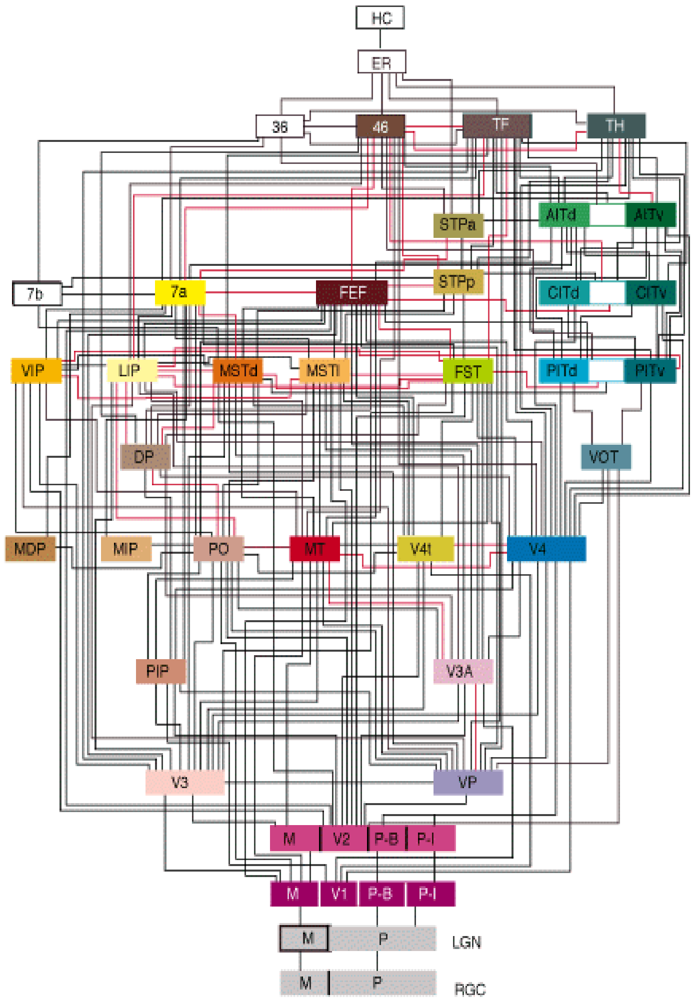
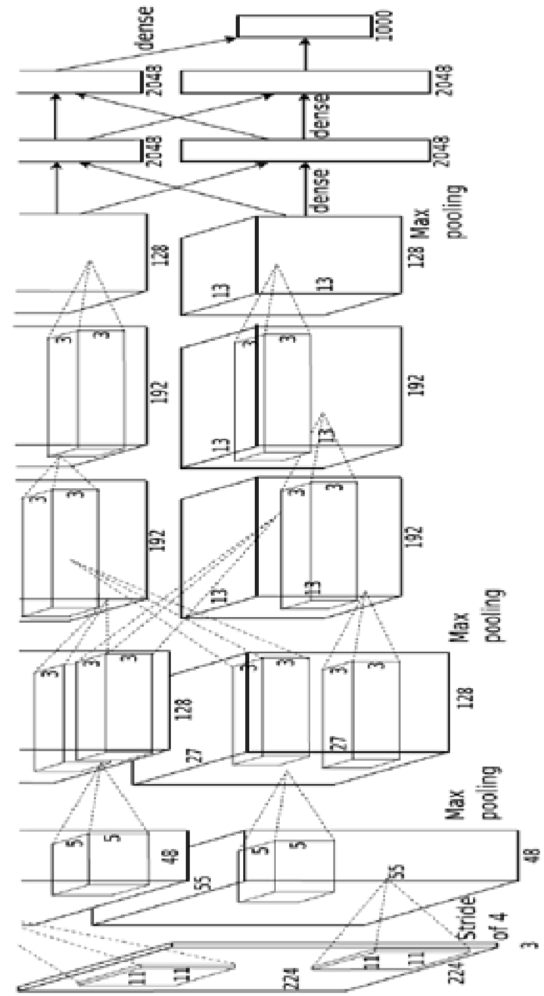
Biological and Artificial Vision

Or

Is the Primate Visual System a Good Model for Understanding ConvNets?



?

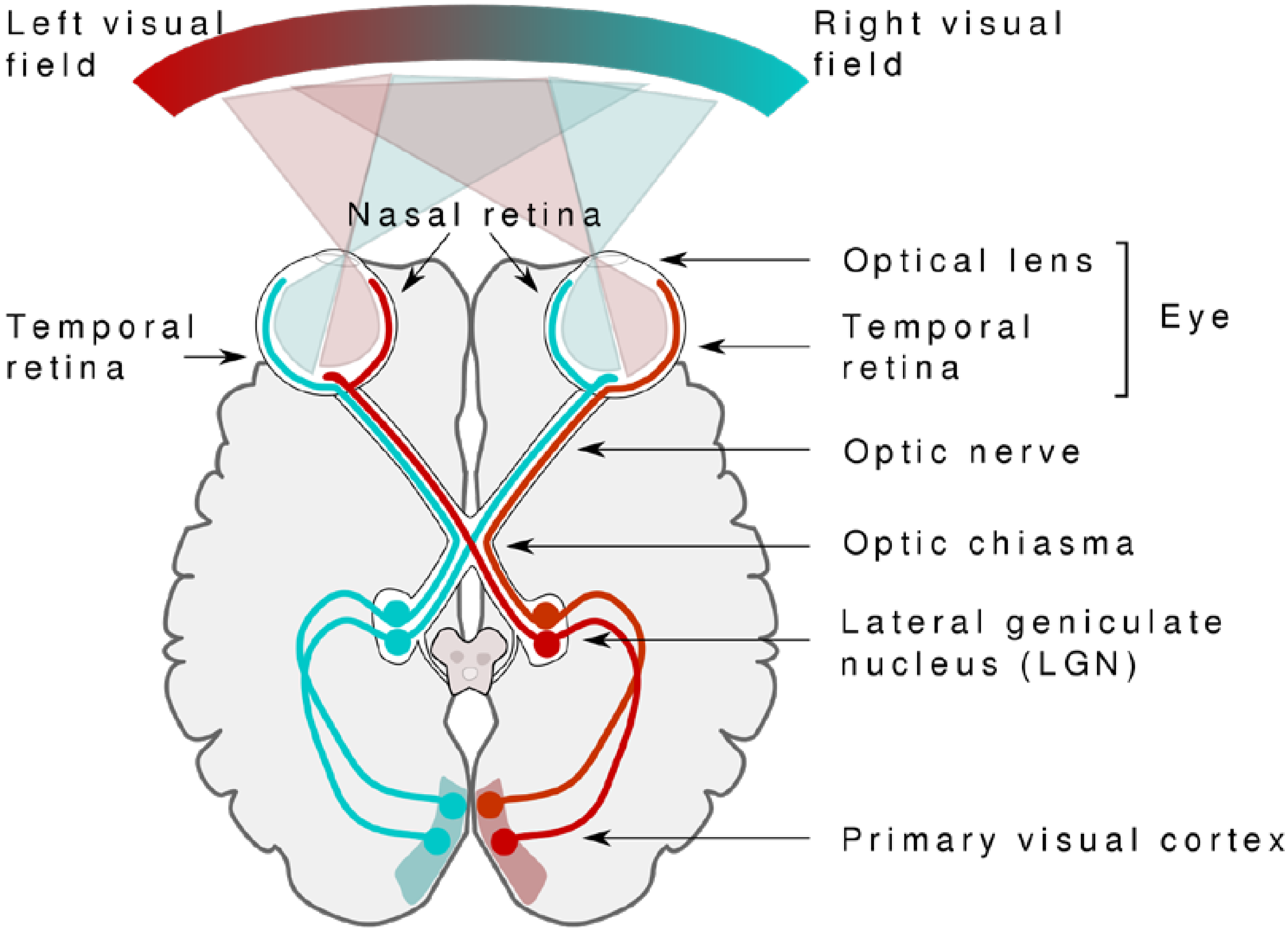


Roadmap

- anatomy of **visual pathways**
- concept of a **receptive field**
- hierarchy: more complex receptive fields can be constructed from simpler ones
- strengths and weaknesses of **CNNs** as a tool for vision science
- what might be missing from current machine vision algorithms?

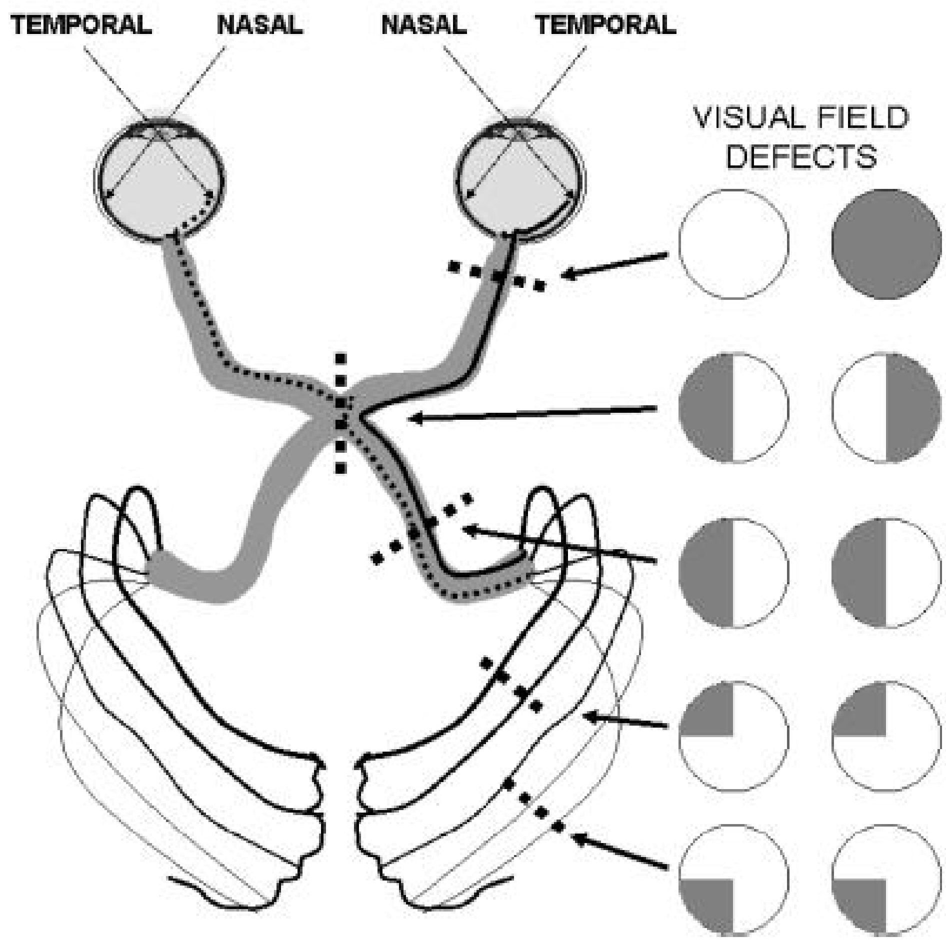
Questions or comments: richard_born@hms.harvard.edu

Anatomy of the visual pathways



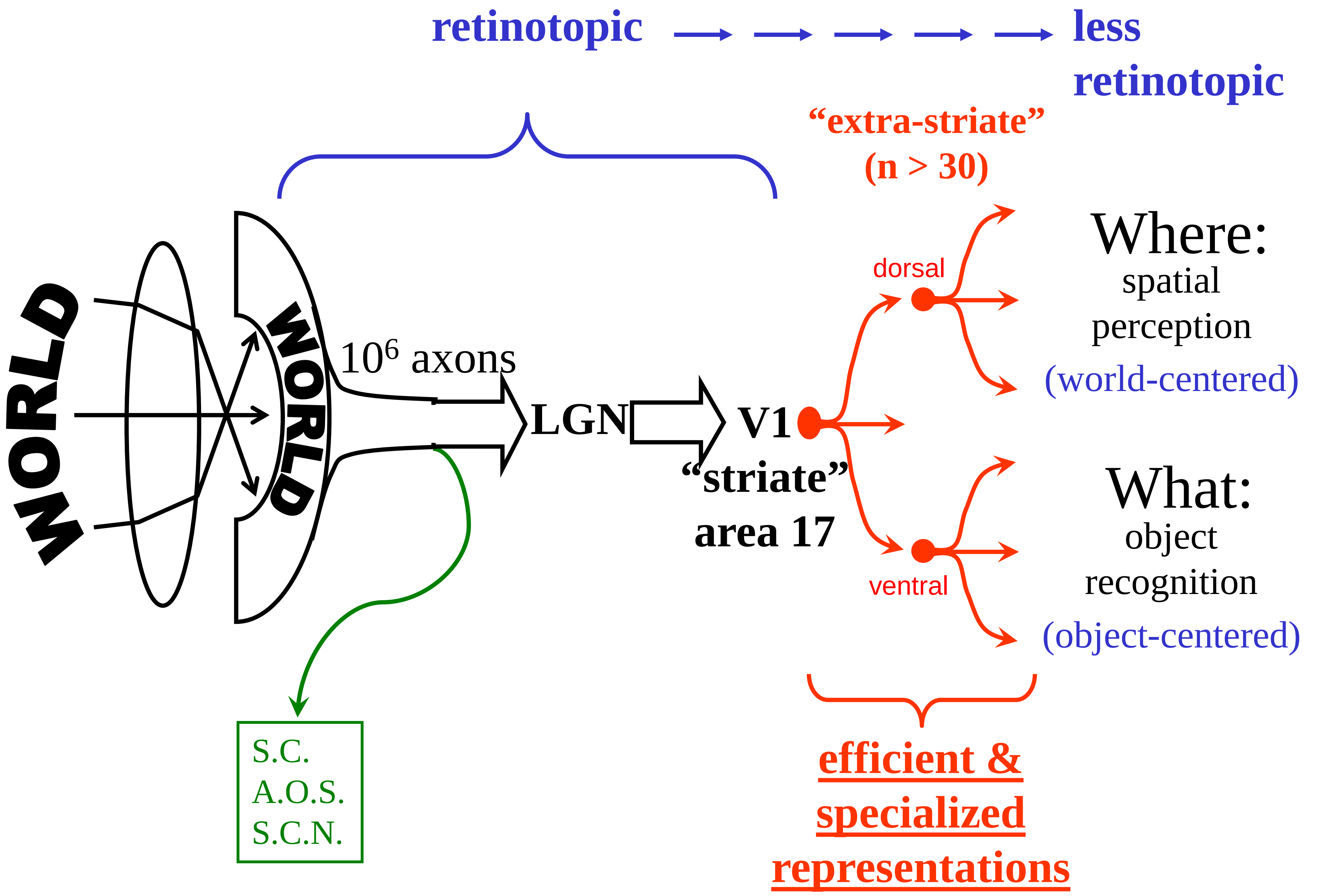
Visual field cuts to localize lesions

Figure 2. The Visual Field Defects Associated With The Various Possible Locations Of A Pathological Lesion

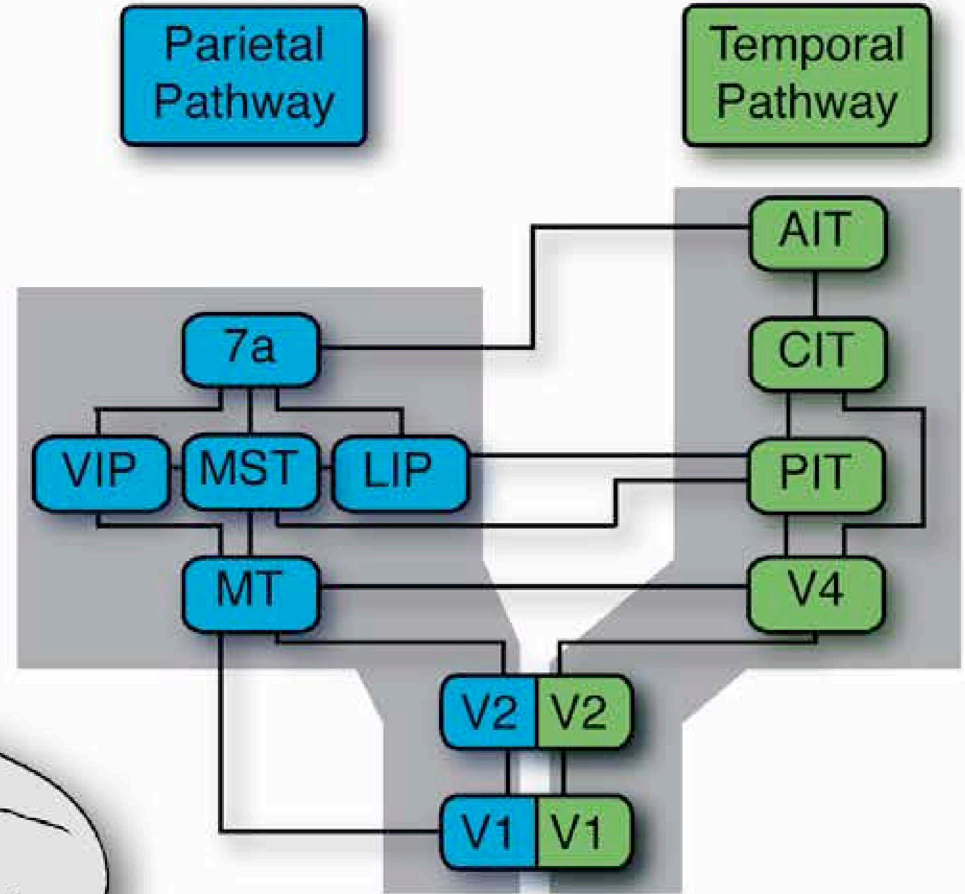


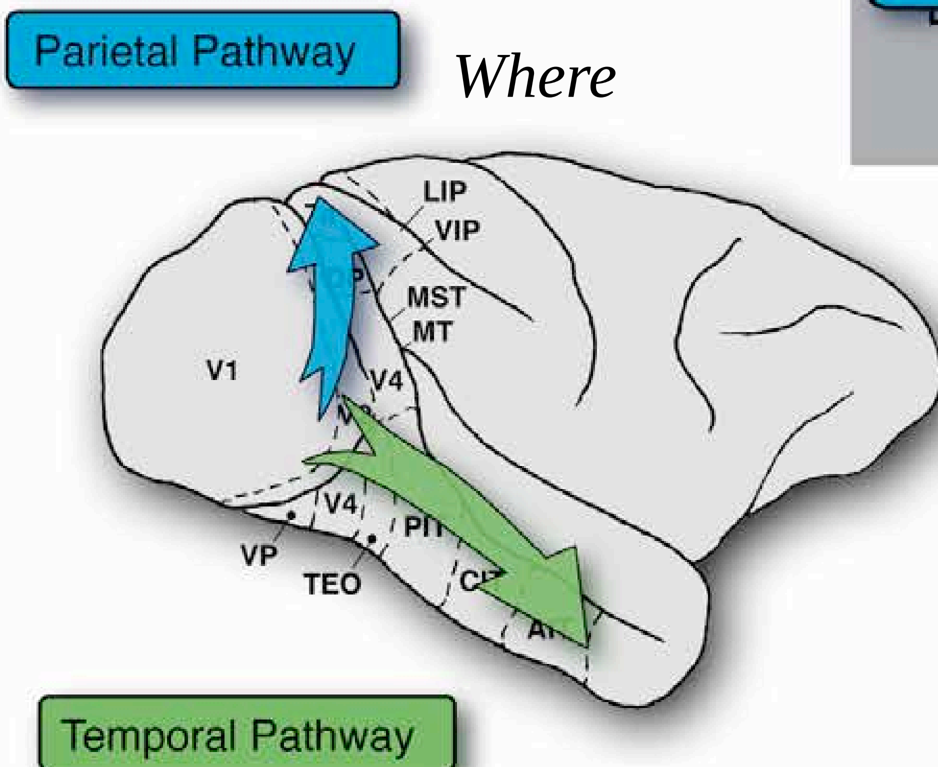
http://www.cksu.com/

How vision works: the Big Picture



Parallel pathways in visual cortex

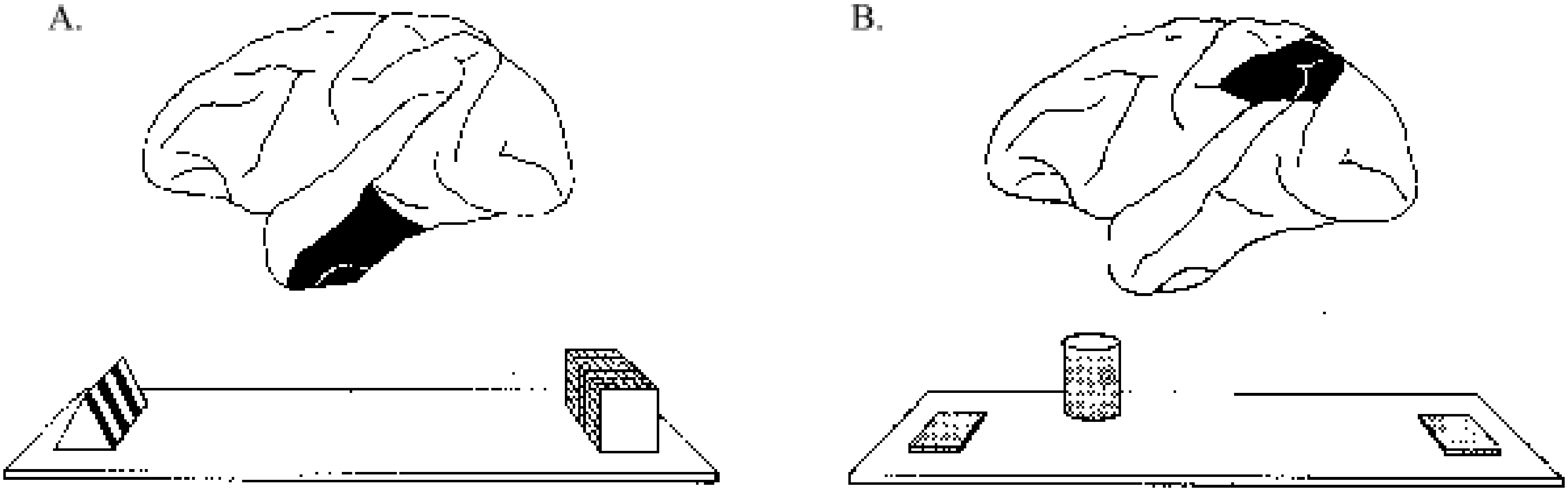




Where

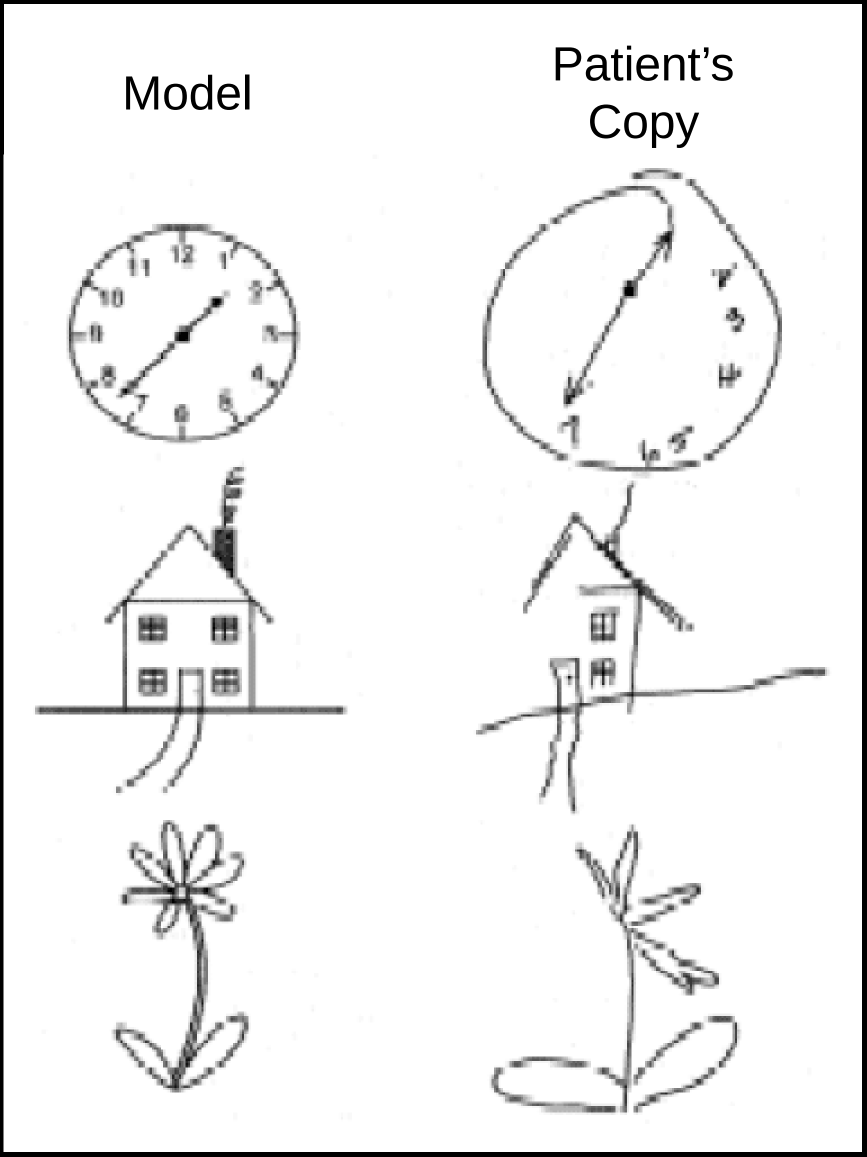
What

What? versus Where?

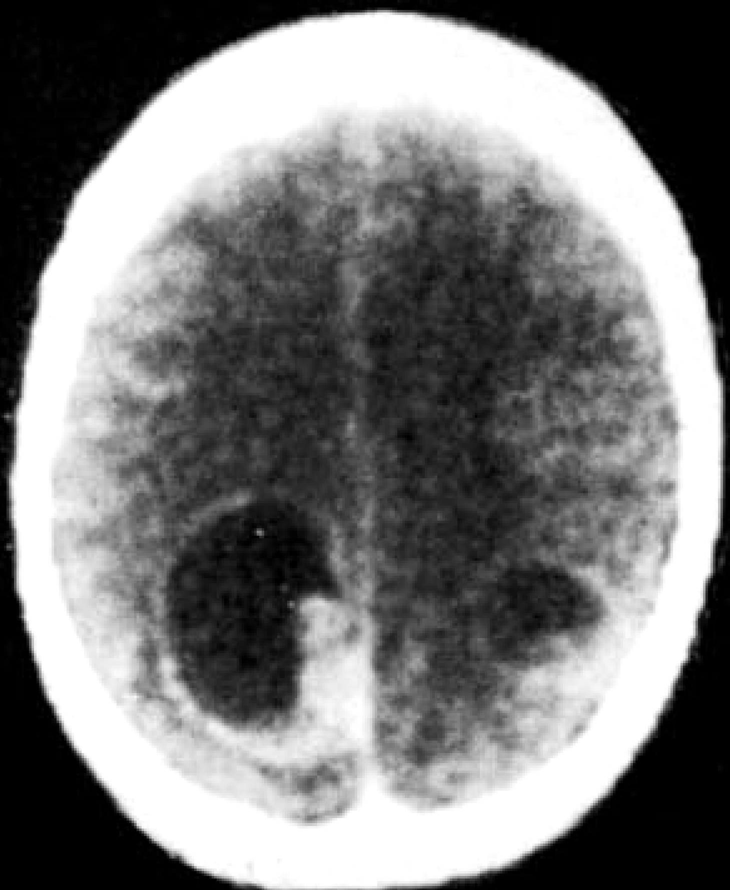


Mishkin et al. 1983

Posterior parietal lesion: hemispatial neglect



Where?: Balint’s syndrome from lesions of posterior parietal cortex

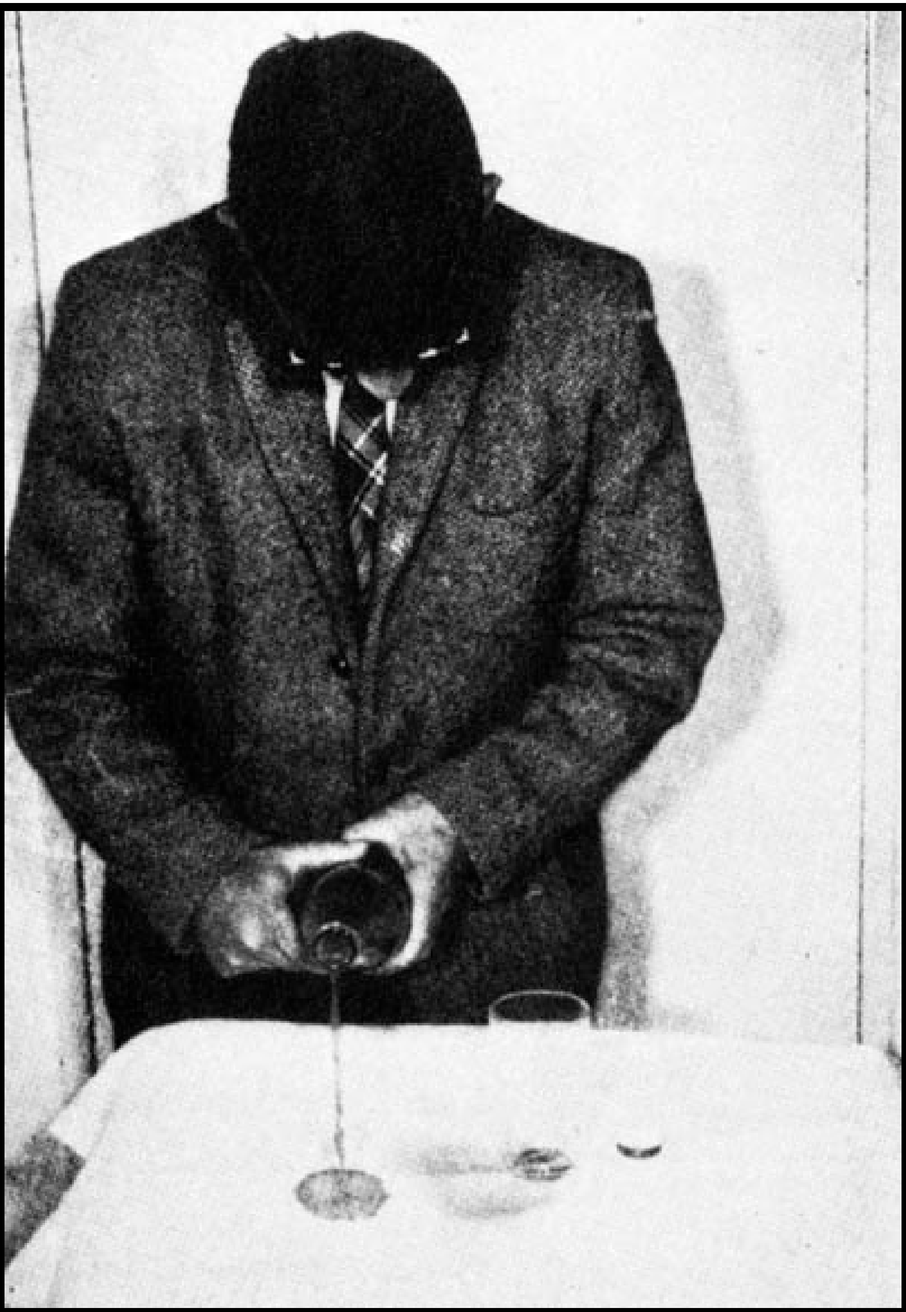


1. Visual Disorientation
(trees, but no forest)

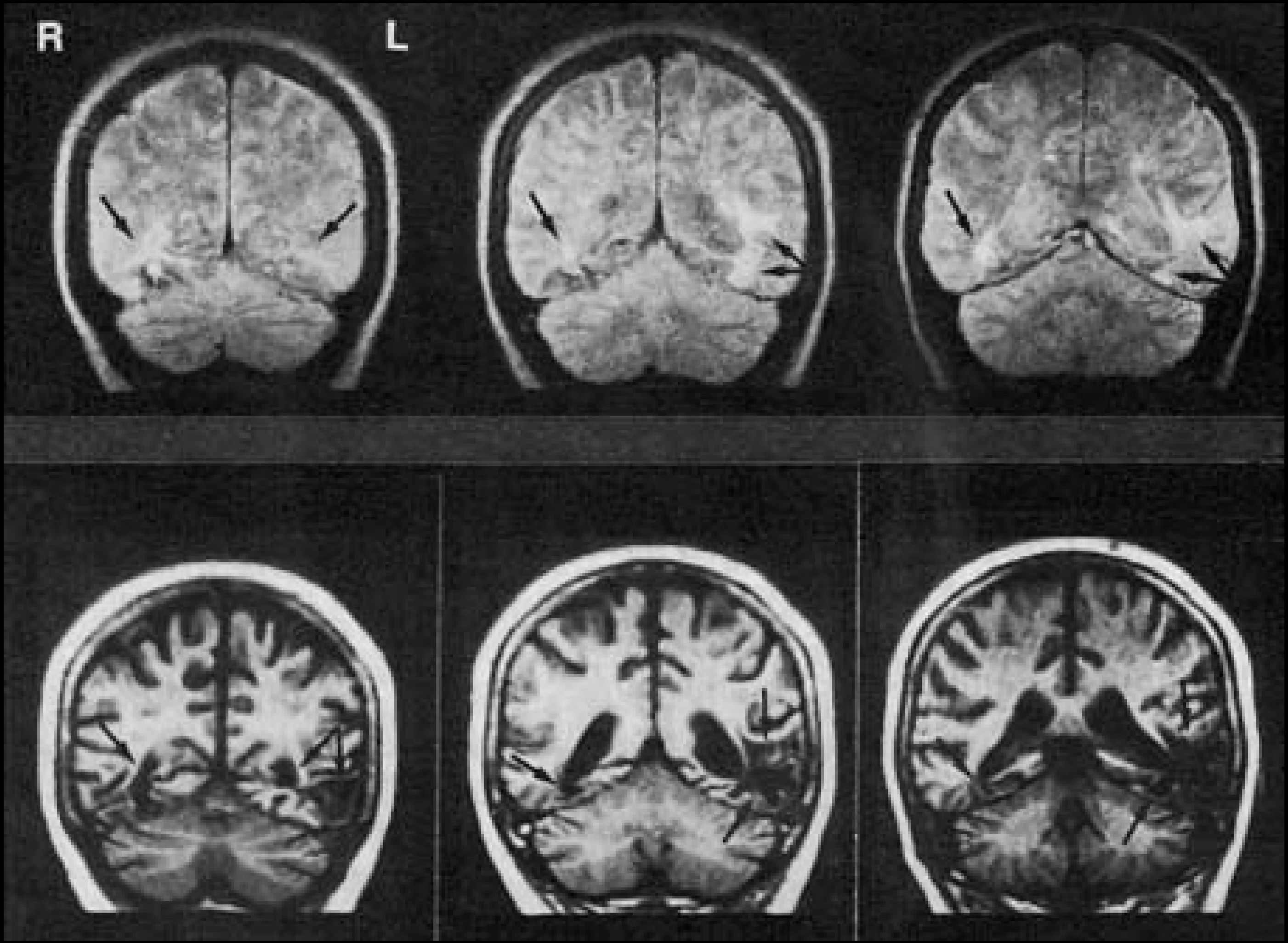
2. Optic Ataxia
(failure of visually guided pointing)

3. Ocular Apraxia
(unable to redirect gaze to targets)

Where?: A person with Balint's syndrome attempting to pour water into a glass



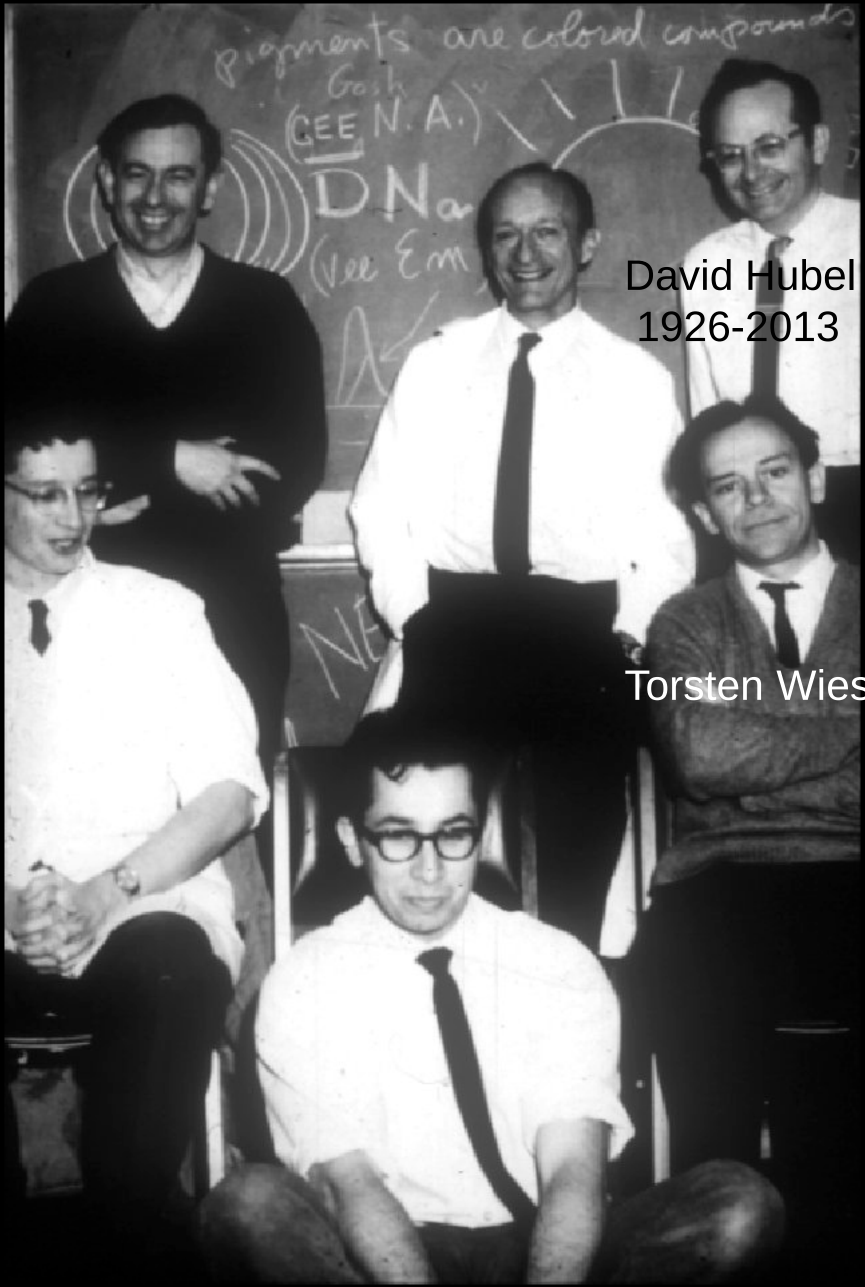
What?: Visual agnosias are produced by lesions of inferior temporal cortex



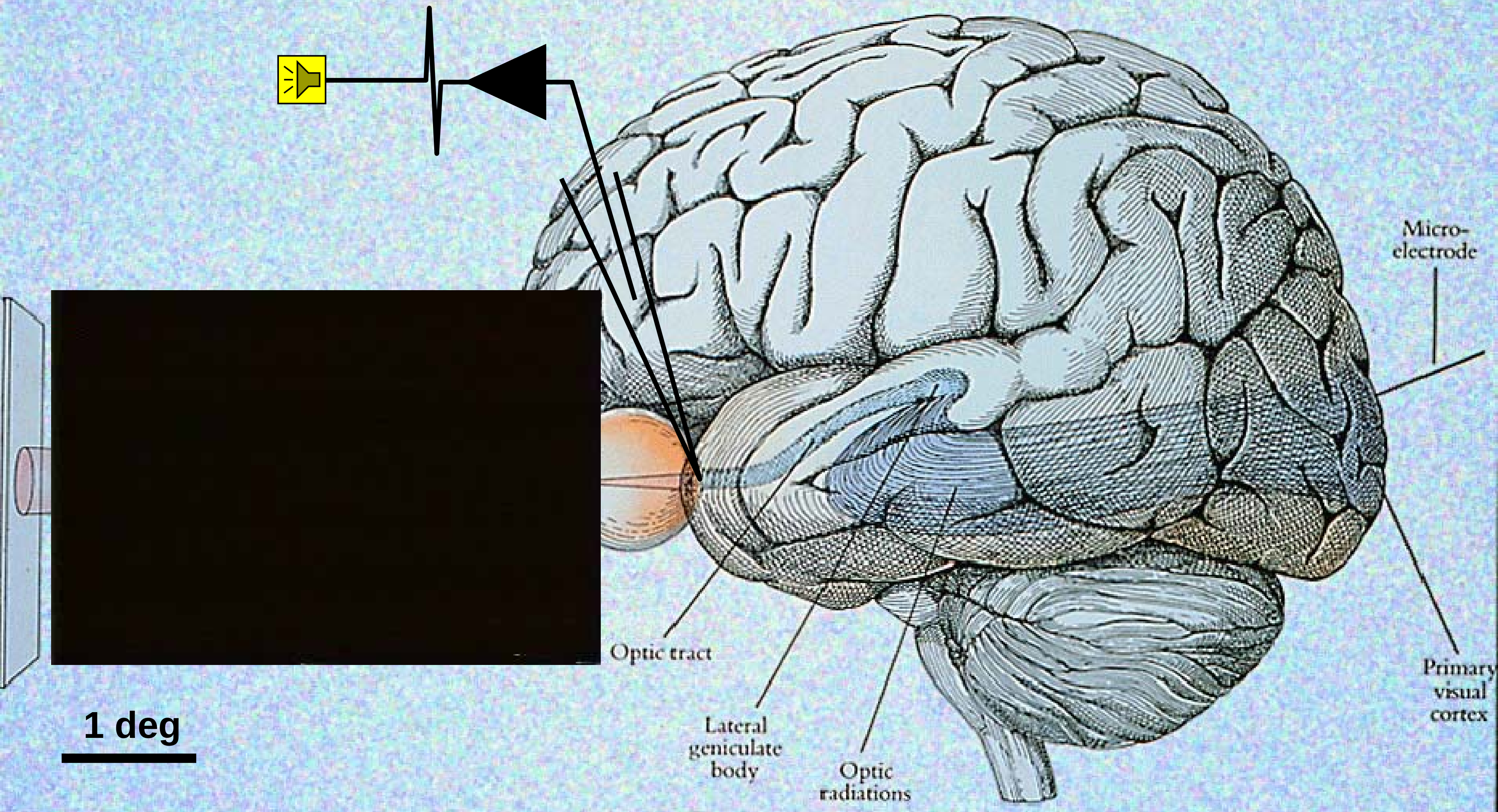
Receptive Field

- the region of a sensory epithelium that can influence a given neuron’s firing rate
- the particular pattern of the stimulus that maximally activates the neuron

Hubel & Wiesel

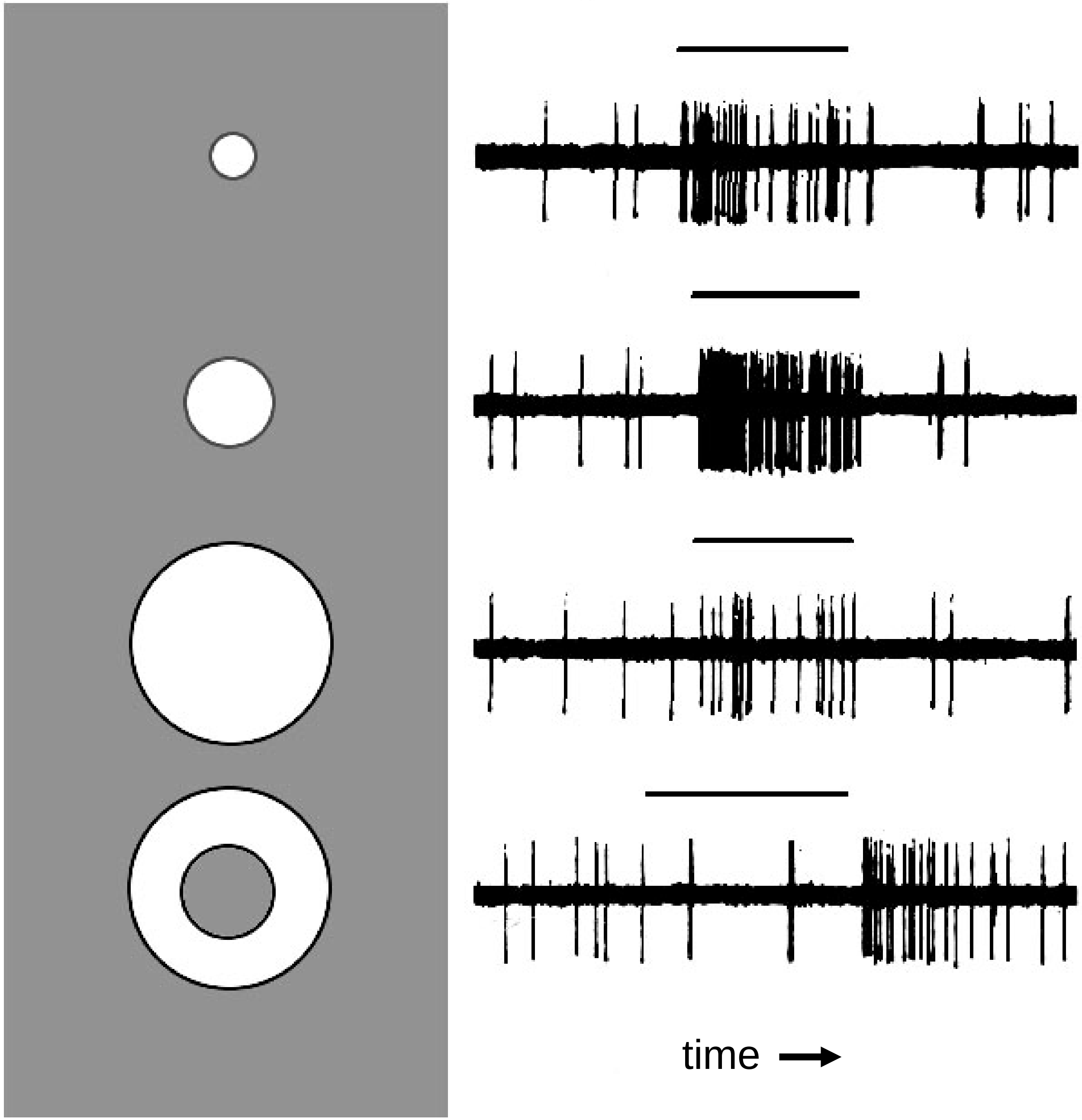


Center-surround organization of RGC receptive fields

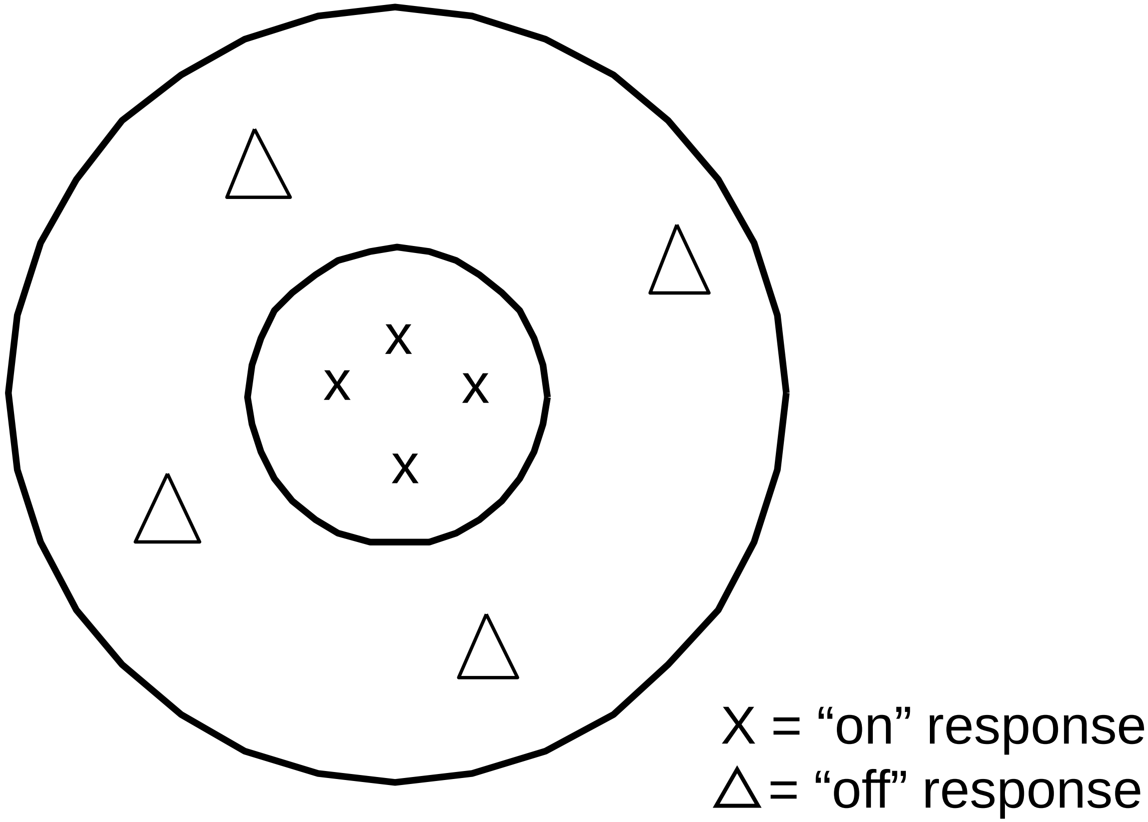


Hubel & Wiesel

Center-surround organization of RGC receptive fields



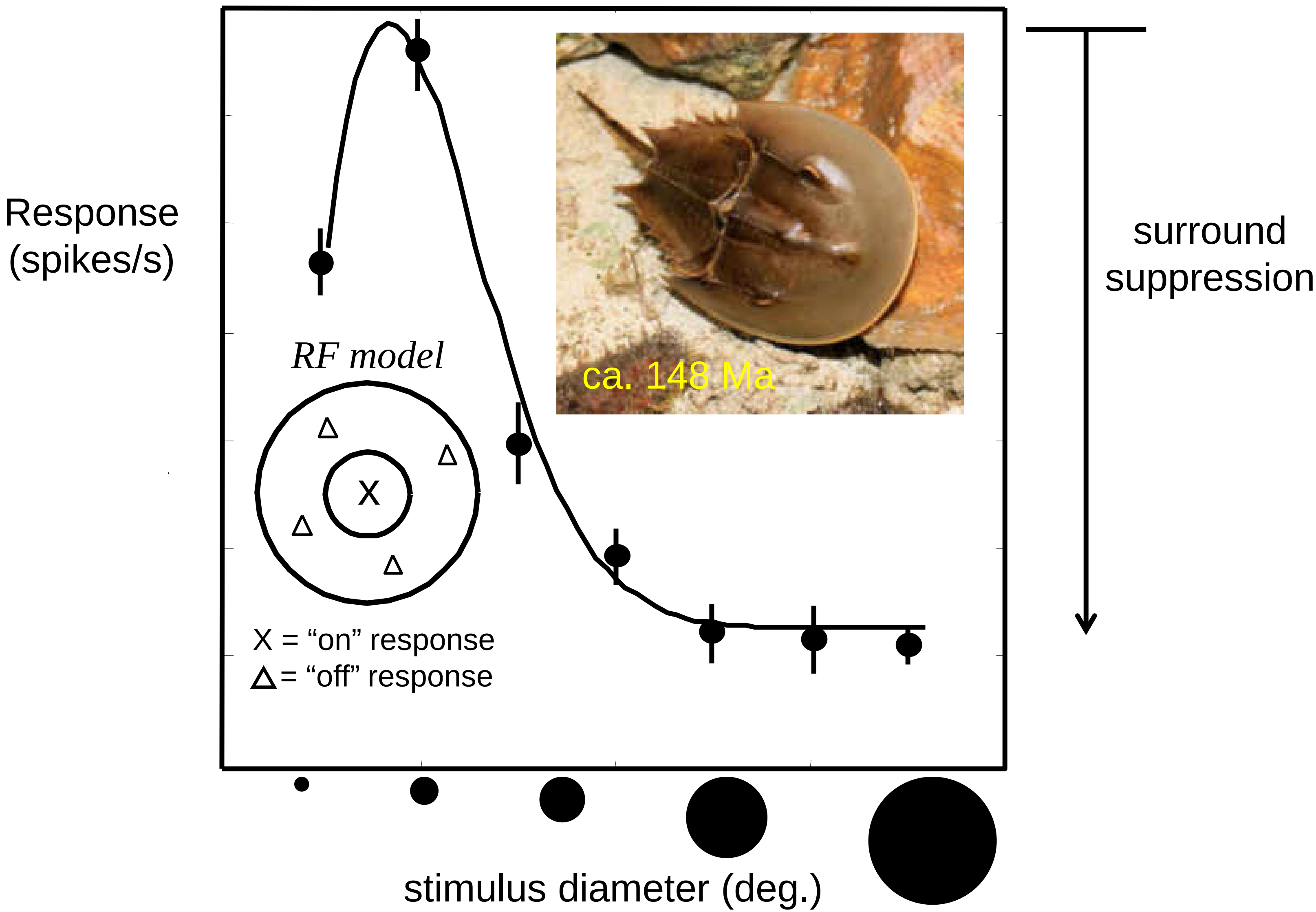
Center-surround organization of RGC receptive fields



“center-surround opponency”

H. K. Hartline 1940s & 50s

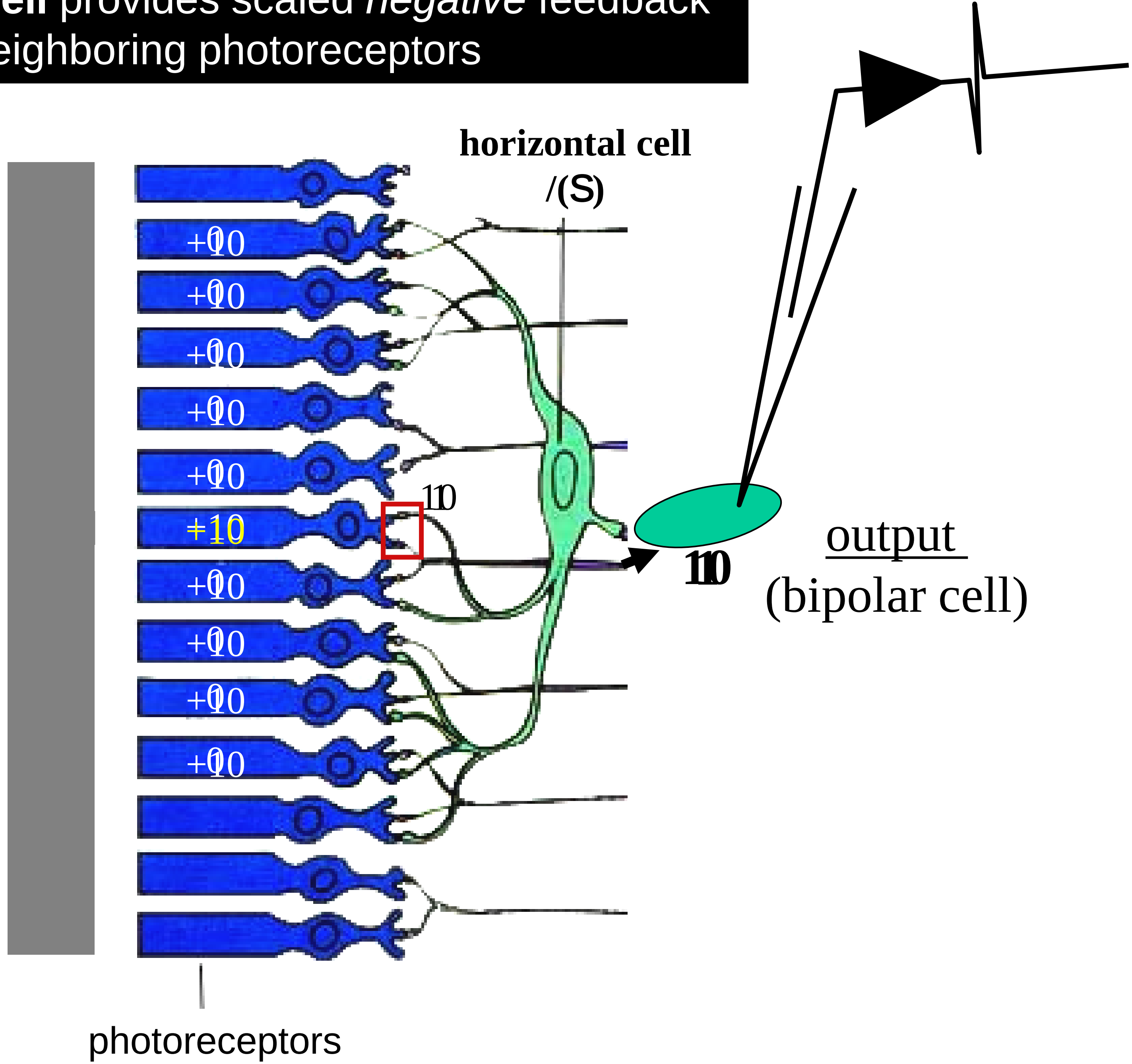
A very old receptive field



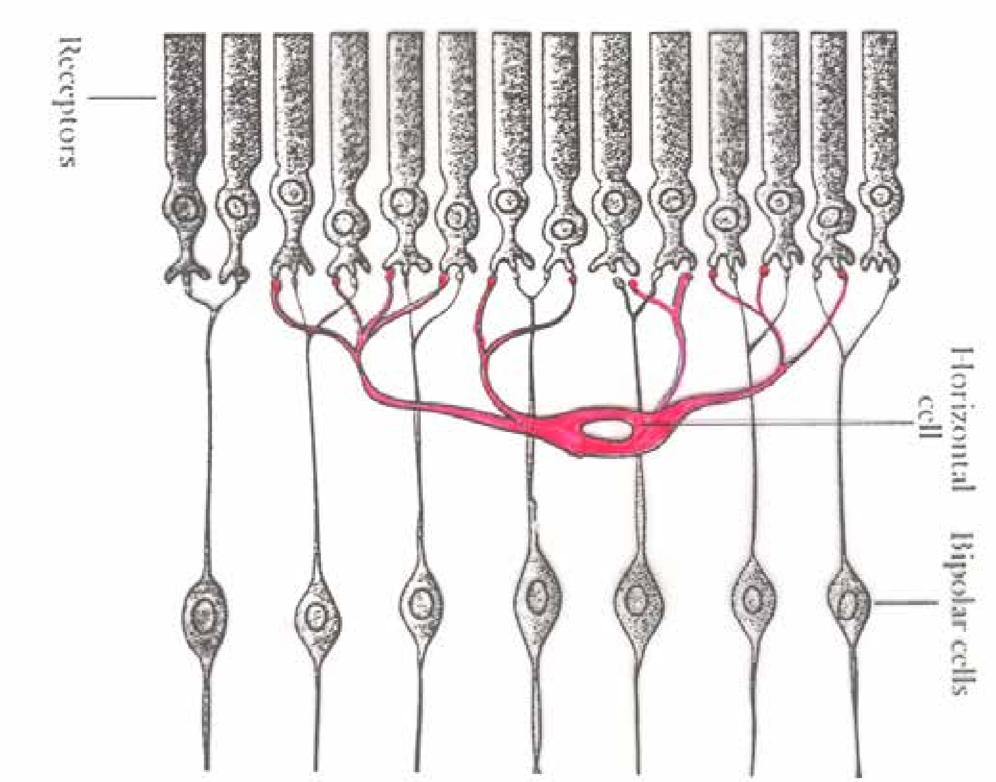
H.K. Hartline 1940s & 50s

The Problem of Limited Dynamic Range

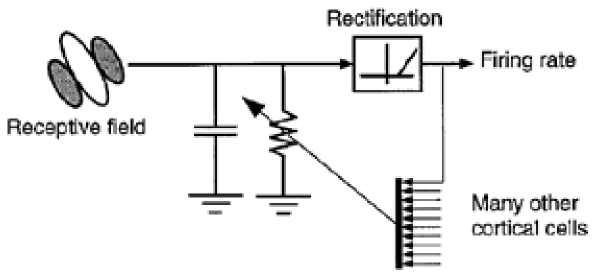
The **Horizontal Cell** provides scaled *negative* feedback from a group of neighboring photoreceptors



Lateral inhibition produces *normalization*

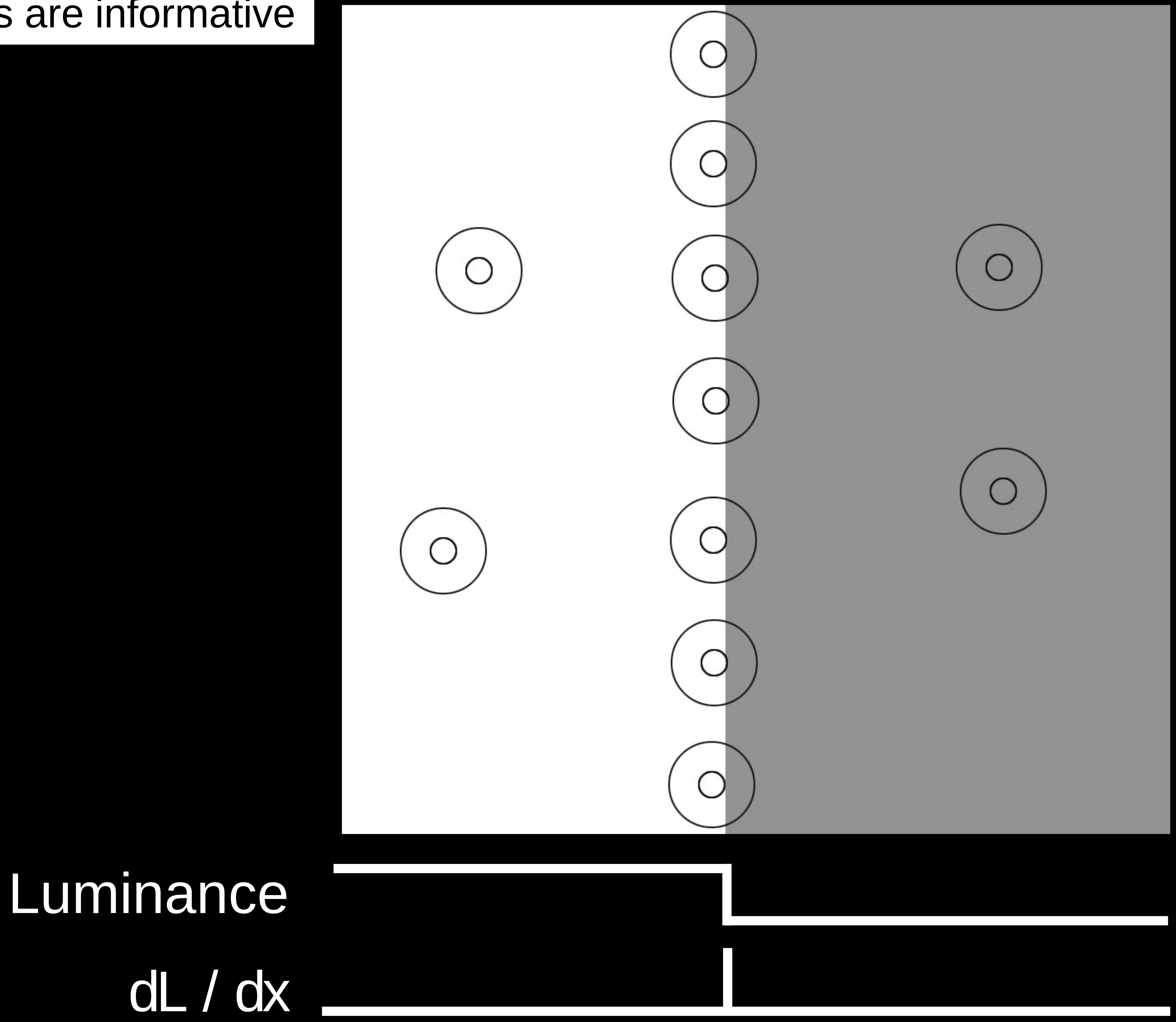


Cortical Normalization



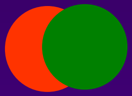
keeps population activity within a stable range

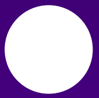
Edges are informative

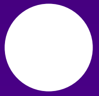


CHANGE

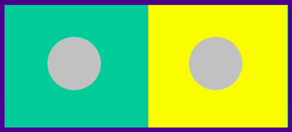
→ space = contrast 

→ wavelength = color contrast 

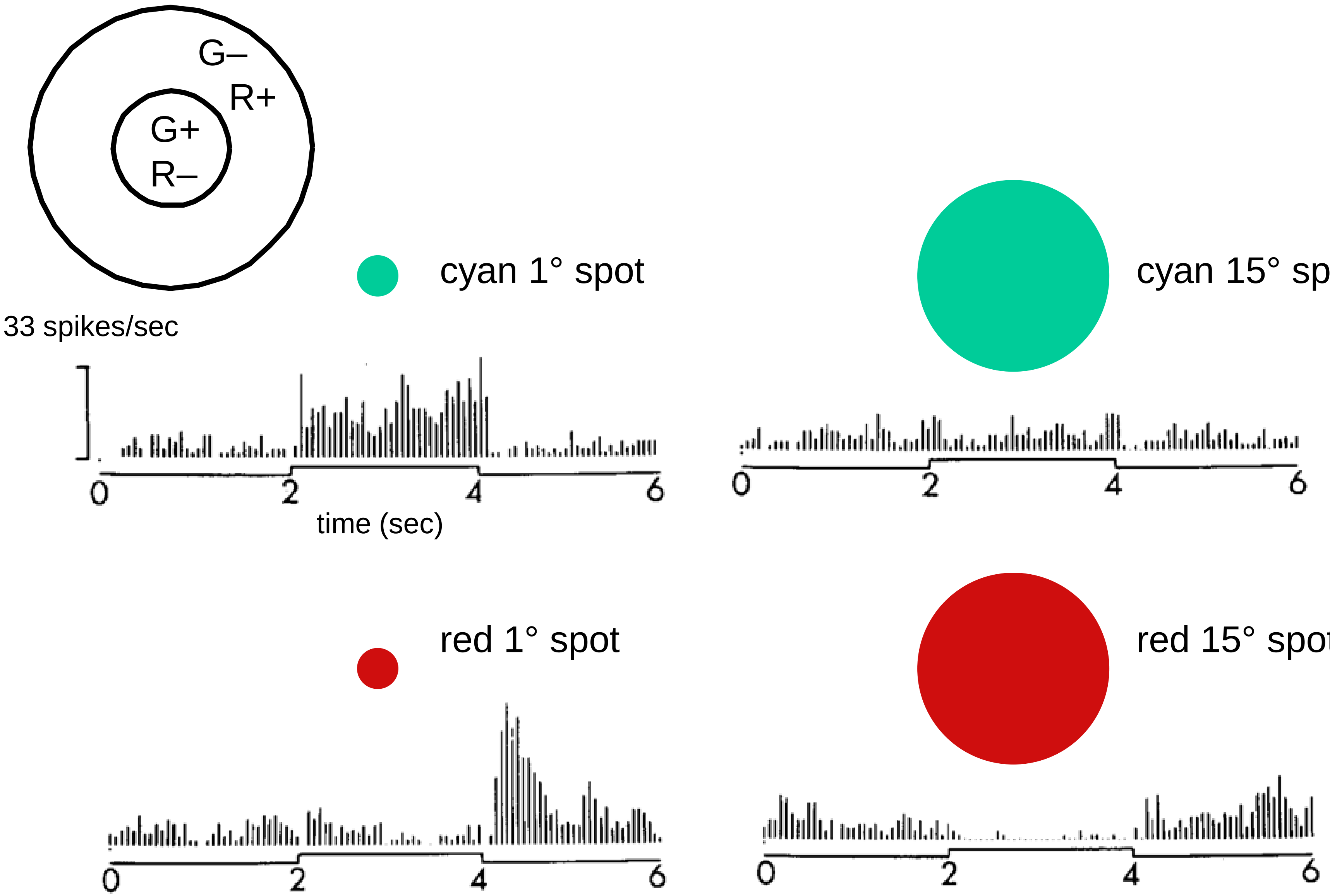
→ time = transience 

→ space/time = motion 

→ space/color = simult. color contrast



Double-opponency: space and chromaticity



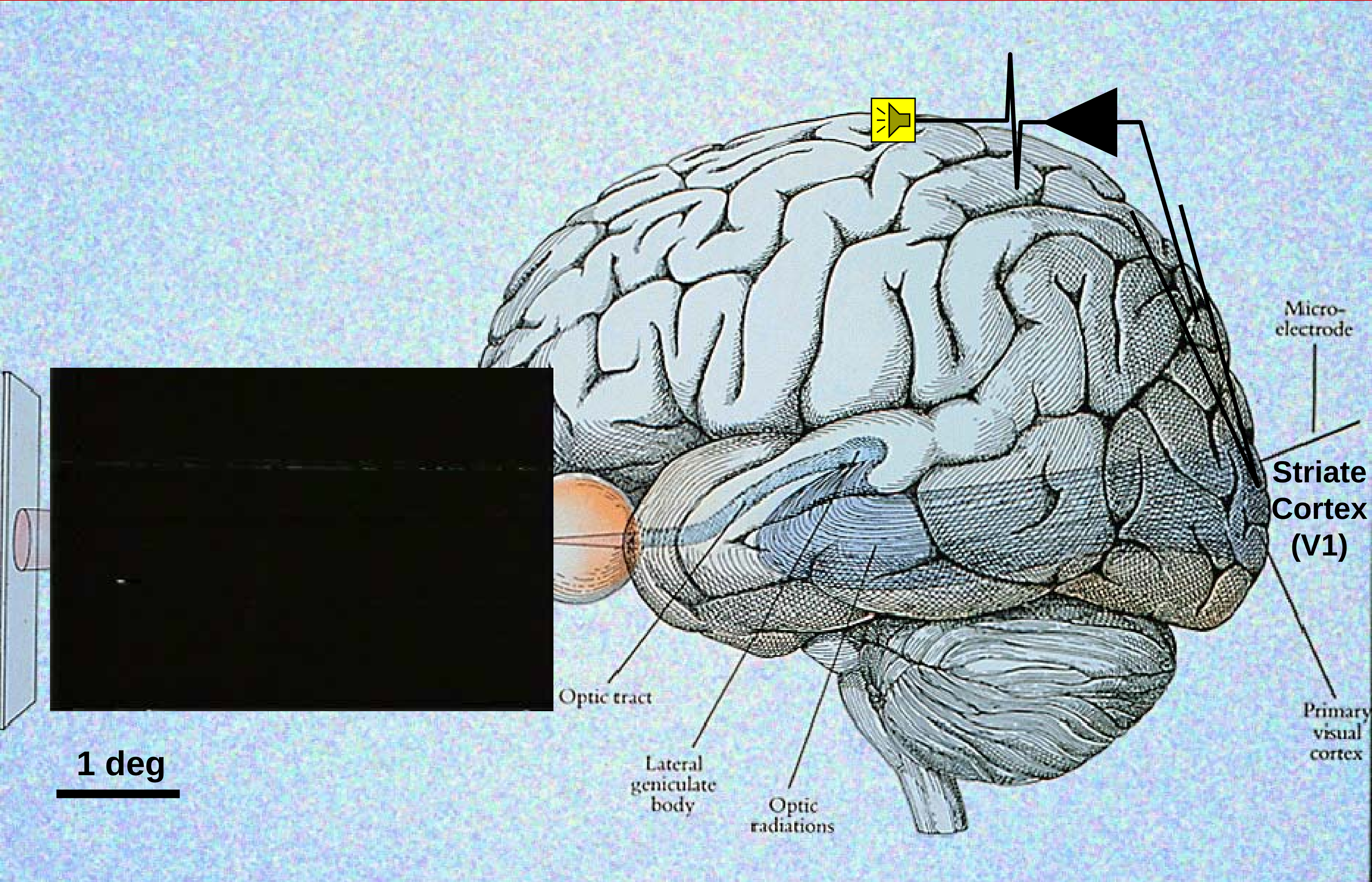
Livingstone & Hubel 1984

HIERARCHY

- “Hierarchical Elaboration of Receptive Fields”

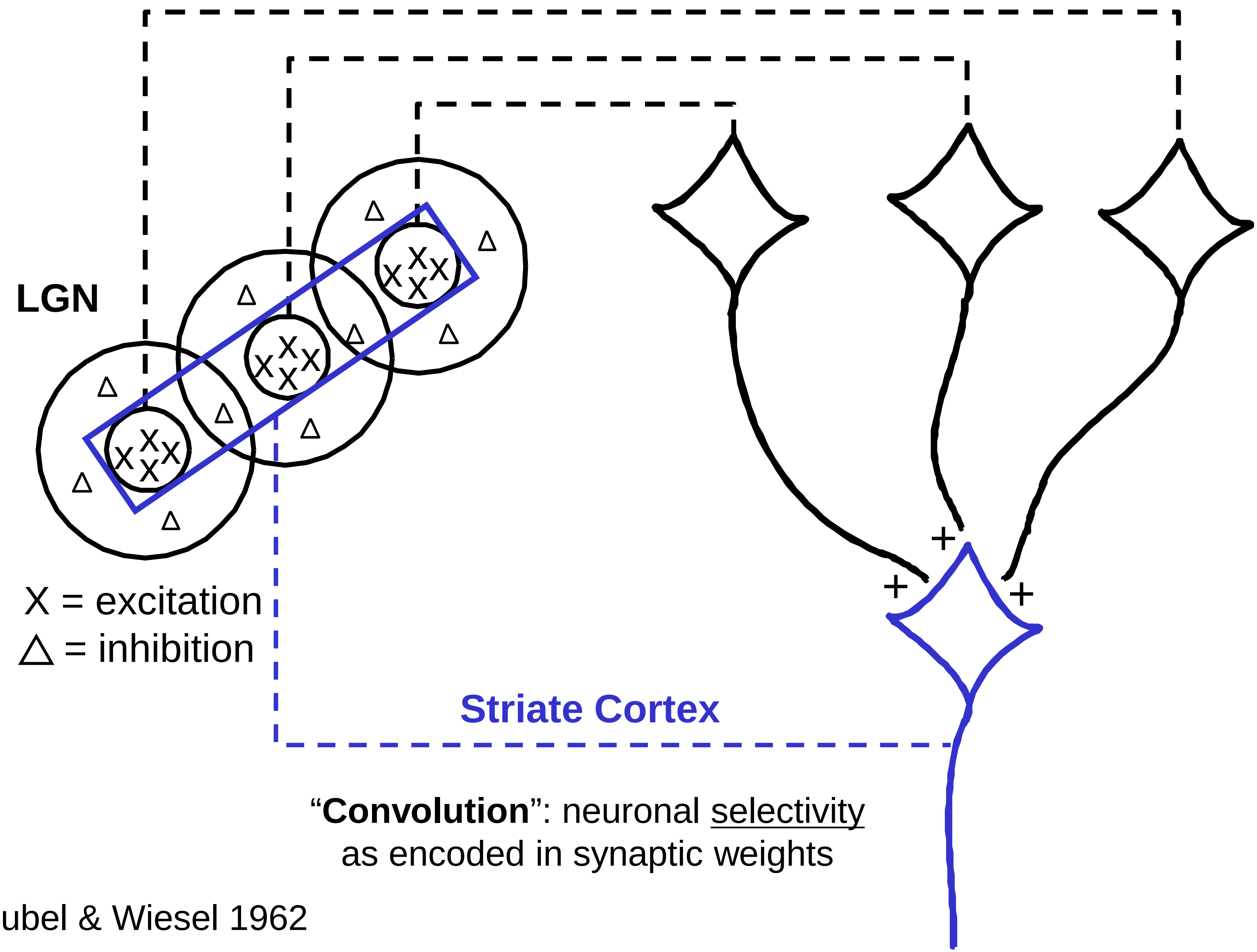
- Larger, more complex receptive fields are constructed by wiring up a bunch neurons with smaller, simpler receptive fields.

Simple cell in primary visual cortex (V1)



Hubel & Wiesel

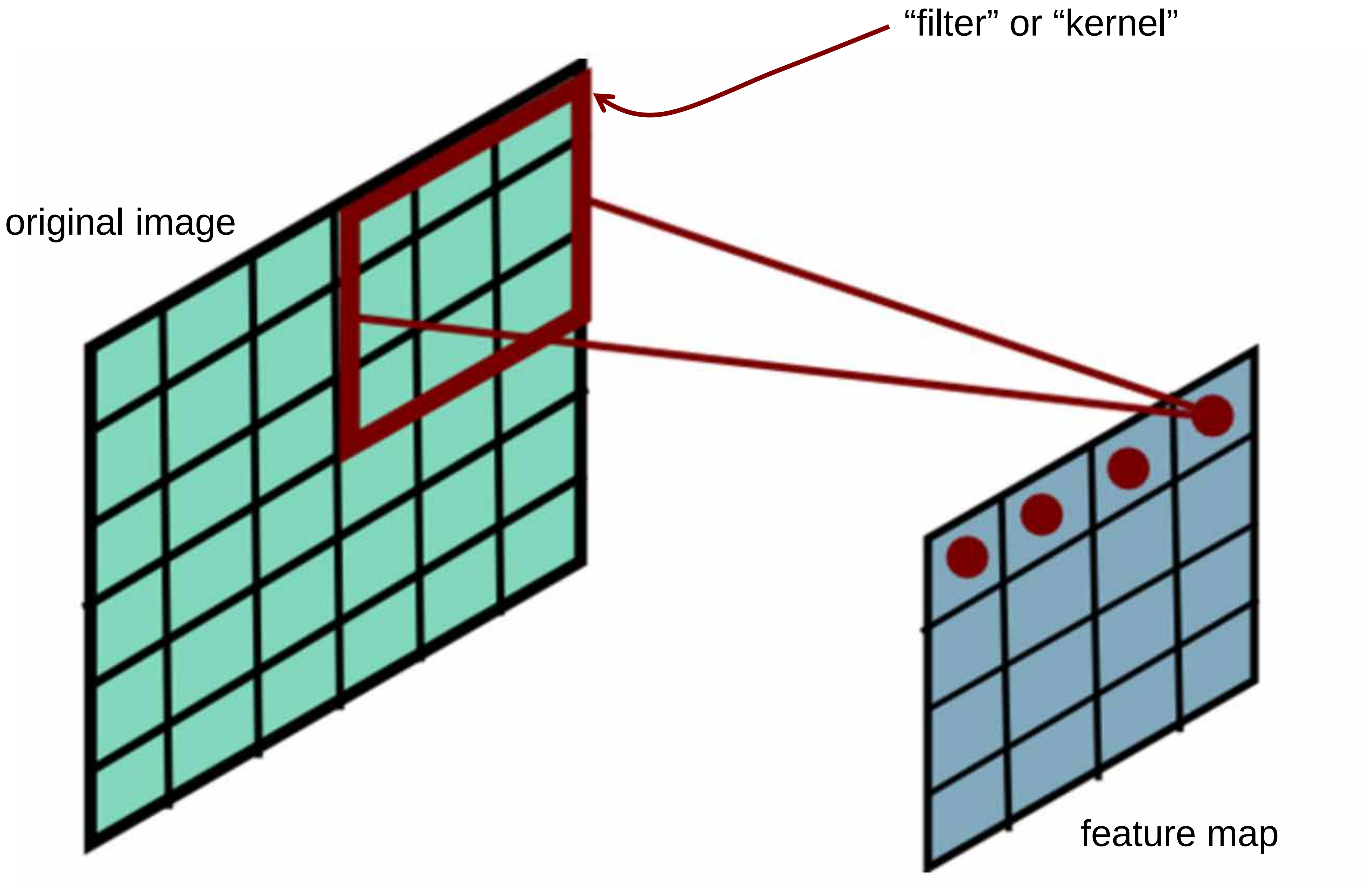
Circuit for a simple cell



"**Convolution**": neuronal <u>selectivity</u>
as encoded in synaptic weights

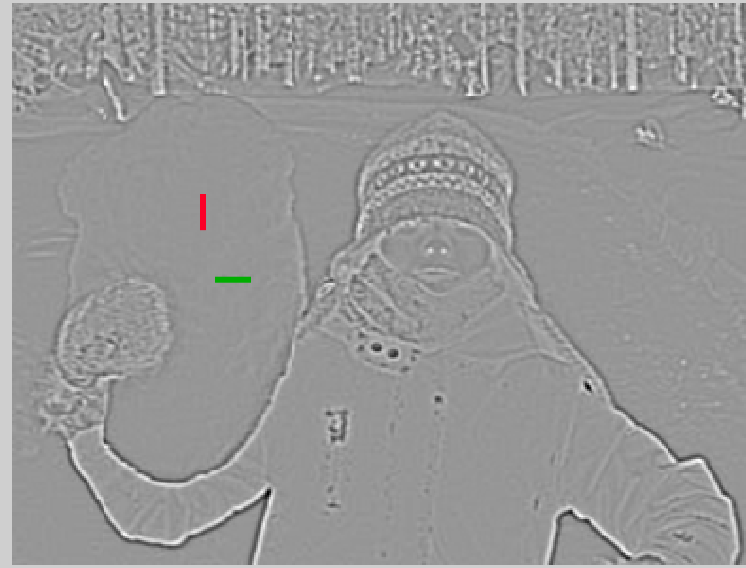
Hubel & Wiesel 1962

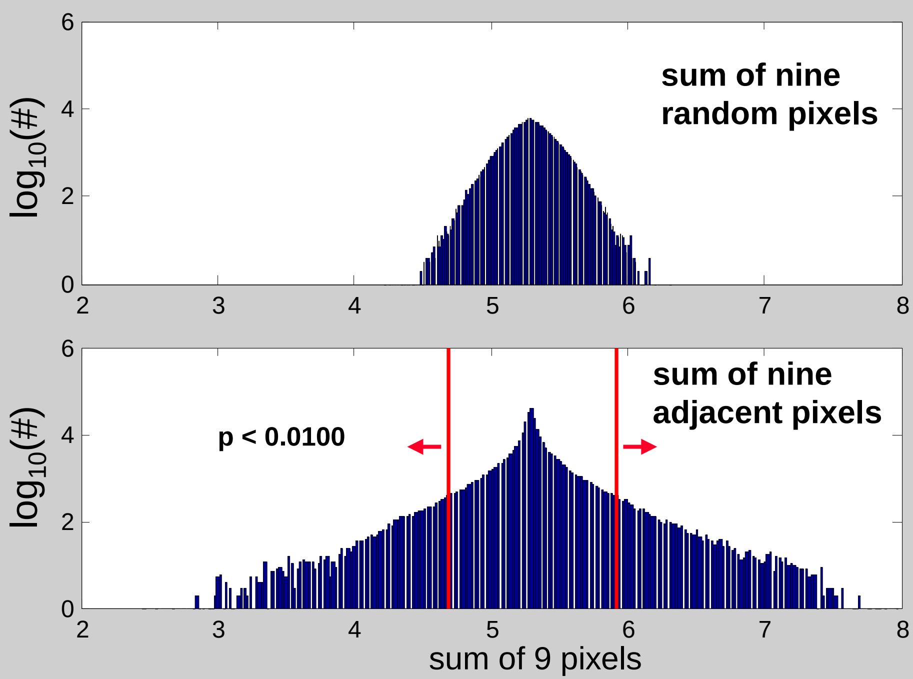
Convolution and filtering



Convolution is simple

"Whitened": Ñ^2×G or what ctr-sur does



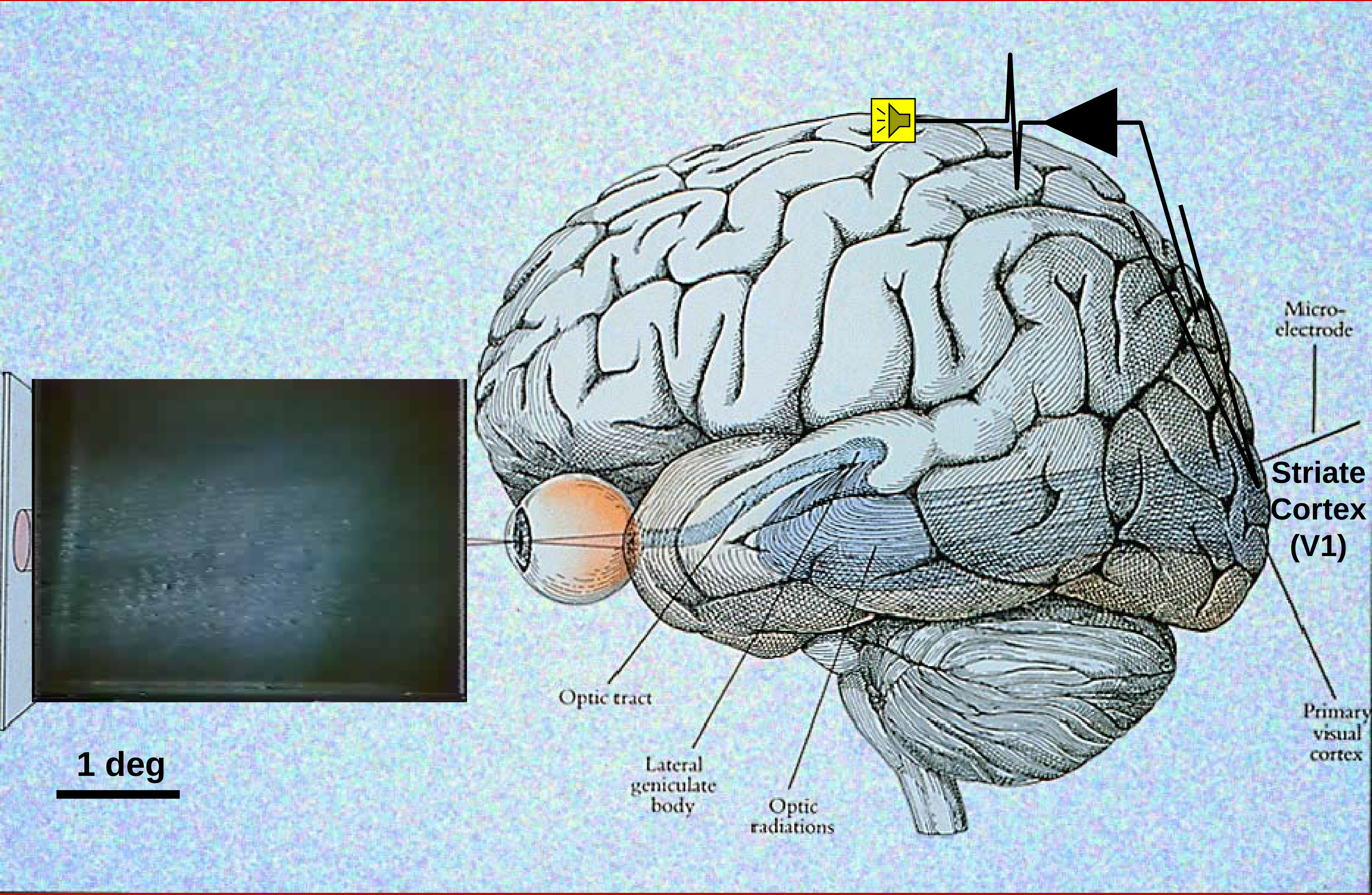


Suspicious Coincidences

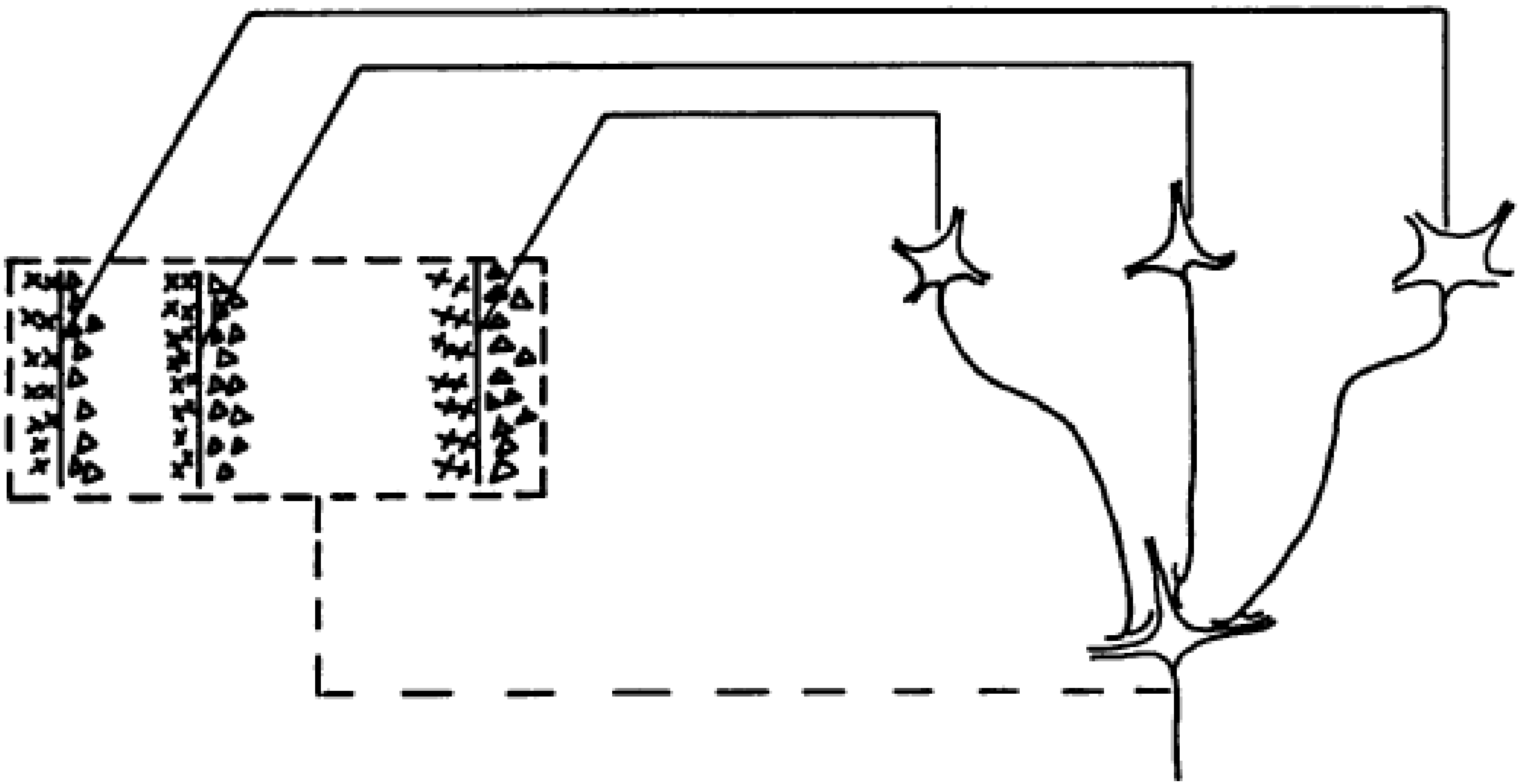
inspired by Horace Barlow

natural_images.m

Complex cell in primary visual cortex (V1)



Circuit for a complex cell



"**Pooling**": combines outputs with similar selectivities to achieve invariance

Hubel & Wiesel 1962

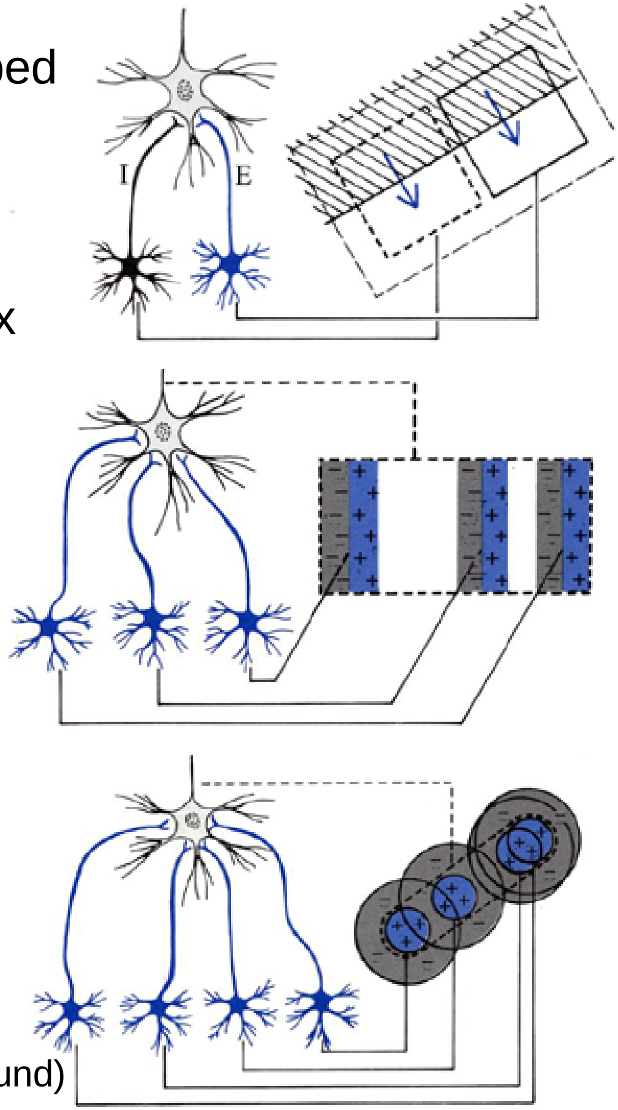
Efficient representations via hierarchical processing

End-stopped

Complex

Simple

LGN
(center-surround)



Hubel & Wiesel

<u>remove known correlations: curvature</u>

↑
↑

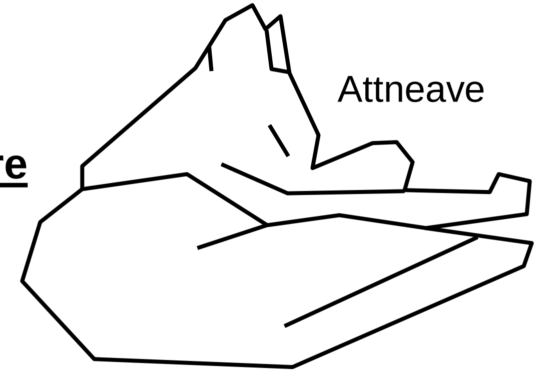
<u>invariance</u>
a) position
b) sign of contrast

↑
↑

<u>represent “suspicious coincidences”</u>

↑
↑

<u>remove known correlations: contrast</u>



Attneave

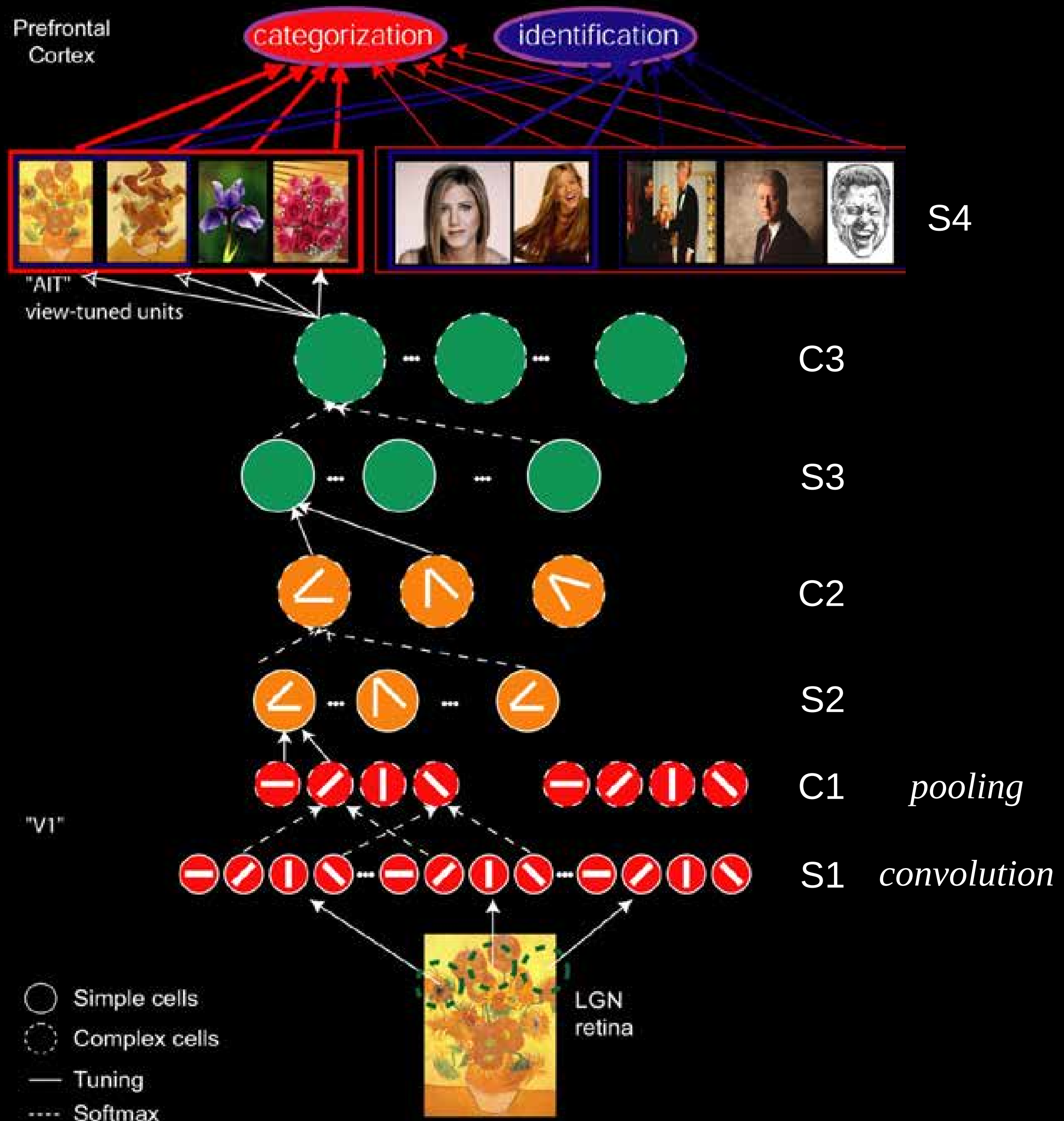
↑
↑
↑
↑

Machine Vision

- “The good, the bad and the ugly” –T. Serre
- Five minute break!

Feedforward models can account for many visual capacities.

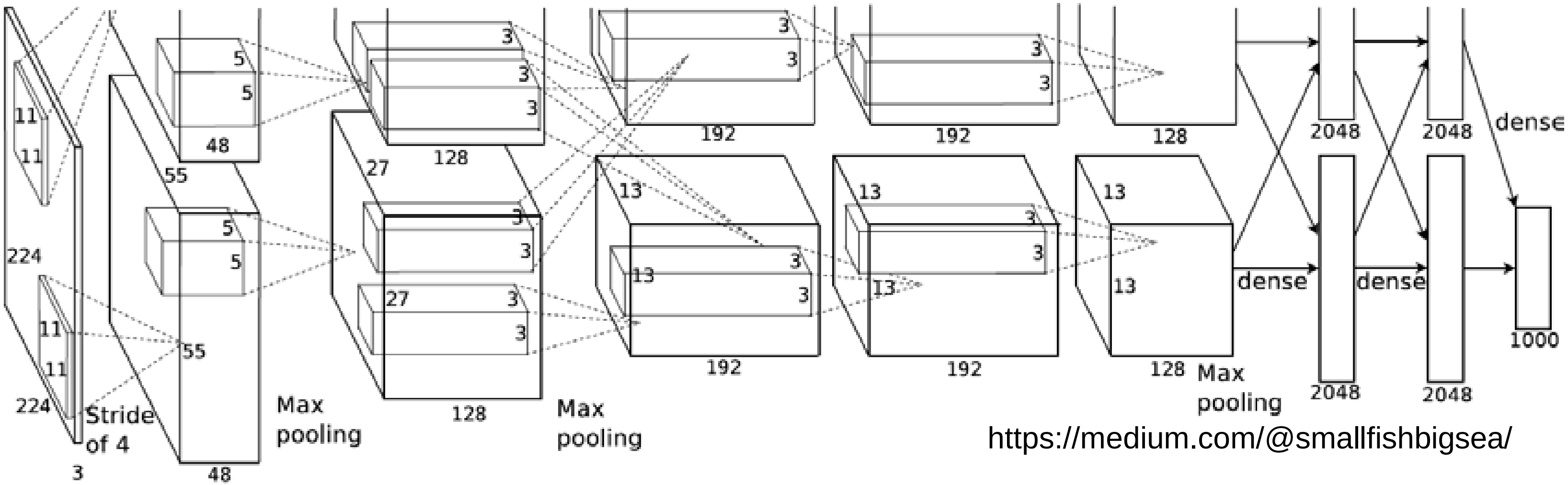
“Cardinal Cells”?
(Barlow 1972)
↑
↑
↑
selectivity
↑
generalization
↑
selectivity
↑
generalization
↑
selectivity



Riesenhuber & Poggio 1999

Convnets

set of functions hierarchically arranged as layers (similar to visual areas)



https://medium.com/@smallfishbigsea/

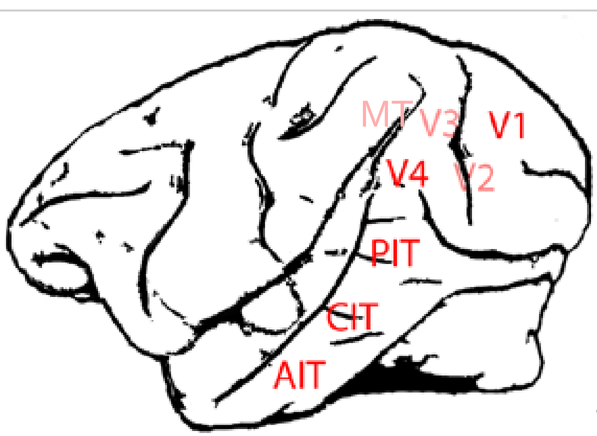
each function is a fairly simple operation (biologically plausible):

1) Convolution

2) Rectification

3) Pooling

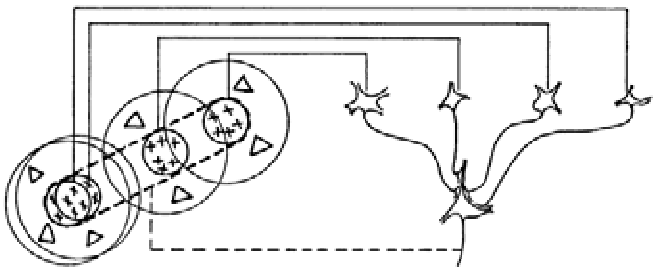
4) Normalization



slide courtesy of Dr. Carlos Ponce

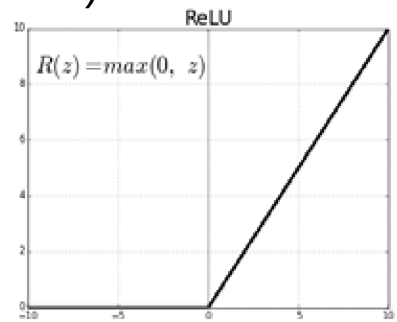
Biologically plausible operations

1) Convolution



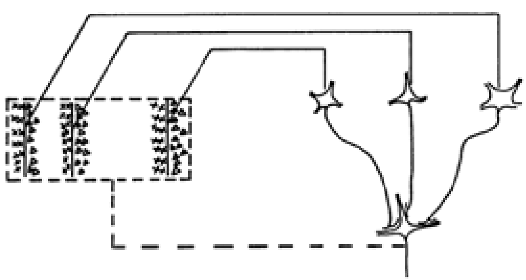
neuronal selectivity as encoded
in synaptic weights

2) Rectification



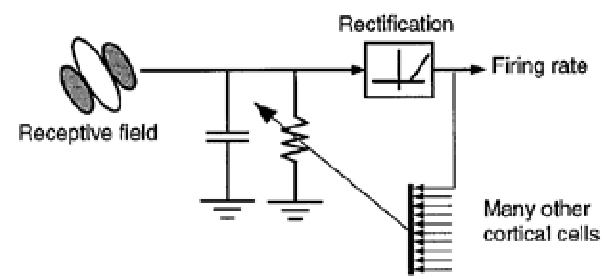
keeps convolution
values positive

3) Pooling



complex cell-like operation, combines
outputs with similar selectivities to
achieve invariance

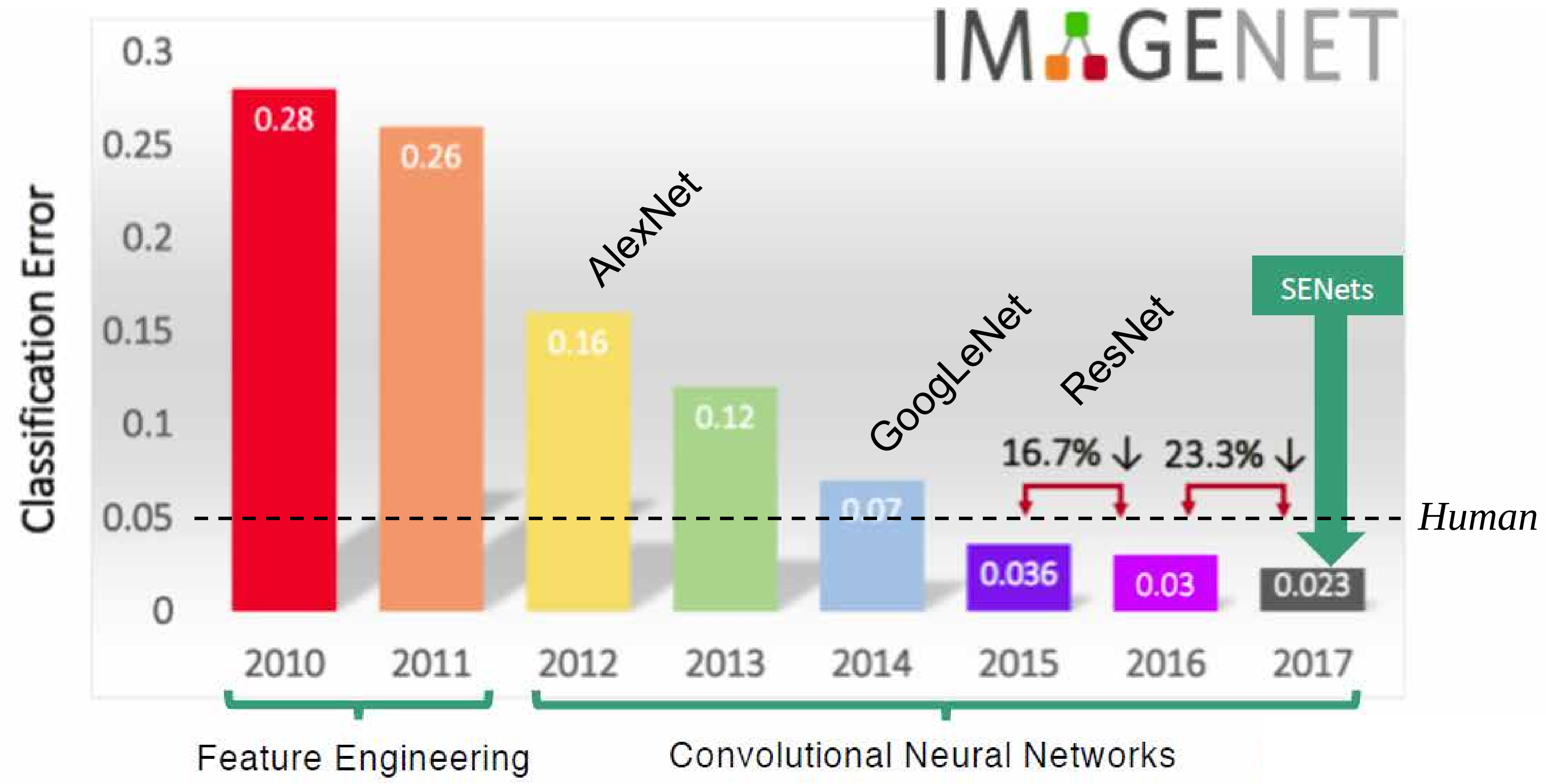
4) Normalization



keeps population activity within a
stable range

slide courtesy of Dr. Carlos Ponce

Convnets crush



Statistics provided by ILSVRC

Convnets: The Good

1. Super-human performance:
   a) ILSVRC (Russakovsky et al. 2015)
   b) face recognition (Phillips et al. 2018)

2. Generalization to other tasks:
   a) *Human scan paths (Bethge lab)

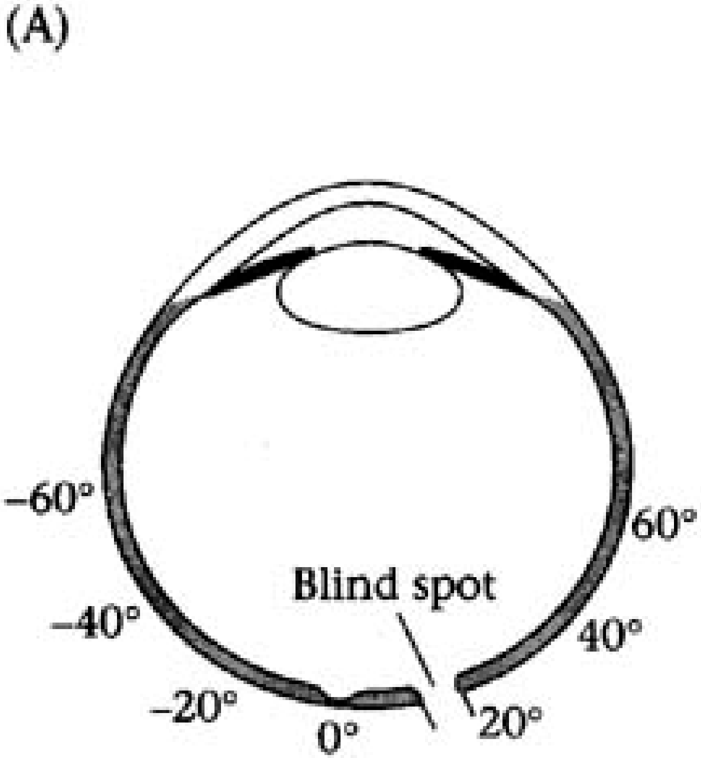
3. In silico electrophysiology:
   a) V1 nonlinearities (Bethge lab)

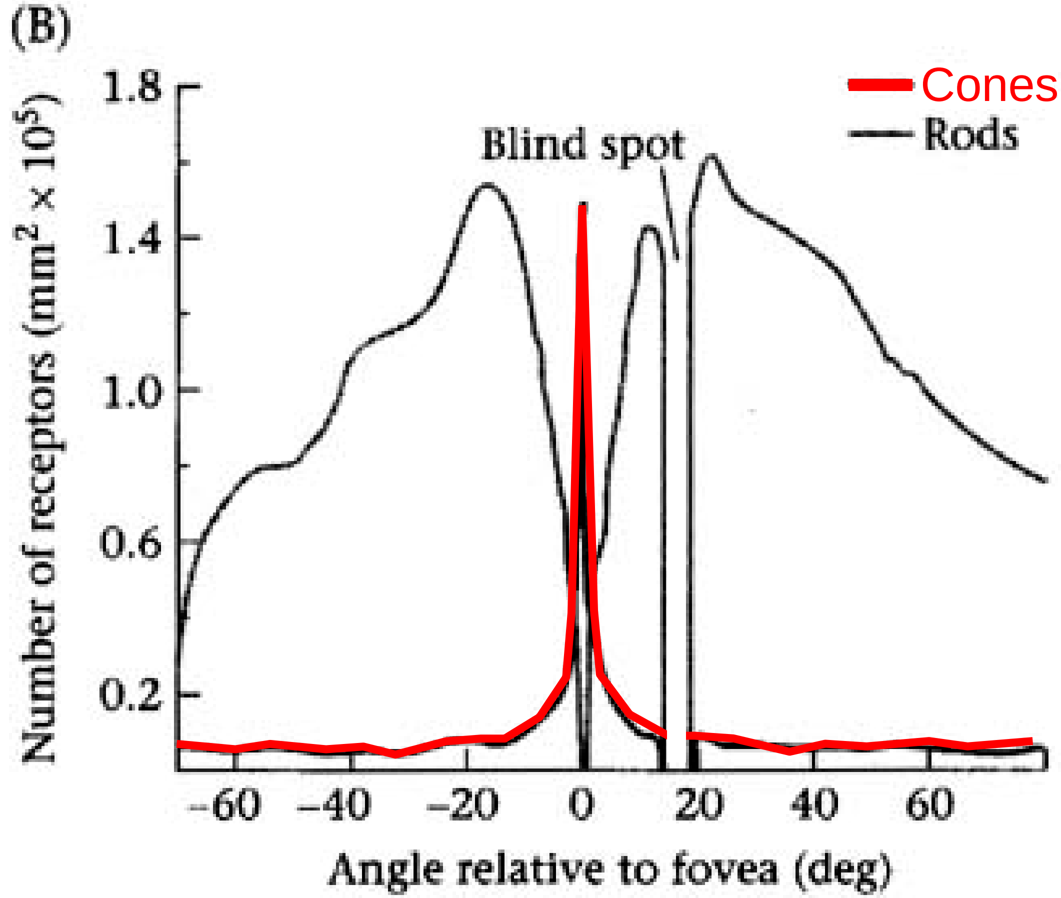
4. Fitting physiology data:
   a) Ventral stream visual areas (Yamins et al. 2014)

5. Tool for exploring high-dimensional feature space:
   a) *Super-natural stimuli (Ponce et al. 2019)

For many more examples, see Serre, T (2019) “Deep learning: The good, the bad and the ugly,” Annu. Rev. Vis. Sci.

Fovea & Acuity





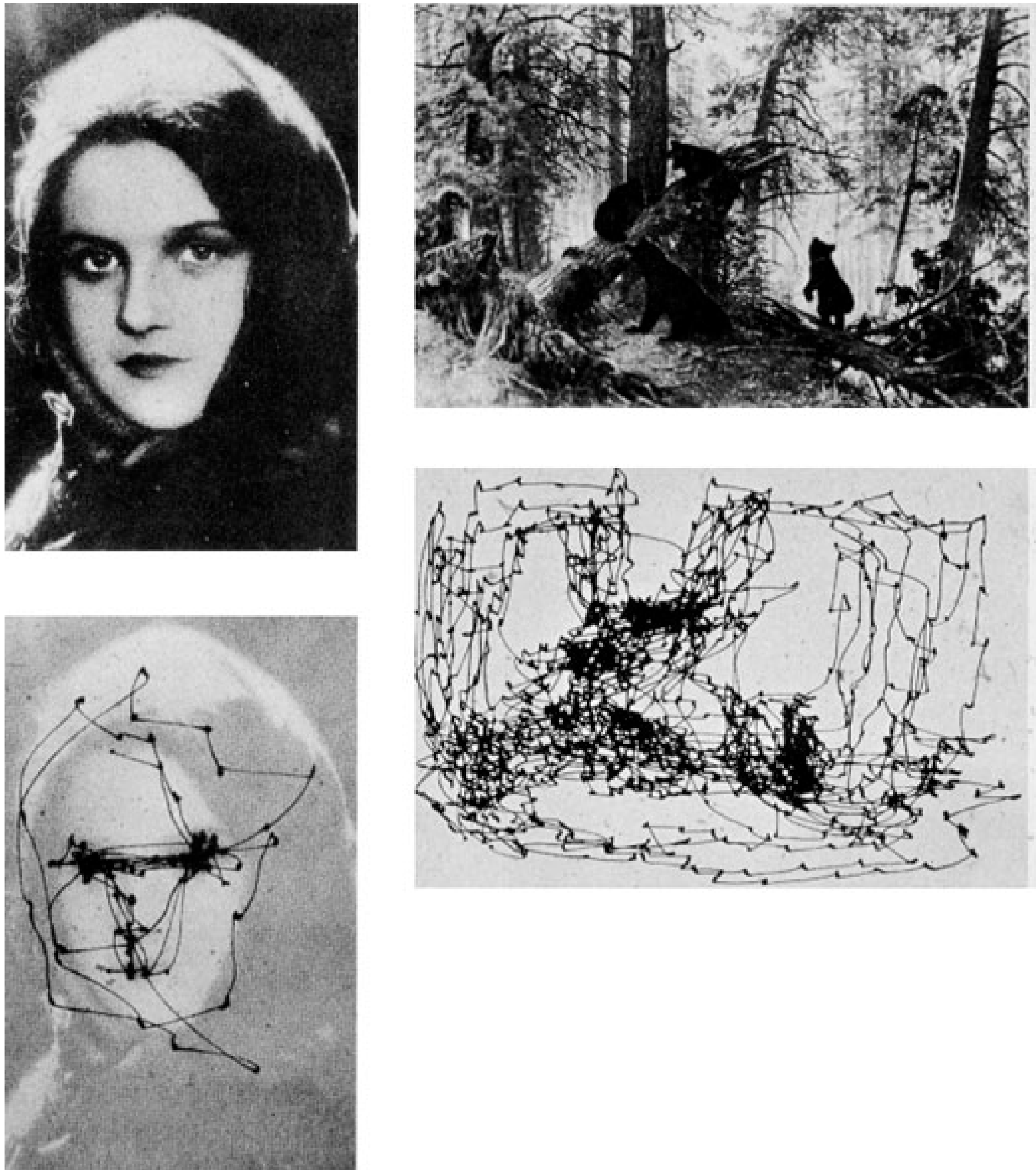
How many letters can you read?

"The muscles were of necessitie provided and given to the eye, that so it might move on every side: for if the eye stoode fast, and immoveable, we should be constrained to turne our head and necke (being all of one peece) for to see: but by these muscles it now moveth it selfe with such swiftnes and nimblenes, without stirring of the head, as is almost incredible …"

Andreas Laurentius (1599)

"A Discourse of the Preservation of Sight: of Melancholike Diseases; of Rheumes, and of Old Age"

Predicting scan paths



Yarbus 1957

Accepted as a workshop contribution at ICLR 2015

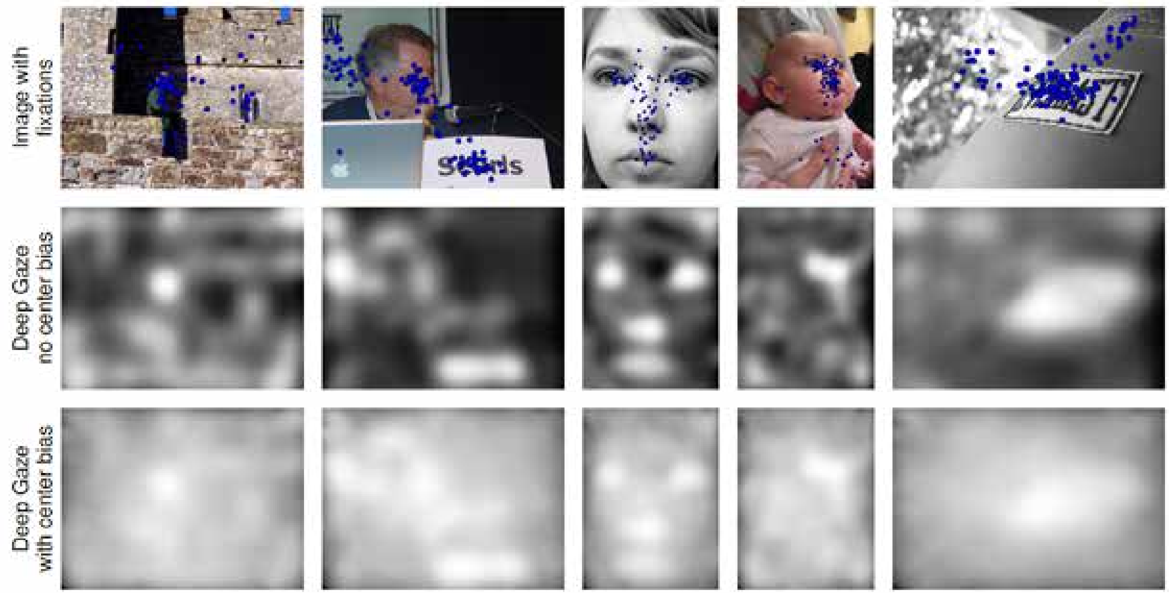
DEEP GAZE I: BOOSTING SALIENCY PREDICTION WITH FEATURE MAPS TRAINED ON IMAGENET

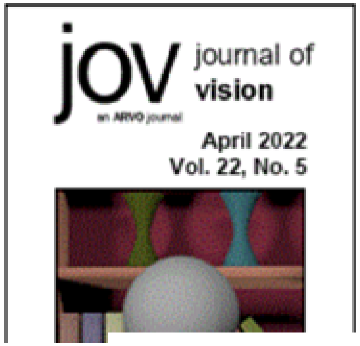
Matthias Kümmerer, Lucas Theis & Matthias Bethge
Werner Reichardt Centre for Integrative Neuroscience
University Tübingen, Germany
{matthias.kuemmerer,lucas,matthias}@bethgelab.org

Recent results suggest that state-of-the-art saliency models perform far from optimal in predicting fixations. This lack in performance has been attributed to an inability to model the influence of high-level image features such as objects. Recent seminal advances in applying deep neural networks to tasks like object recognition suggests that they are able to capture this kind of structure. However, the enormous amount of training data necessary to train these networks makes them difficult to apply directly to saliency prediction. <u>We present a novel way of reusing existing neural networks that have been pretrained on the task of object recognition in models of fixation prediction. Using the well-known network of Krizhevsky et al. (2012), we come up with a new saliency model that significantly outperforms all state-of-the-art models on the MIT Saliency Benchmark.</u> The structure of this network allows new insights in the psychophysics of fixation selection and potentially their neural implementation. To train our network, we build on recent work on the modeling of saliency as point processes.

Kümmerer, Theis & Bethge 2015

Predicting scan paths with AlexNet





OPEN ACCESS

Article | April 2022

DeepGaze III: Modeling free-viewing human scanpaths with deep learning

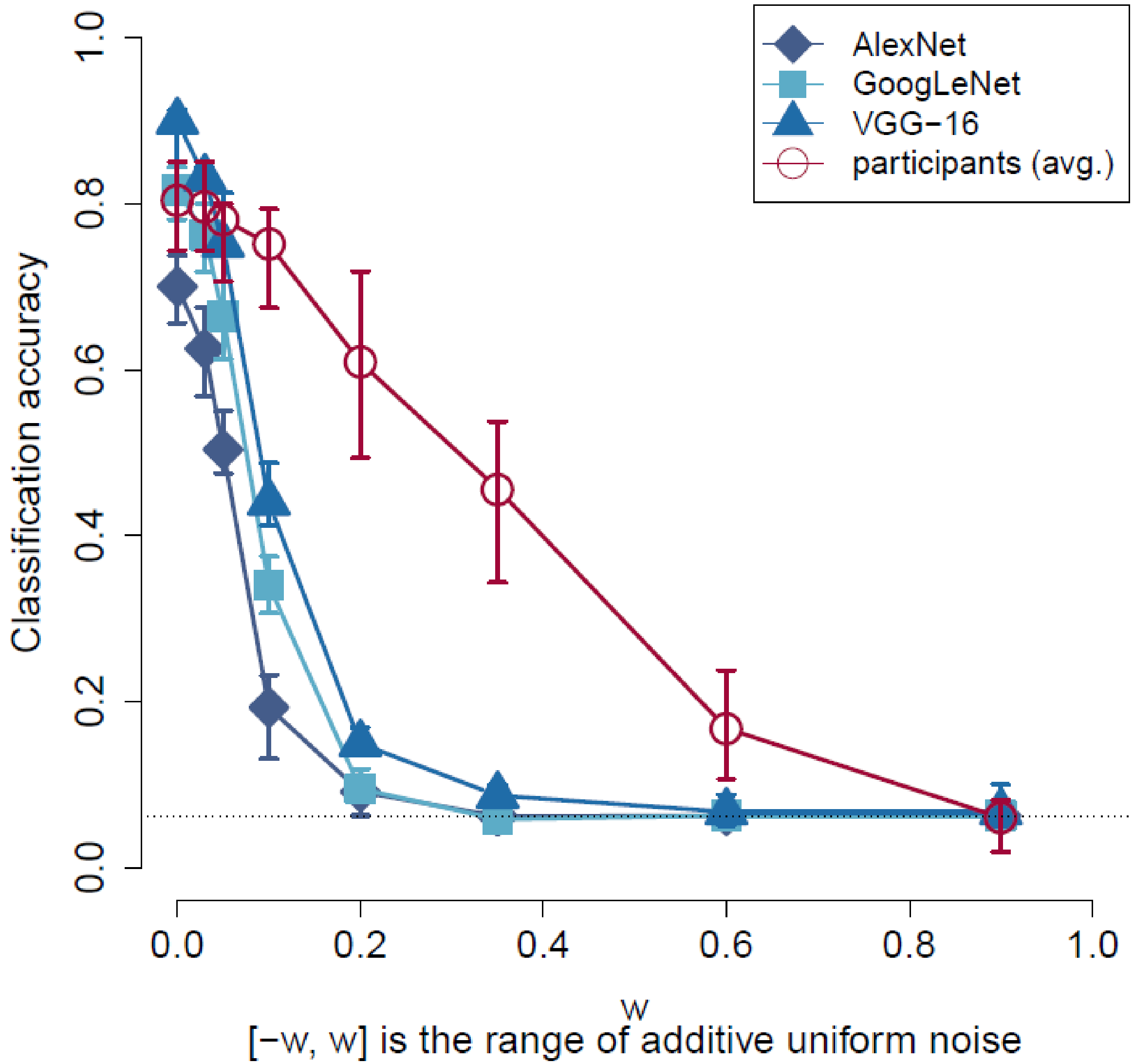
Matthias Kümmerer; Matthias Bethge; Thomas S. A. Wallis

We make the code and trained model parameters for DeepGaze III publicly available at

https://github.com/matthias-k/DeepGaze

Kümmerer, Theis & Bethge 2015

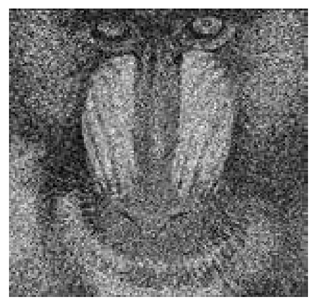
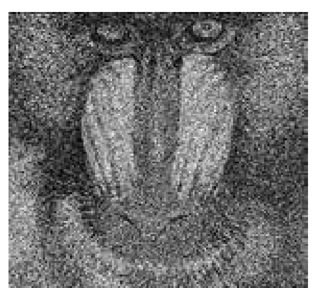
The Bad: Poor noise immunity



Geirhos et al. 2018

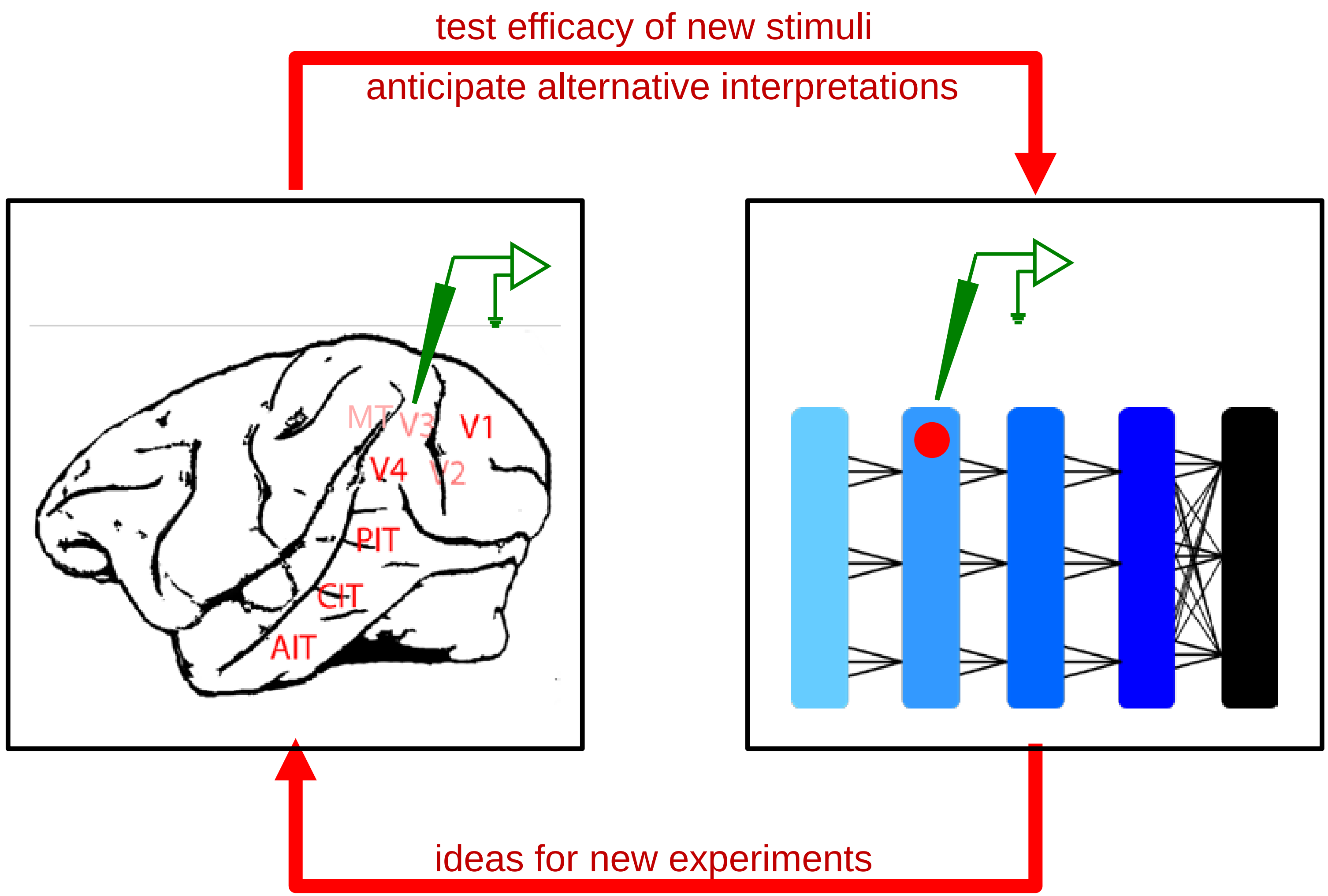
The Ugly: Failure to generalize

Classification performance of **ResNet-50** trained from scratch on (potentially distorted) ImageNet images.

Train		Test		Performance
standard			*standard*	*Super-human*
additive uniform noise			*additive uniform noise*	*Super-human*
salt & pepper noise			*additive uniform noise*	***Chance level***

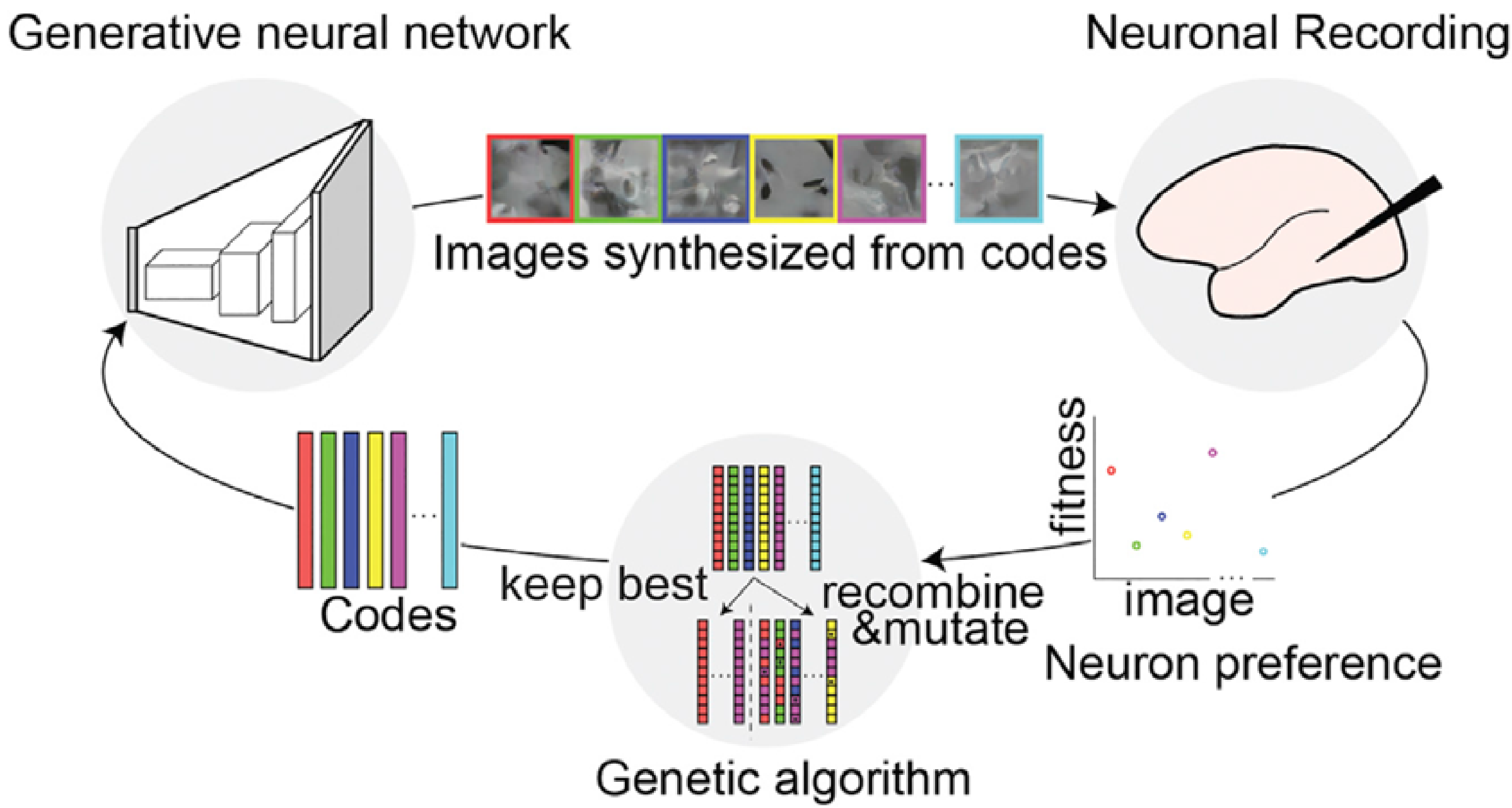
Geirhos et al., NeurIPS 2018

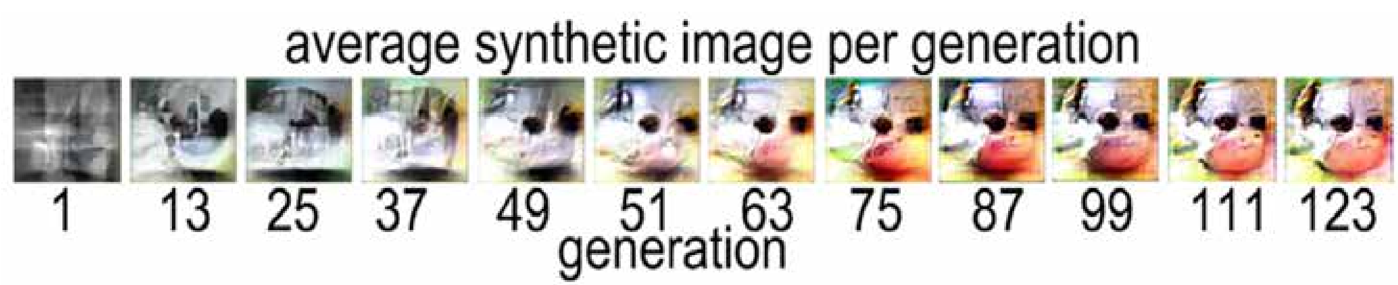
CNNs: predictive models for neuroscience?



slide courtesy of Dr. Carlos Ponce

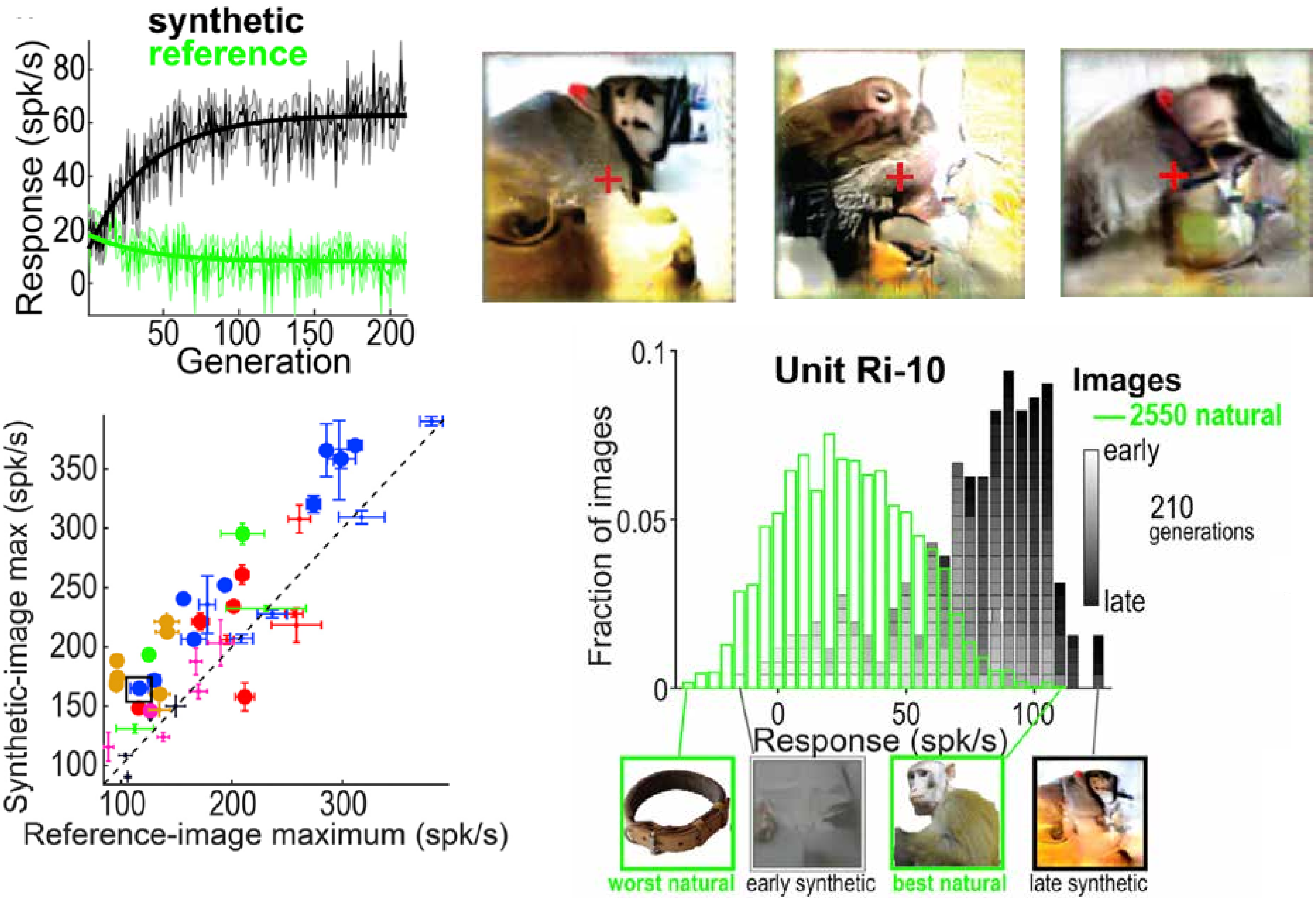
Using convnets to study neurons





Ponce et al. 2019

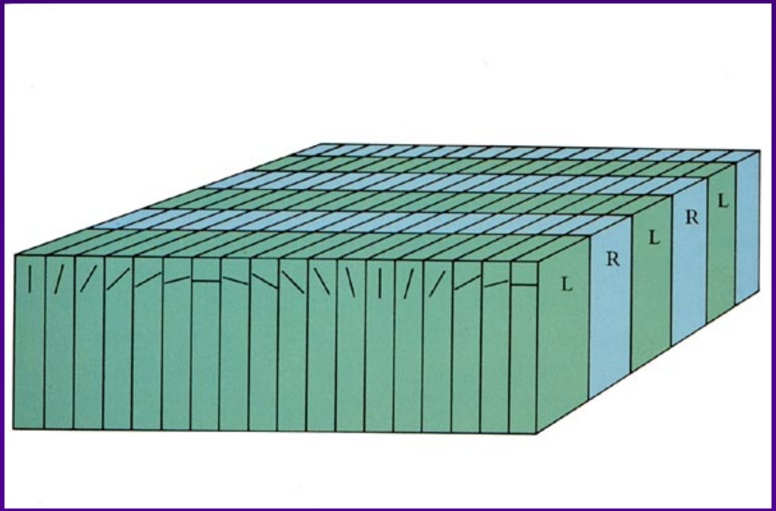
Using convnets to study neurons



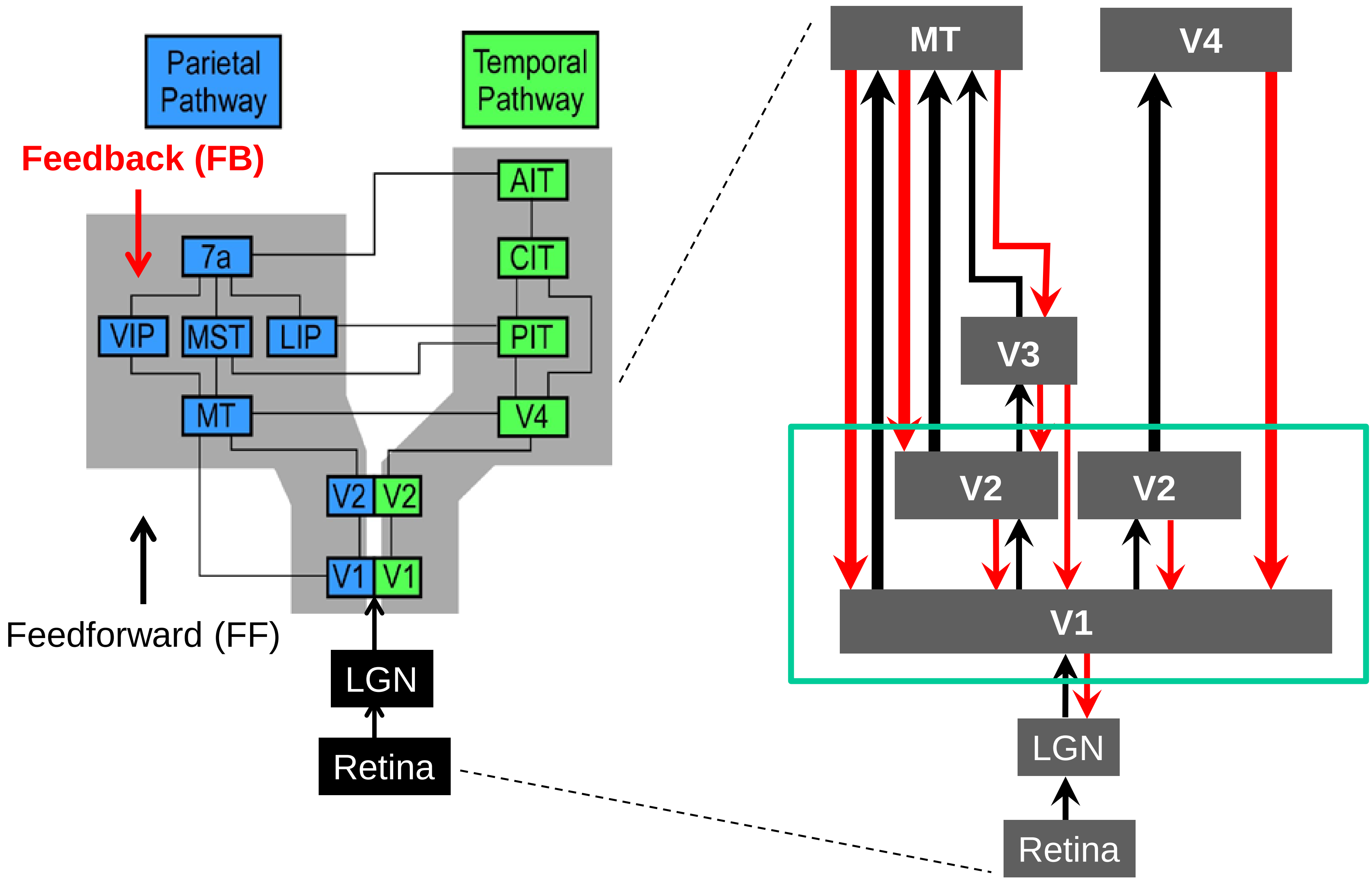
Ponce et al. 2019

Machine Vision

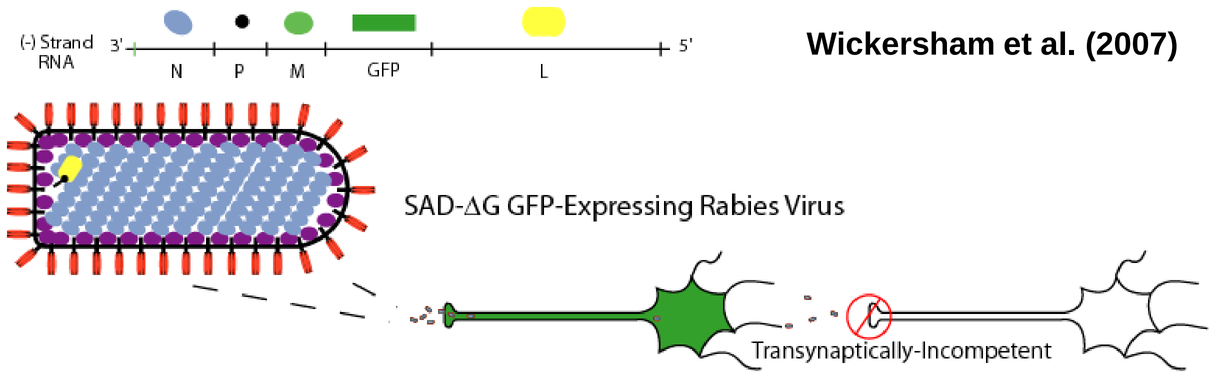
- What might be missing?
  - Feedback?
  - ~~Functional architecture?~~



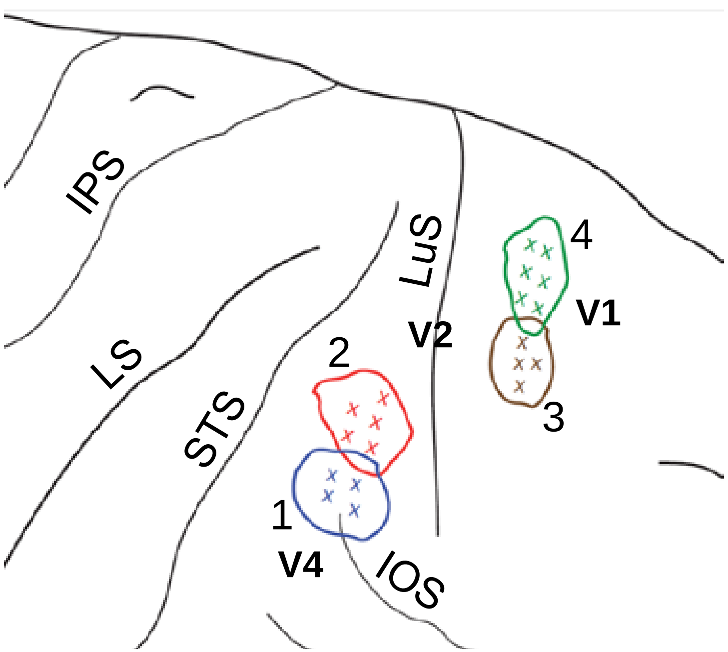
Feedback connections carry information in the opposite direction: top-down

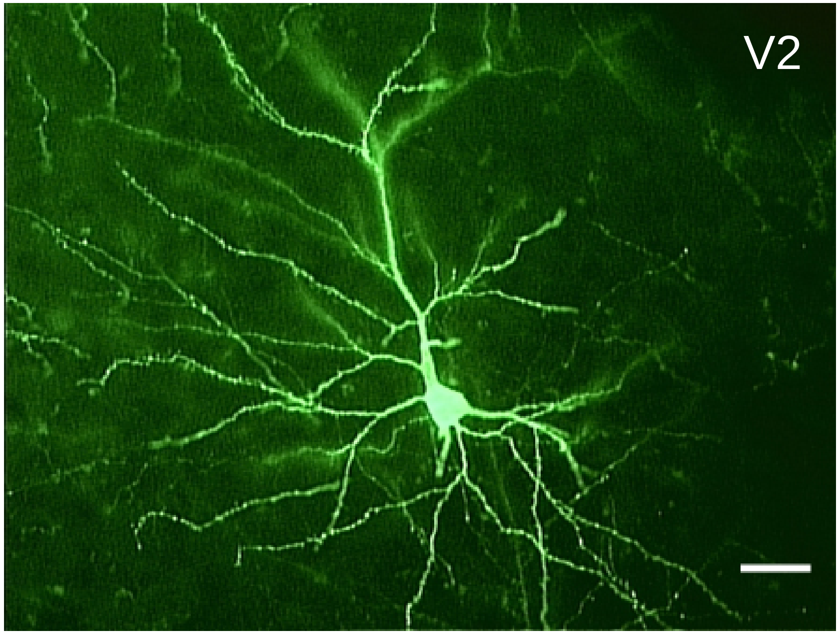


Retrograde tracing with G-deleted rabies virus



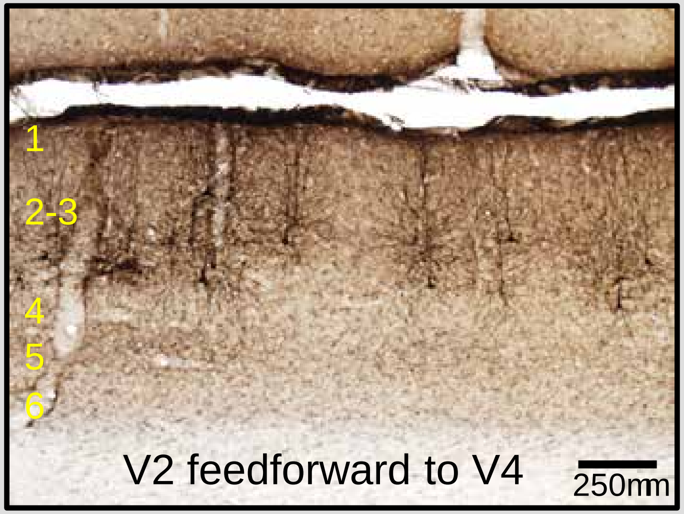
Wickersham et al. (2007)



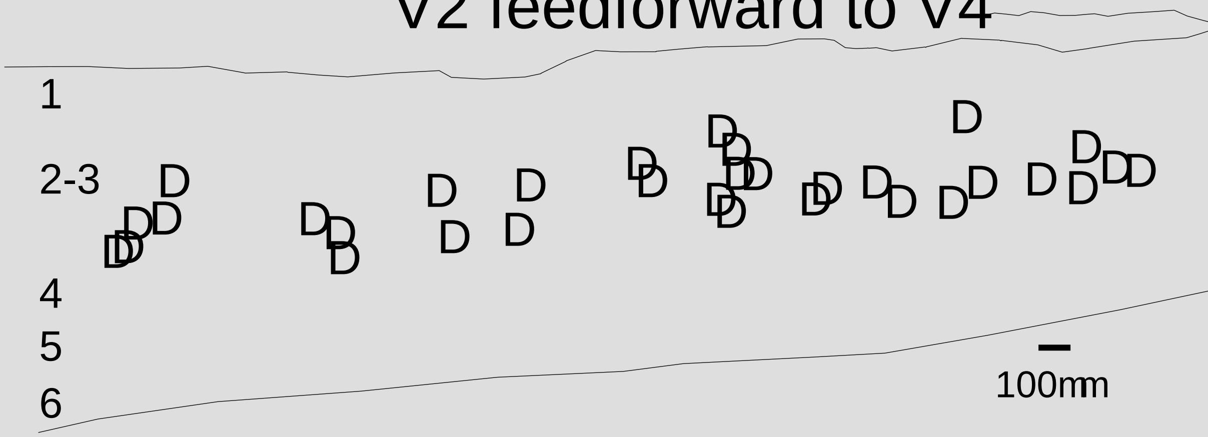


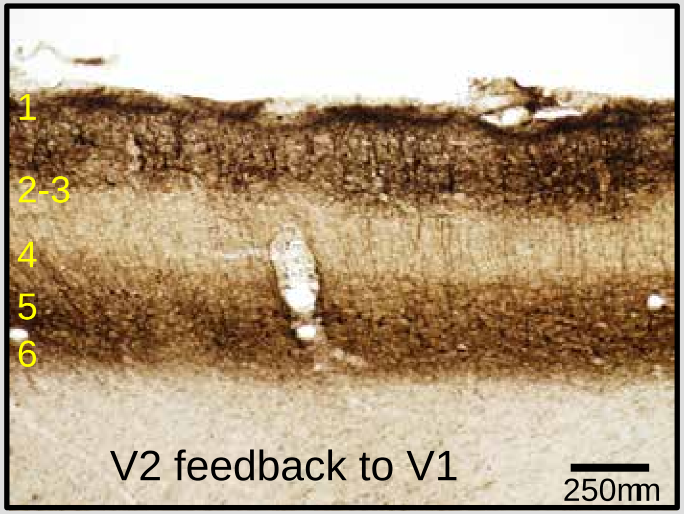
JoJo Nassi

Monkey: FB neurons are numerous and segregated in different layers

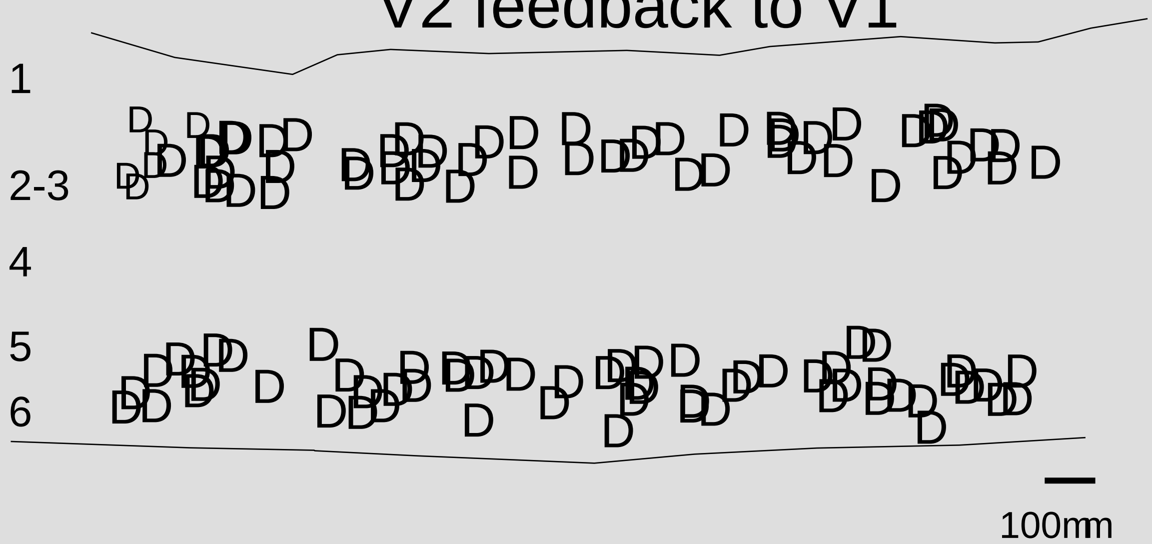


V2 feedforward to V4



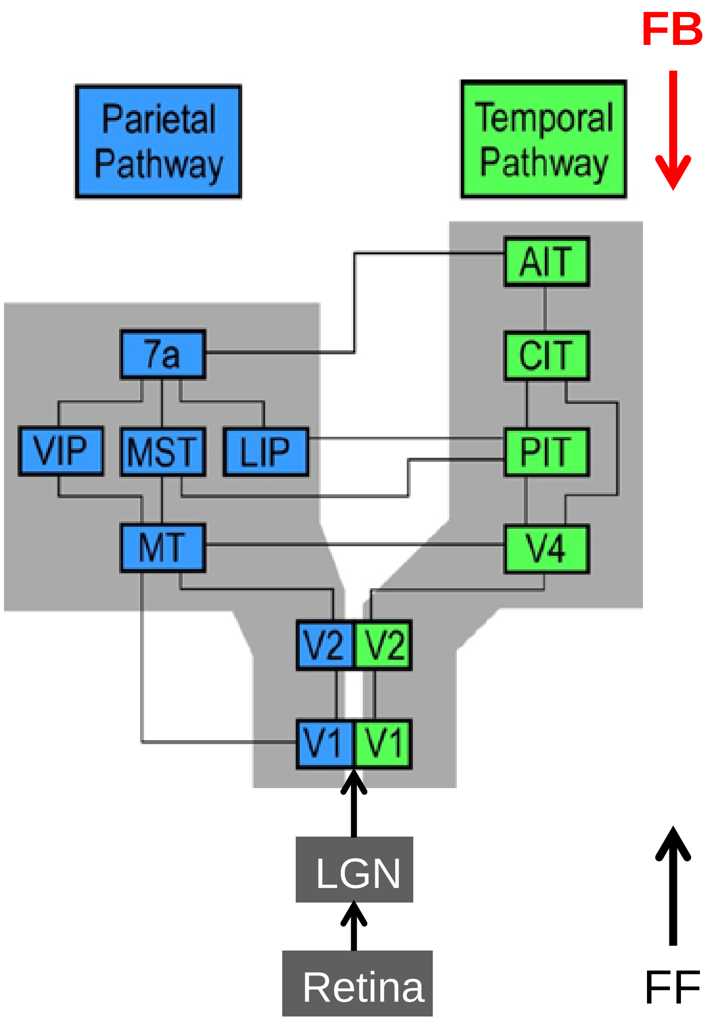


V2 feedback to V1



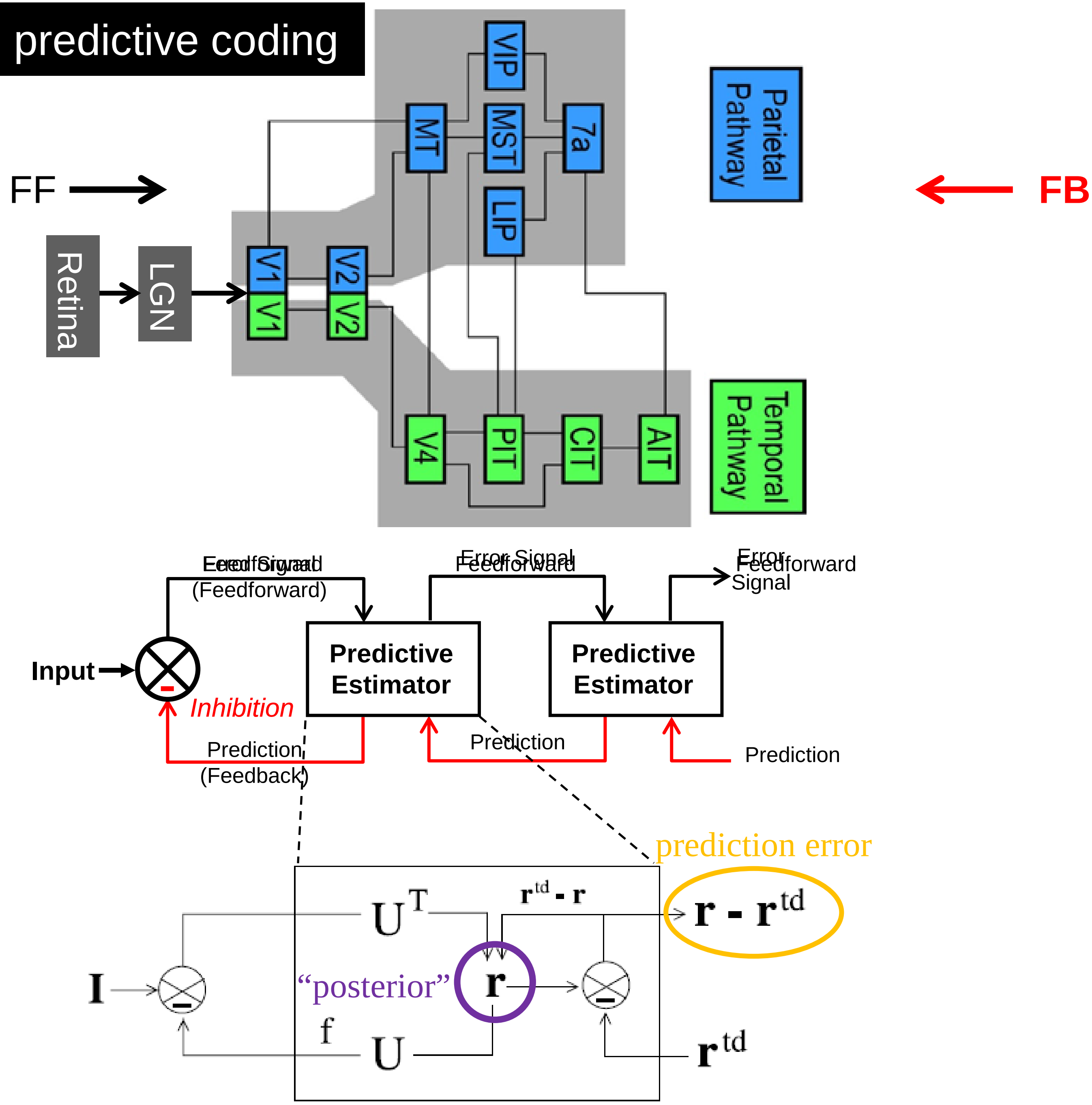
JoJo Nassi & Vladimir Berezovskii

Feedback and predictive coding



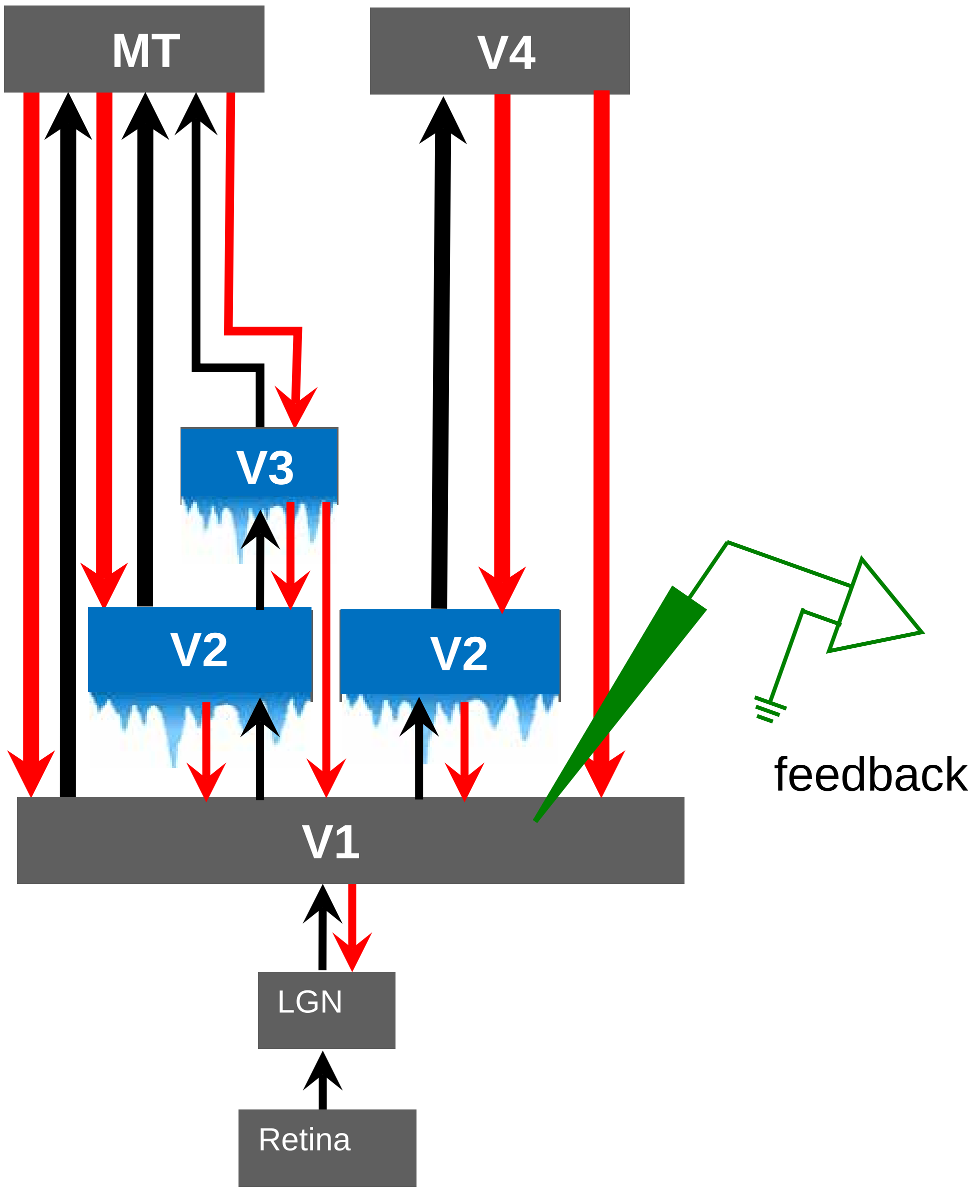
Rao & Ballard 1999

Feedback and predictive coding

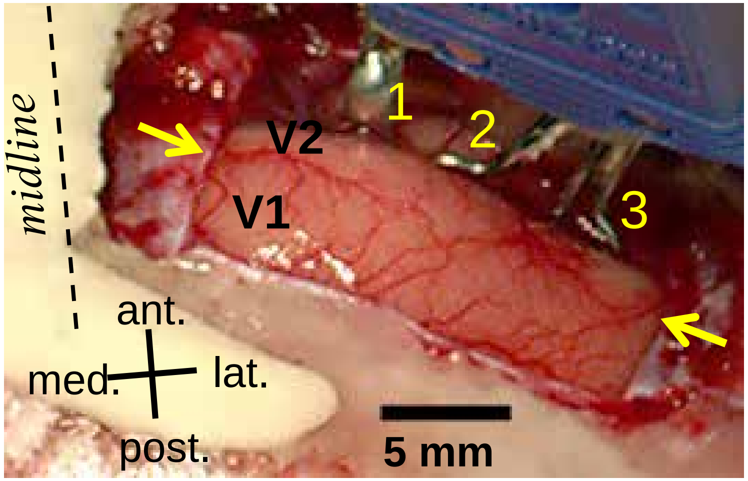


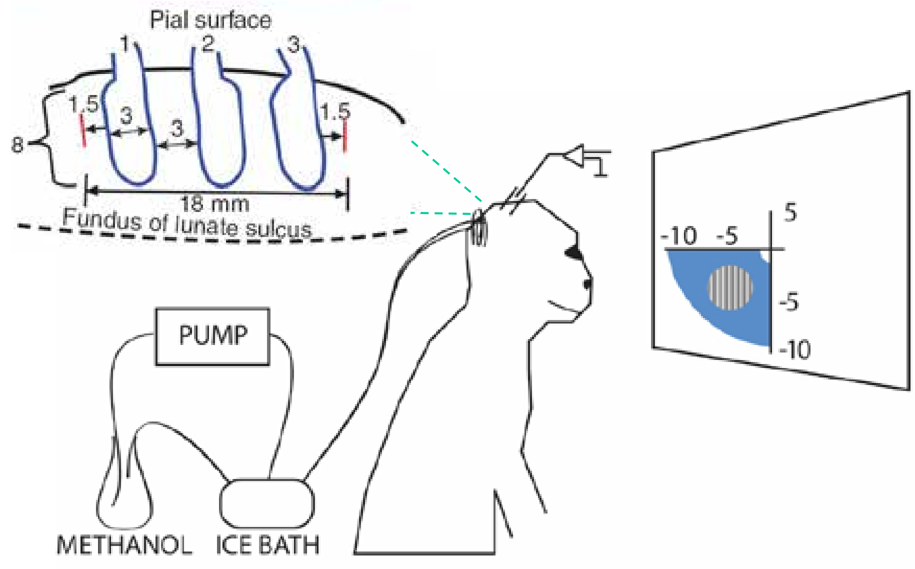
Rao & Ballard 1999

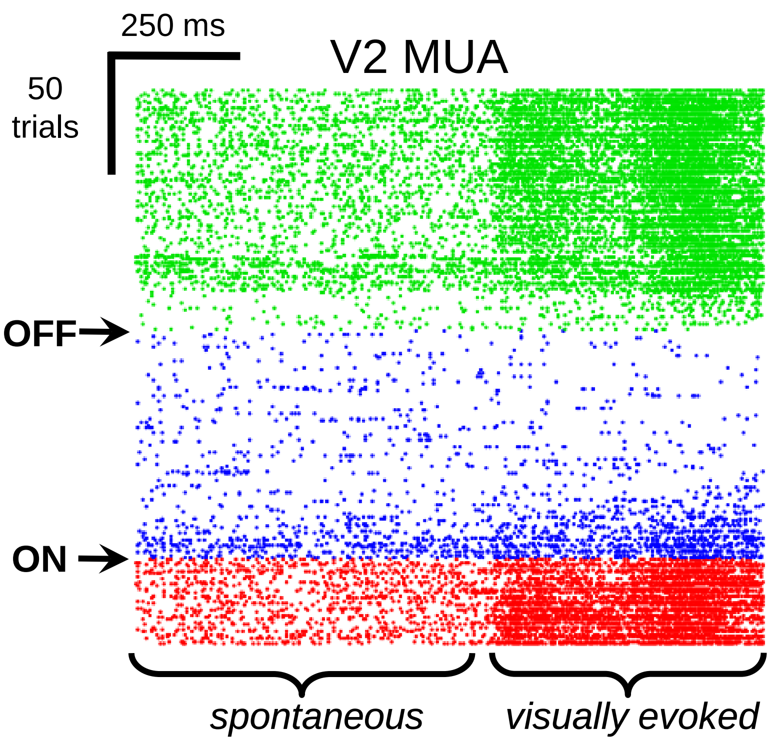
Feedback inactivation

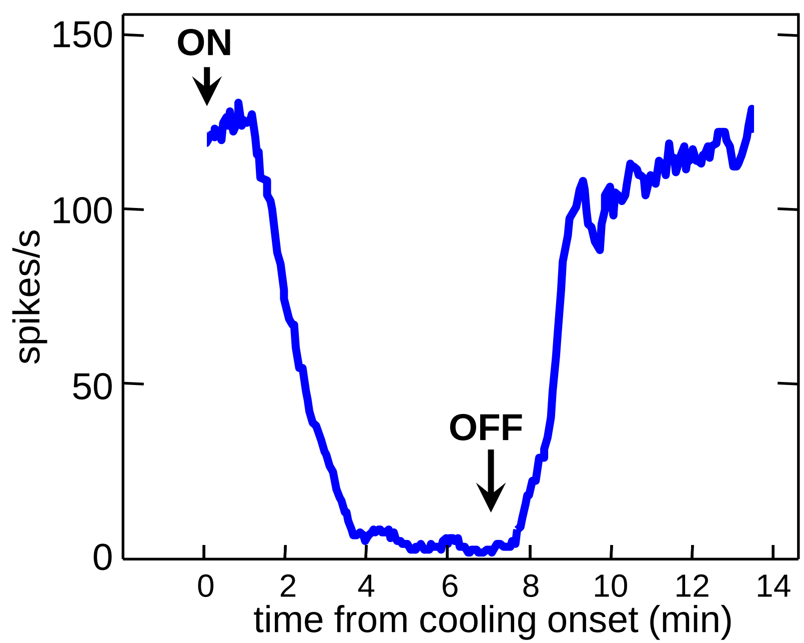


Cooling with cryoloops to reversibly inactivate feedback



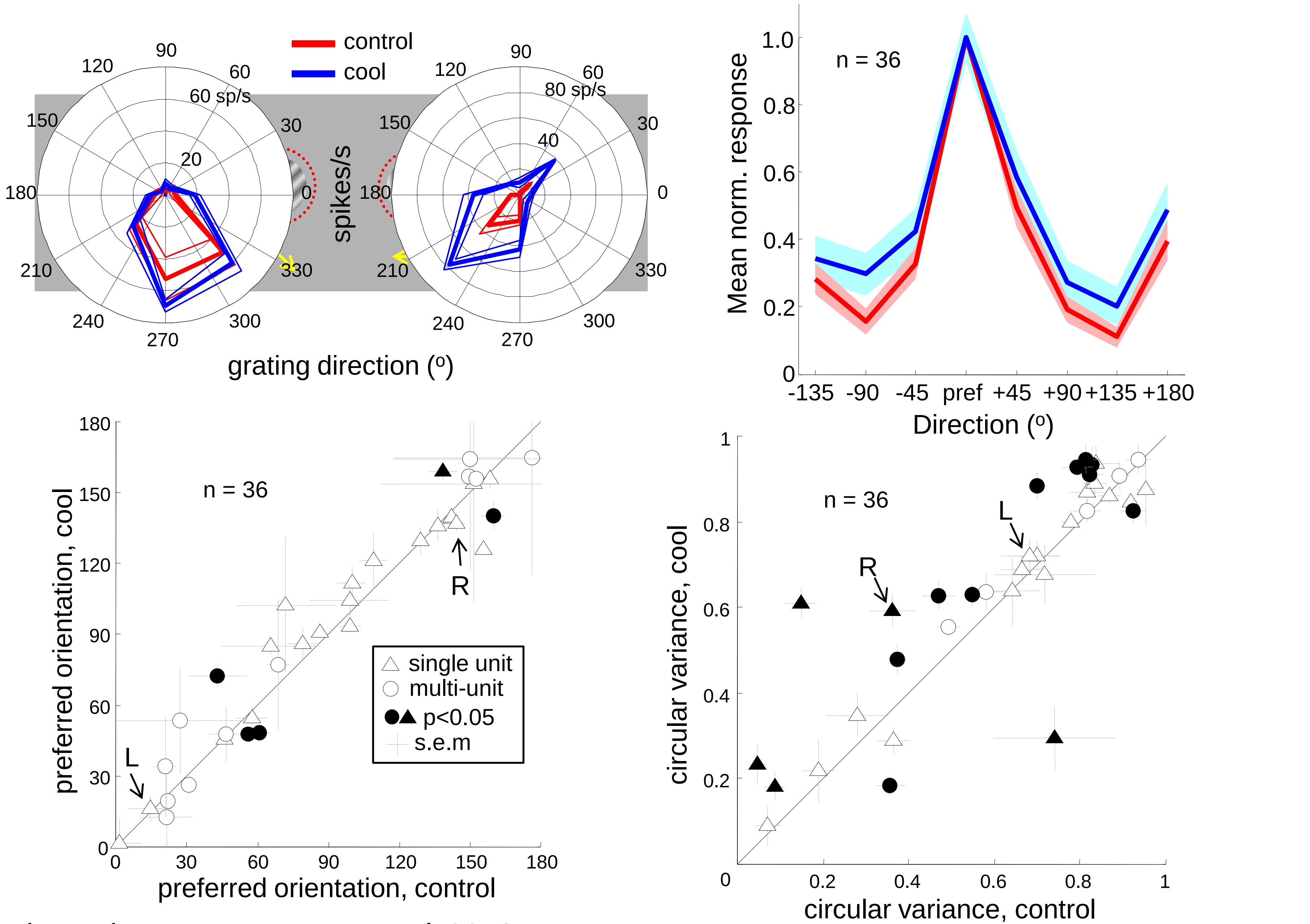






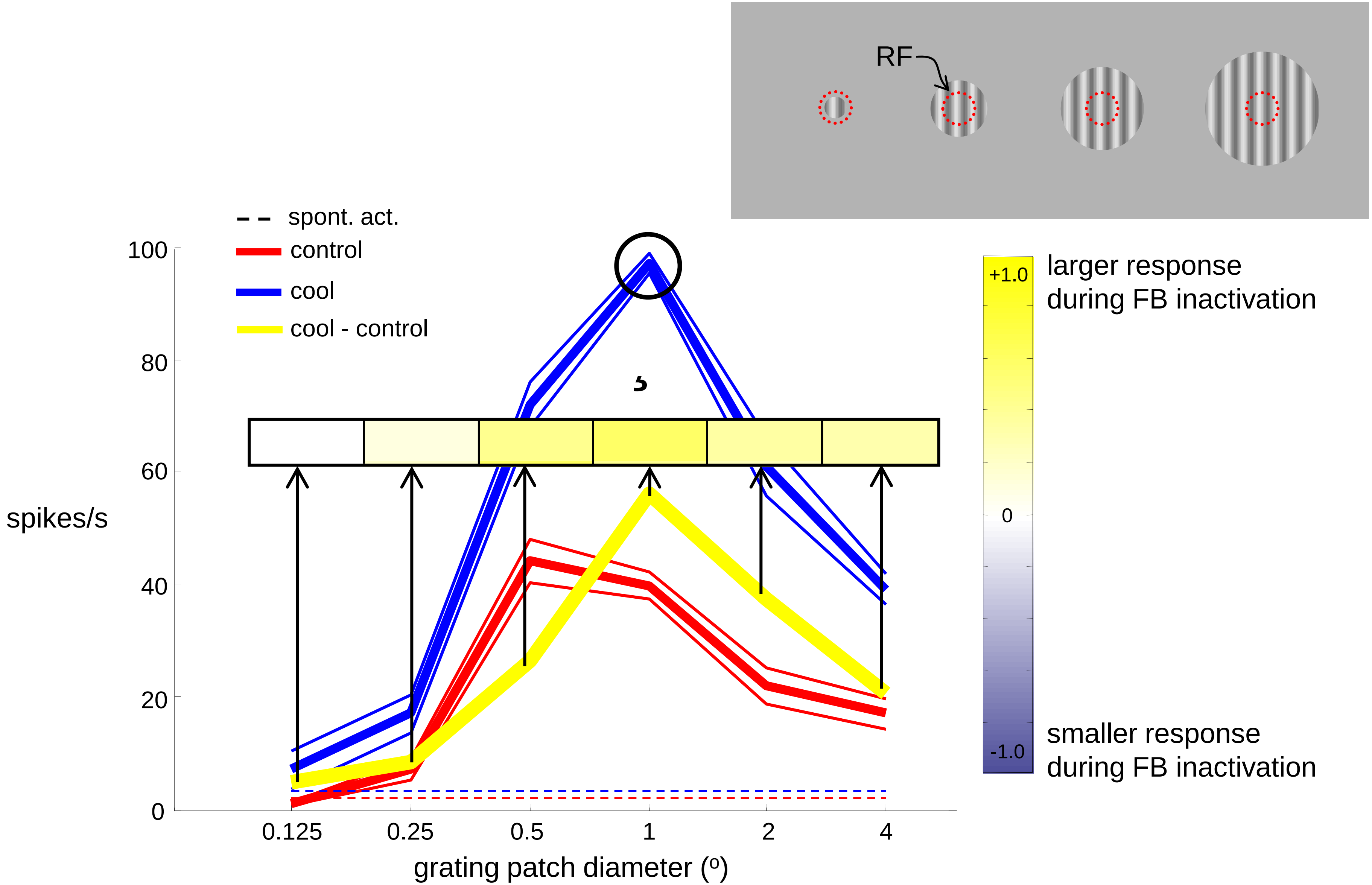
Ponce et al., *Nat. Neurosci.* 2008

Orientation tuning in V1 during FB inactivation: Not much changes.



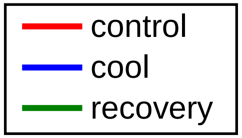
Nassi, Lomber & Born, *J. Neurosci*. 2013

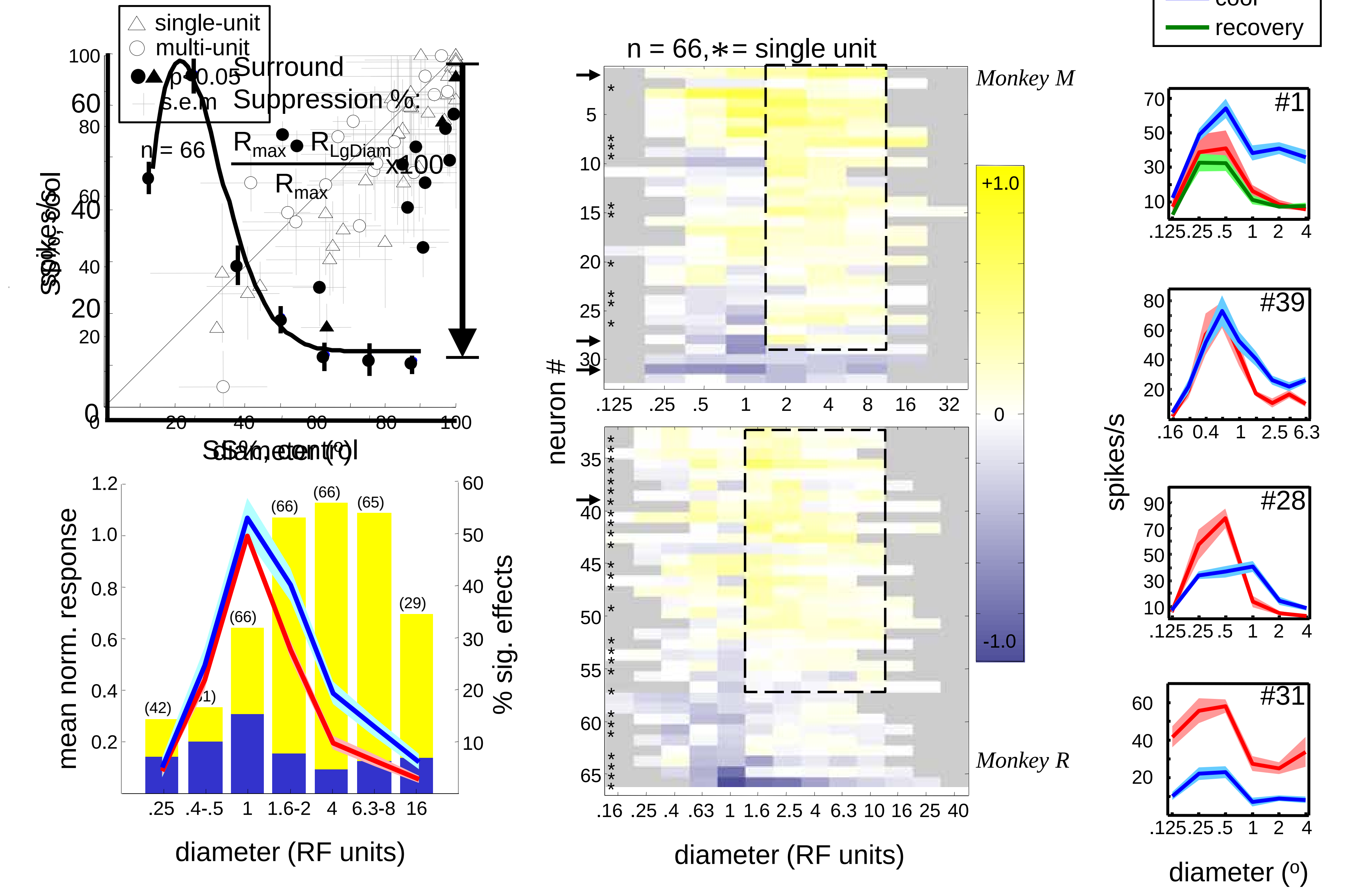
Surround suppression is weaker during inactivation of FB



Nassi, Lomber & Born, *J. Neurosci*. 2013

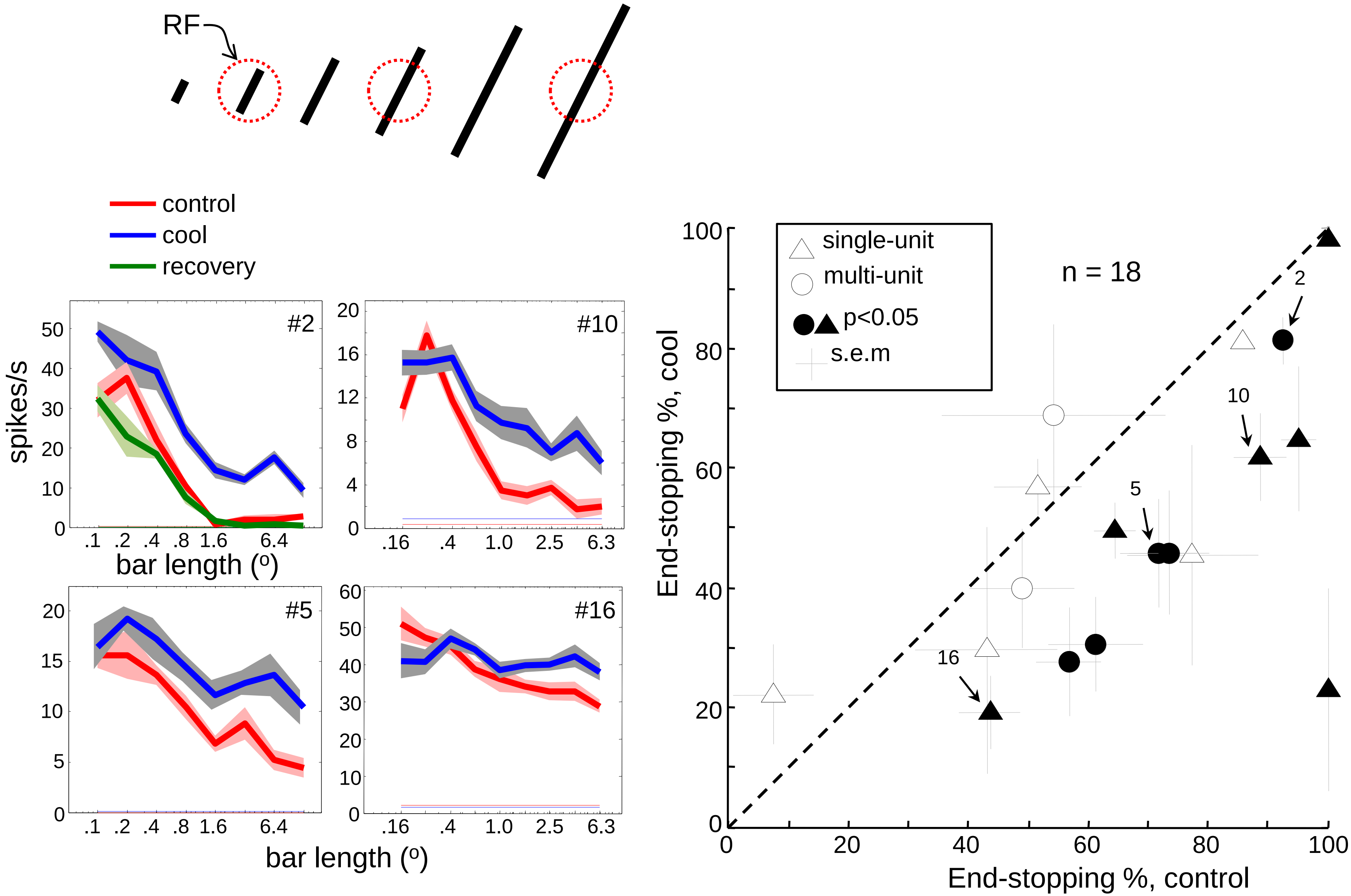
Release of suppression during inactivation of feedback





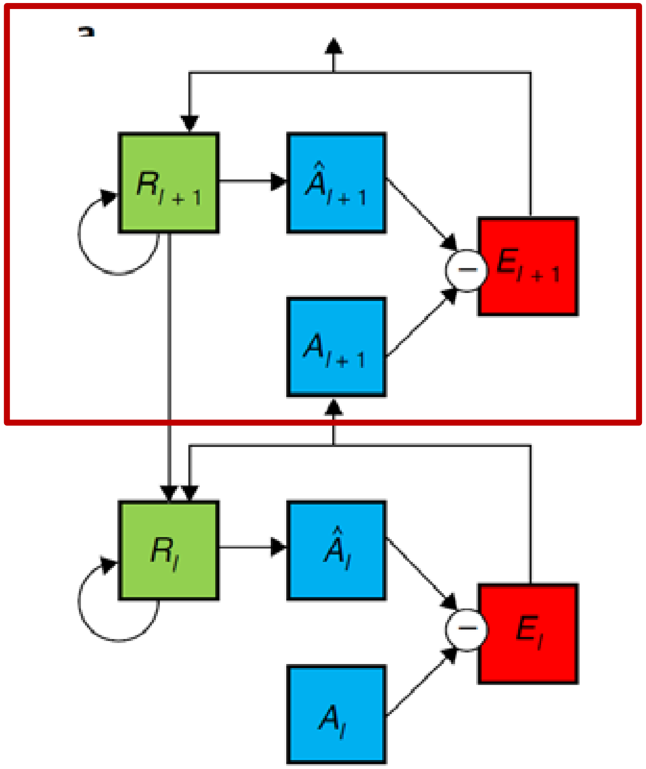
Nassi, Lomber & Born, *J. Neurosci*. 2013

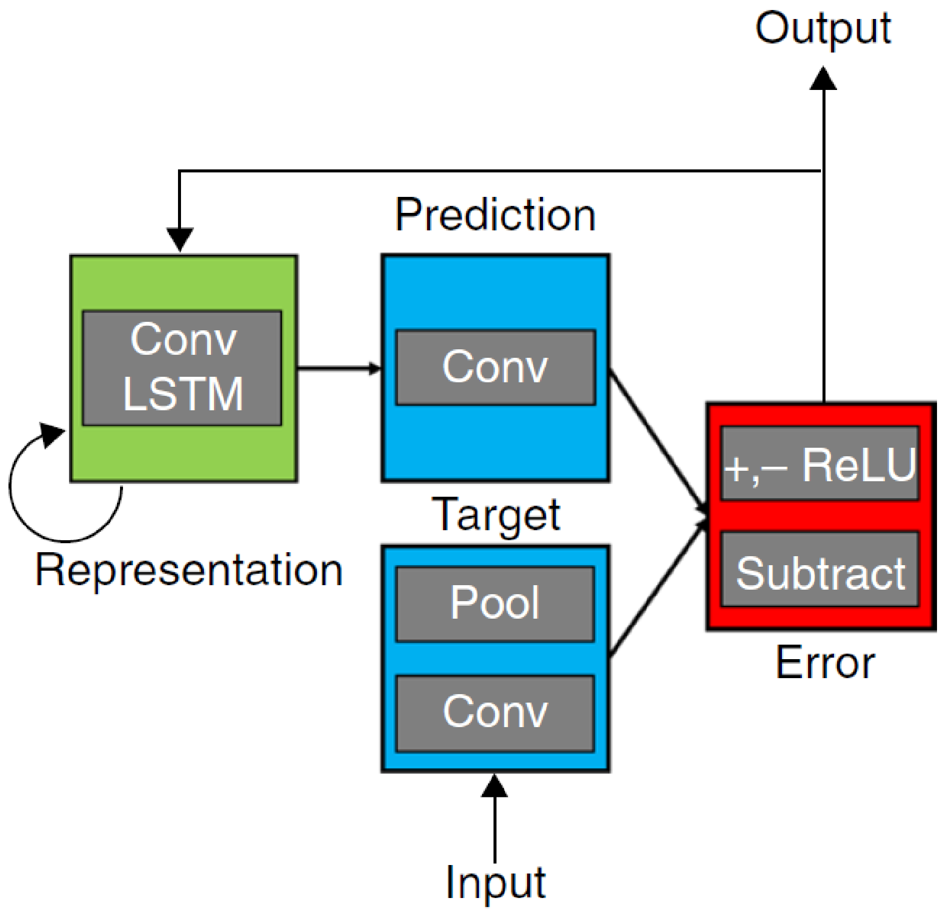
End-stopping is also decreased during inactivation of feedback



Nassi, Lomber & Born, *J. Neurosci*. 2013

Deep neural networks that predict: PredNet

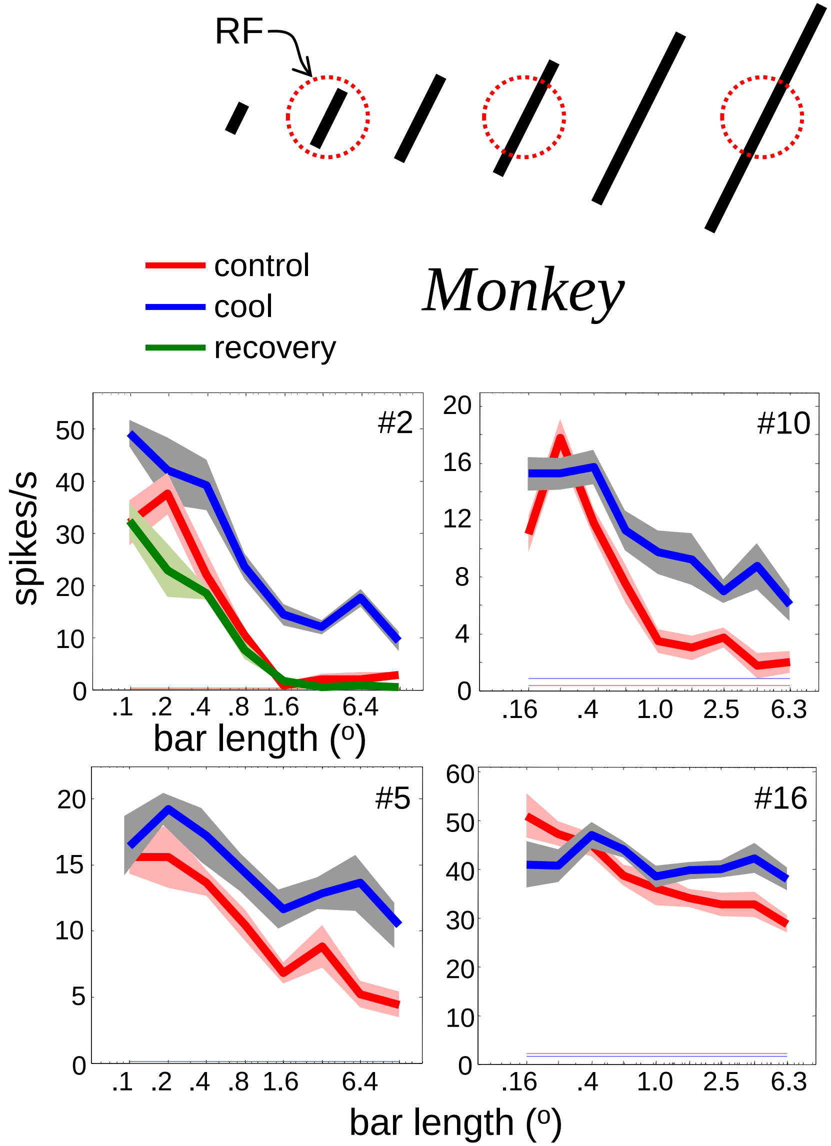


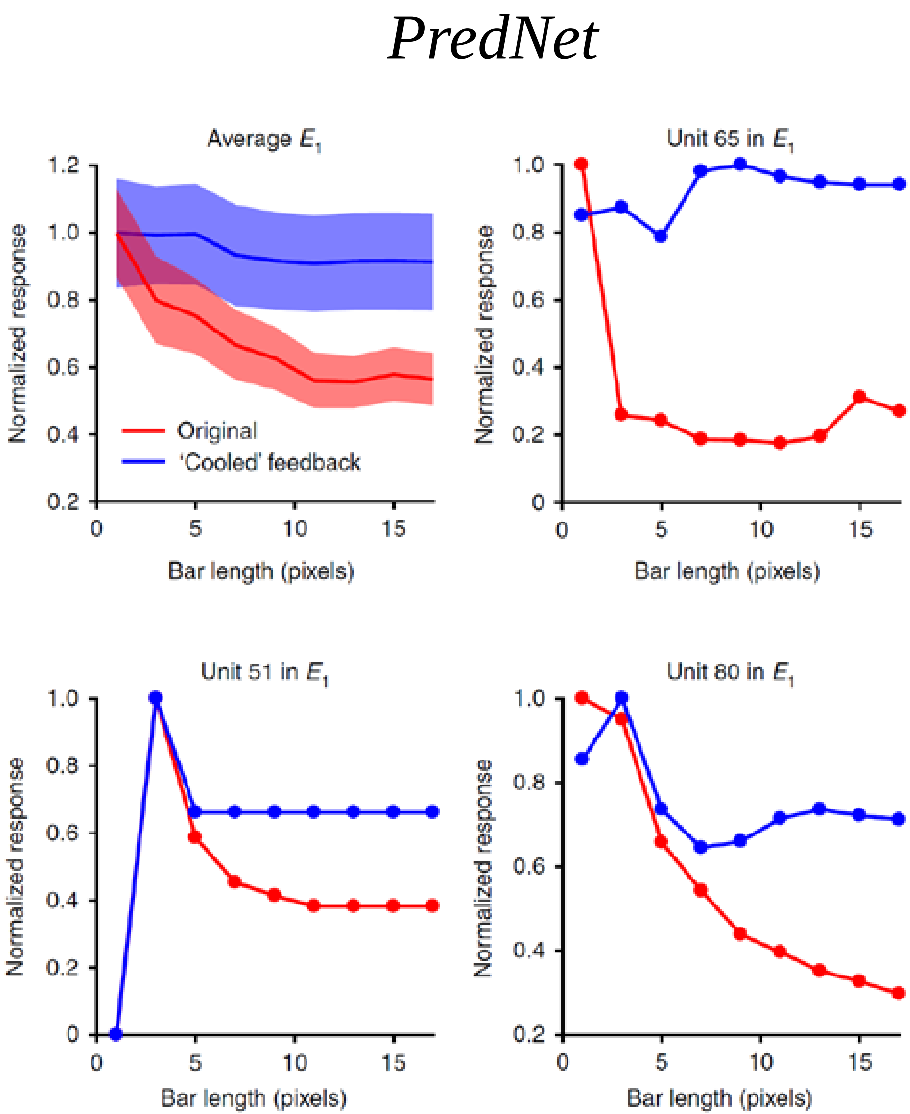


The network is trained to minimize the activations of the error units across the training set using (truncated) backpropagation through time, with the error units at each layer contributing to the total loss. Similar to the original work, the results presented here use a model trained for next-frame prediction on a car-mounted camera dataset (KITTI). Thus, the model is **trained in an unsupervised or 'self '-supervised manner** that does not require any external labels or other forms of supervision.

Lotter, Kreiman & Cox, *Nat. Mach. Intell.* 2020

Deep neural networks that predict: PredNet



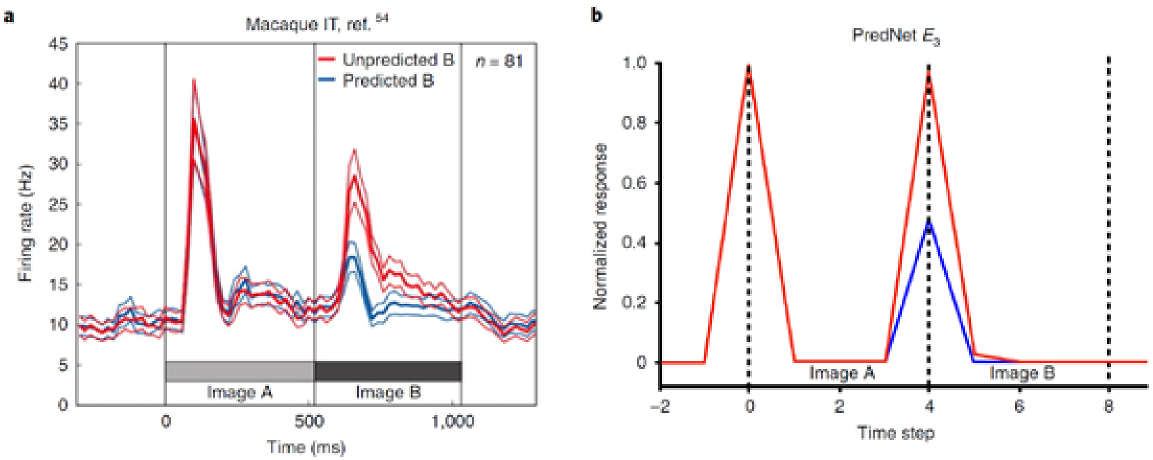


Nassi, Lomber & Born, *J. Neurosci.* 2013

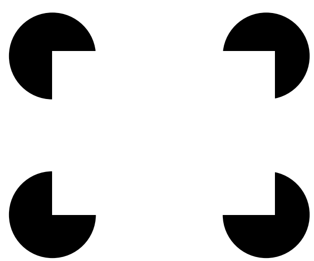
Lotter, Kreiman & Cox, *Nat. Mach. Intell.* 2020

PredNet has other interesting functional properties

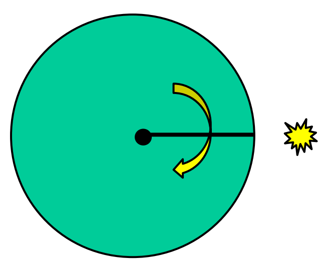
1. IT-like responses of high levels to learned image sequences:



2. V1/V2-like responses of low levels to illusory contours

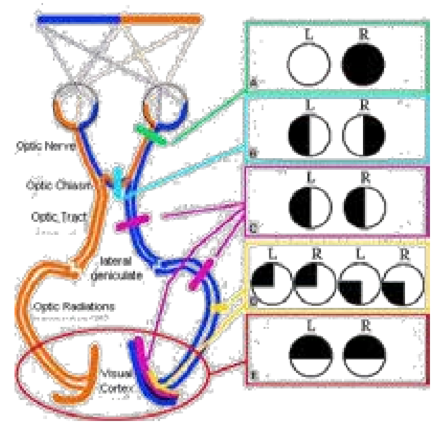


3. Flash-lag effect

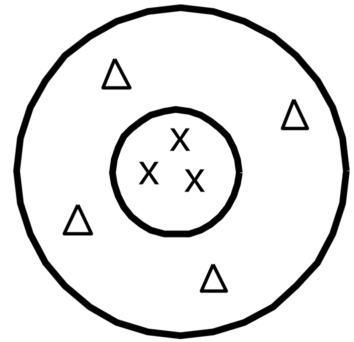


Lotter, Kreiman & Cox, *Nat. Mach. Intell.* 2020

Roadmap1

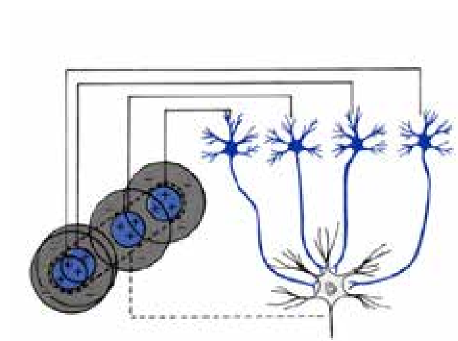


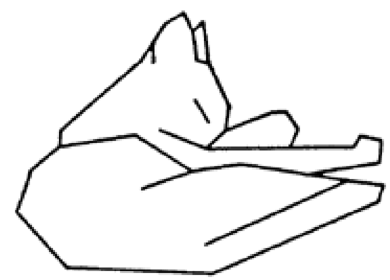
- anatomy of the **visual pathways**



- concept of a **receptive field**

- hierarchy: more complex receptive fields are constructed from simpler ones

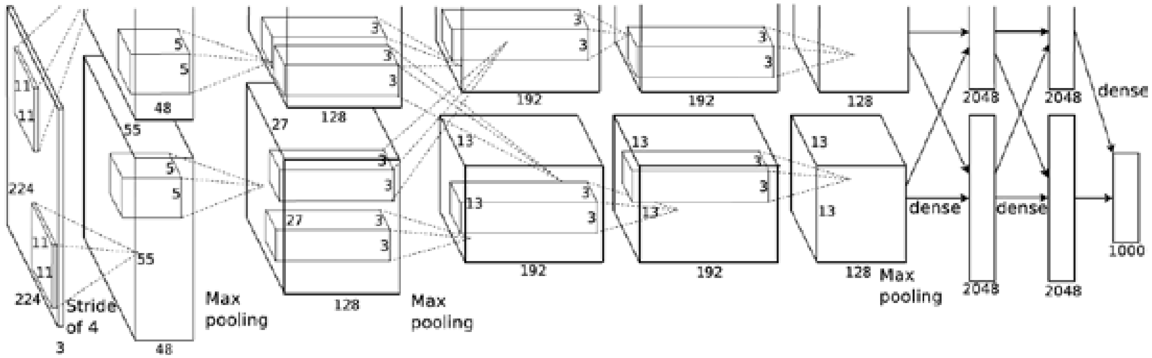




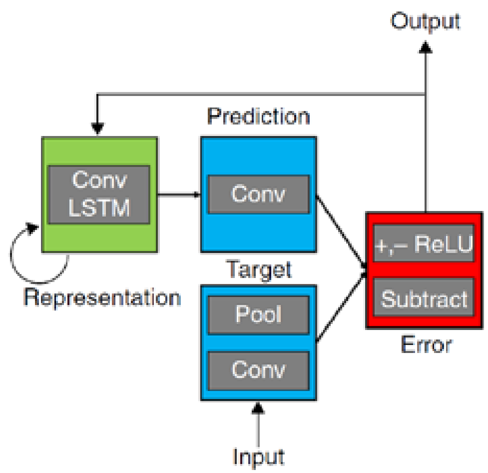
Questions or comments: richard_born@hms.harvard.edu

Roadmap2

• strengths and weaknesses of **CNNs** as a tool for vision science



• what might be missing from current machine vision algorithms?



Questions or comments: richard_born@hms.harvard.edu