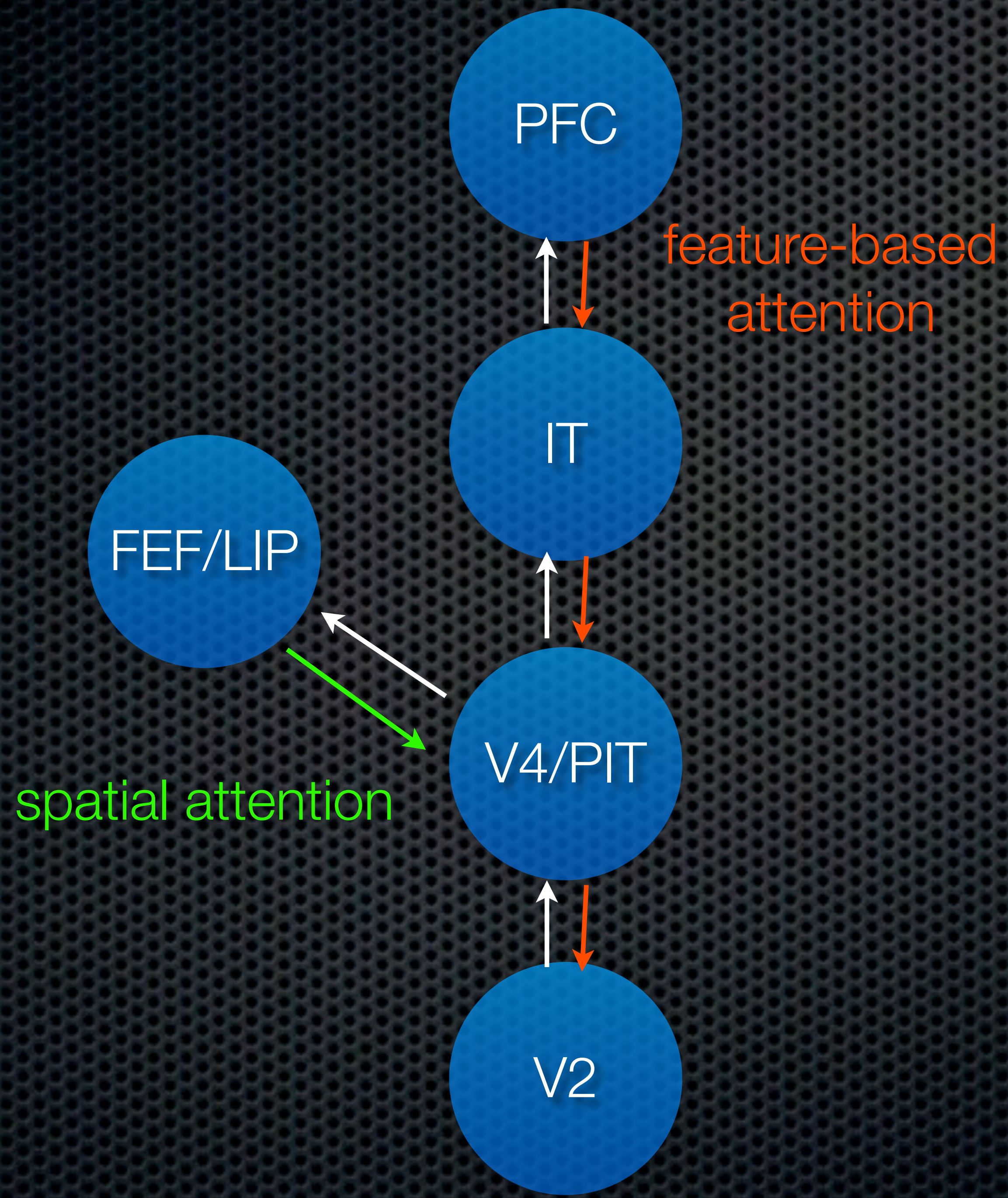
Attention as Bayesian inference



see also Rao 2005; Lee & Mumford 2003

Chikkerur Serre & Poggio in prep

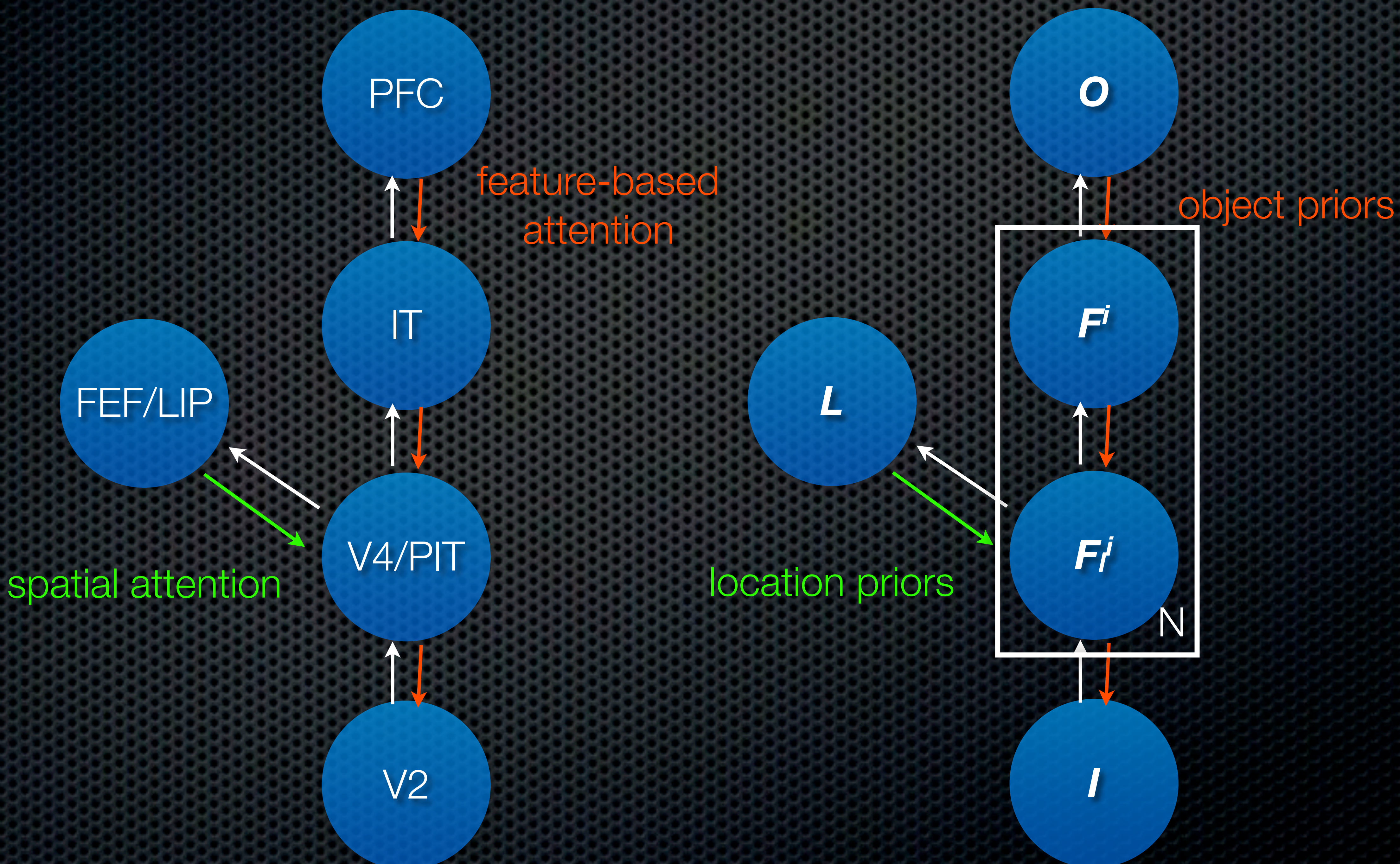
Attention as Bayesian inference



see also Rao 2005; Lee & Mumford 2003

Chikkerur Serre & Poggio in prep

Attention as Bayesian inference



see also Rao 2005; Lee & Mumford 2003

Chikkerur Serre & Poggio in prep