Deep networks:
The good, the bad, and the ugly

The future:
Neuroscience-inspired deep nets

Thomas Serre

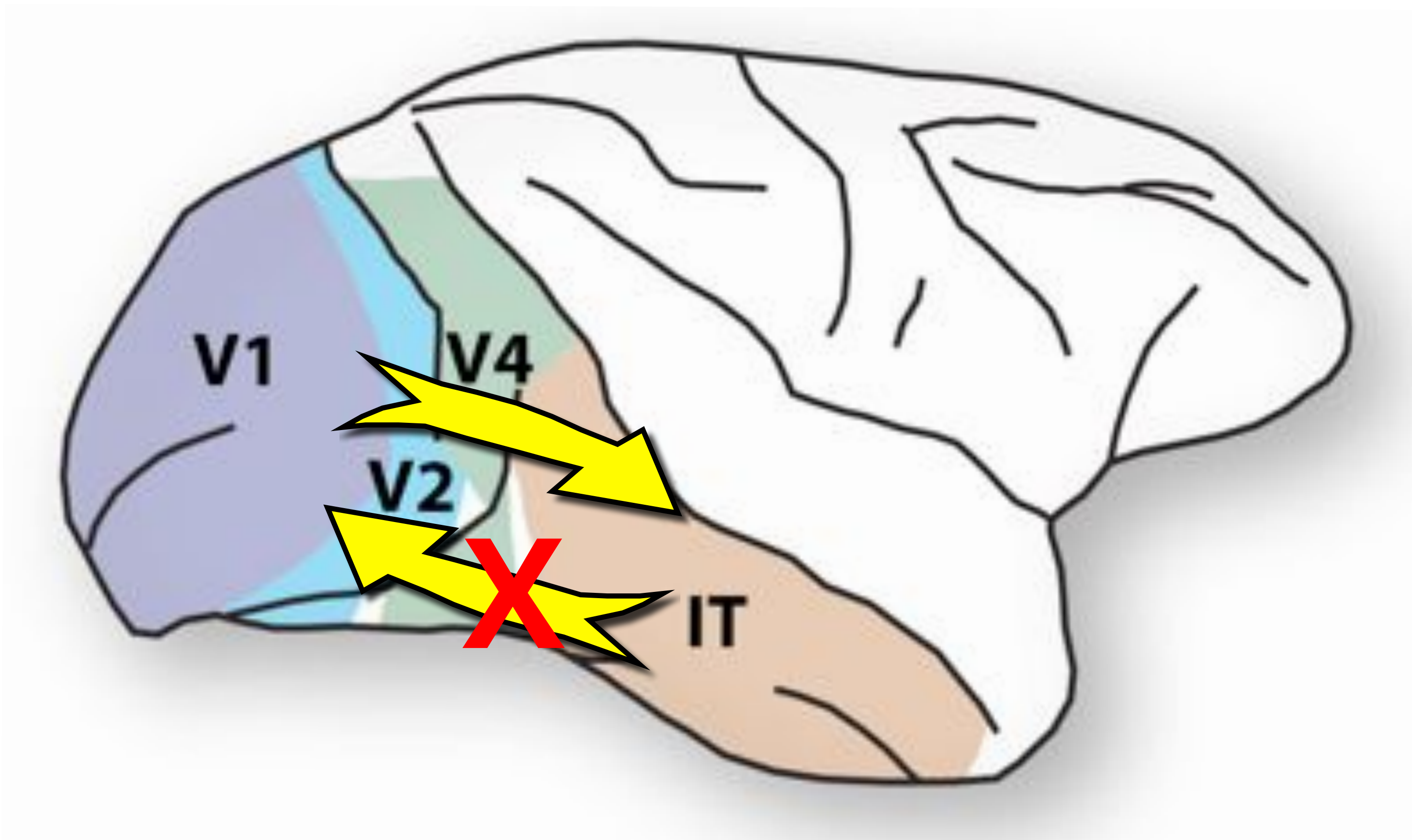
Brown University
Cognitive Linguistic & Psychological Sciences
Carney Institute for Brain Science

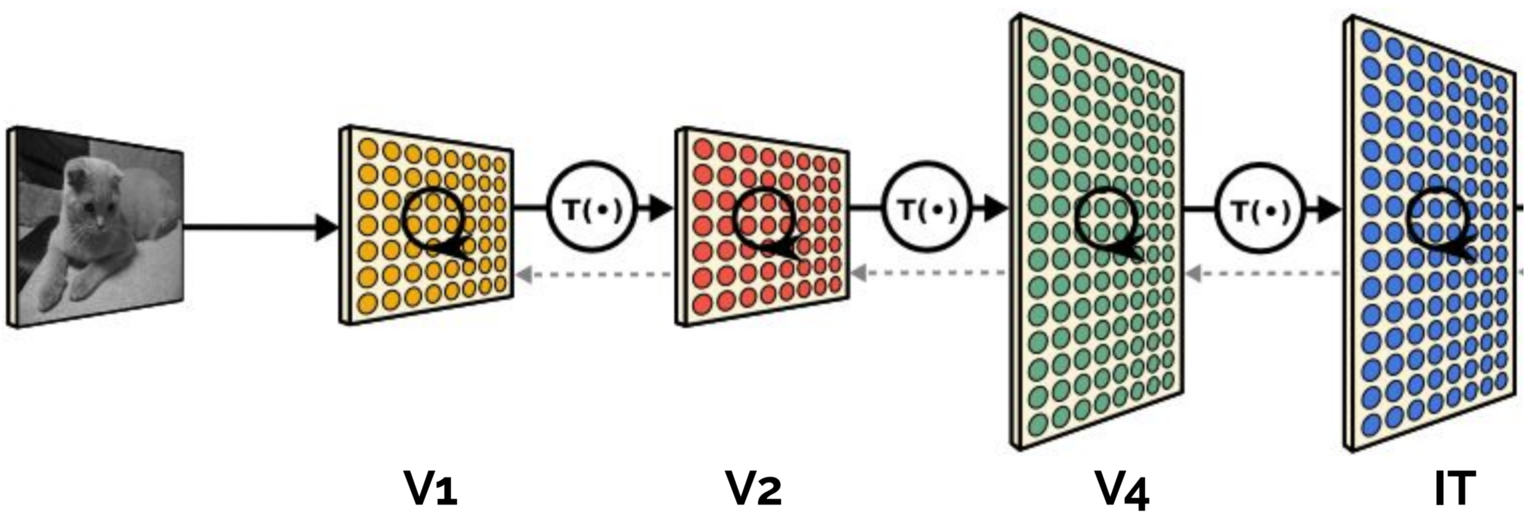
ANR-3IA Artificial and Natural Intelligence Toulouse Institute

Rapid categorization

Cauchoix Crouzet* Fize & Serre NeuroImage 2016*

Ventral stream
of the visual cortex





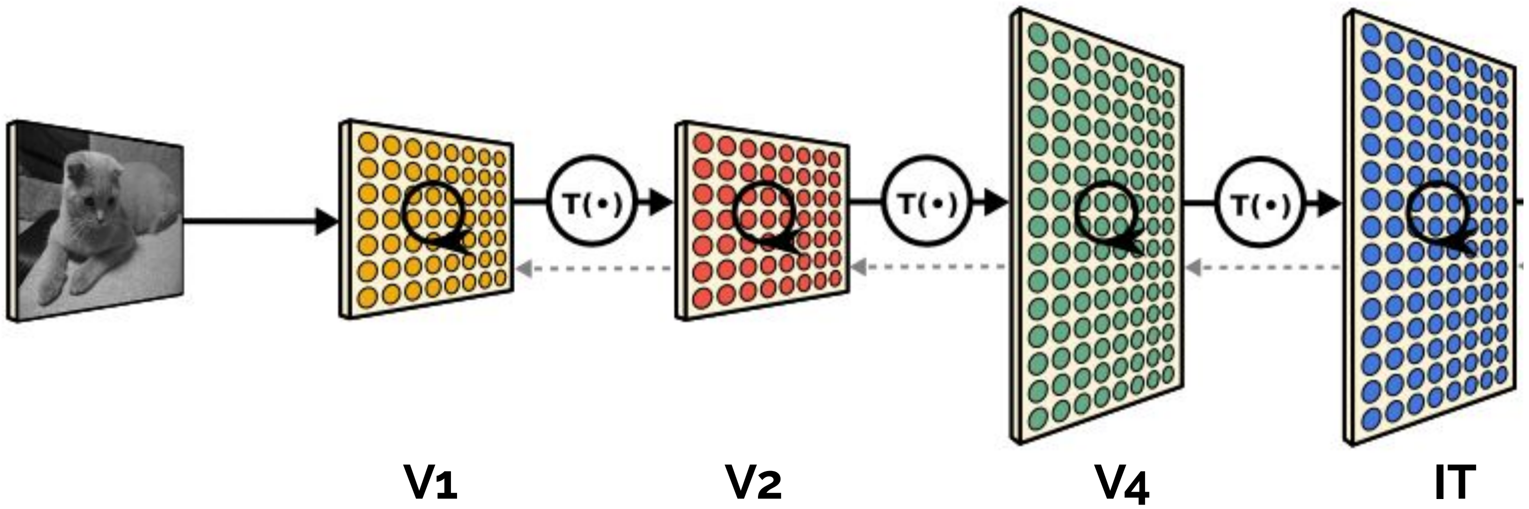
- Very fast latency <50-70 ms
- Category information <90-100 ms
- **Primarily feedforward (w| limited feedback)**
- **5-8 processing stages**

Image source: Jim Dicarlo

Cauchoix Crouzet* Fize & Serre NeuroImage 2016; see also Hung Kreiman et al 2005*

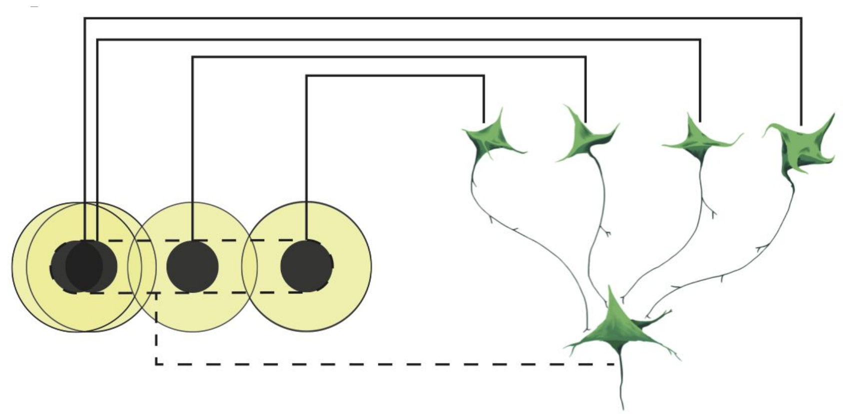
Computational models

- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)

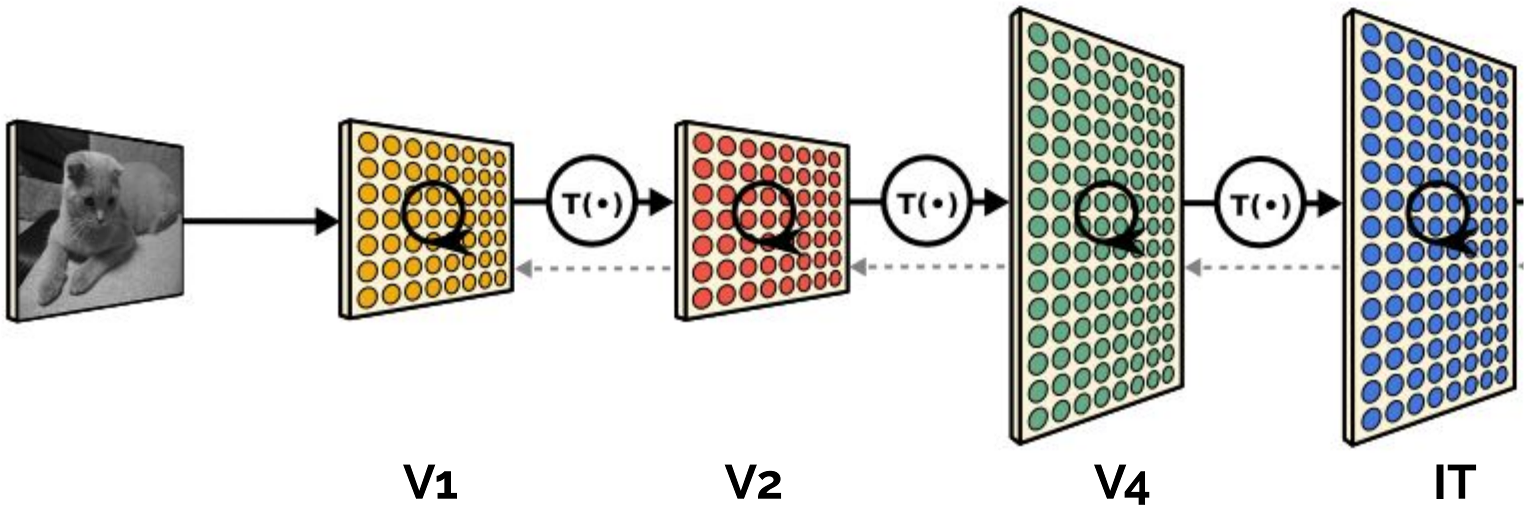


See Serre Ann Rev of Vision Science 2019 for a review

Computational
models

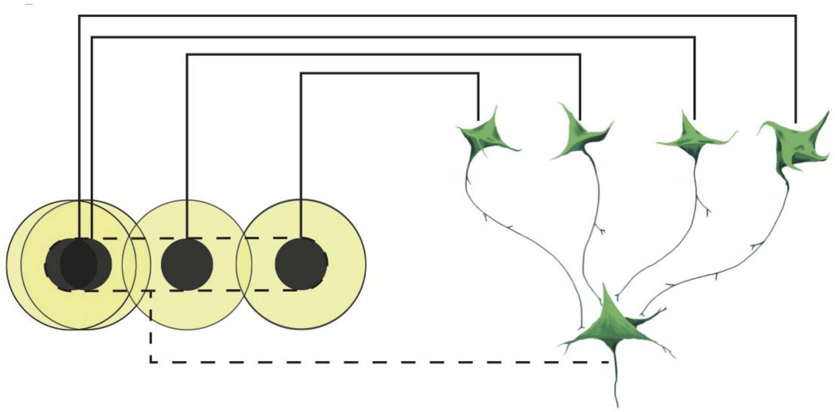


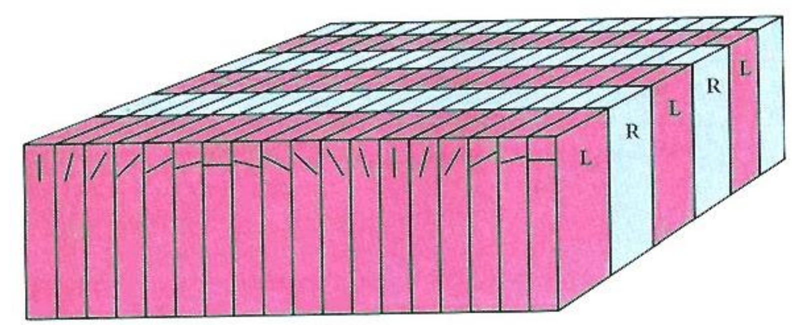
- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)



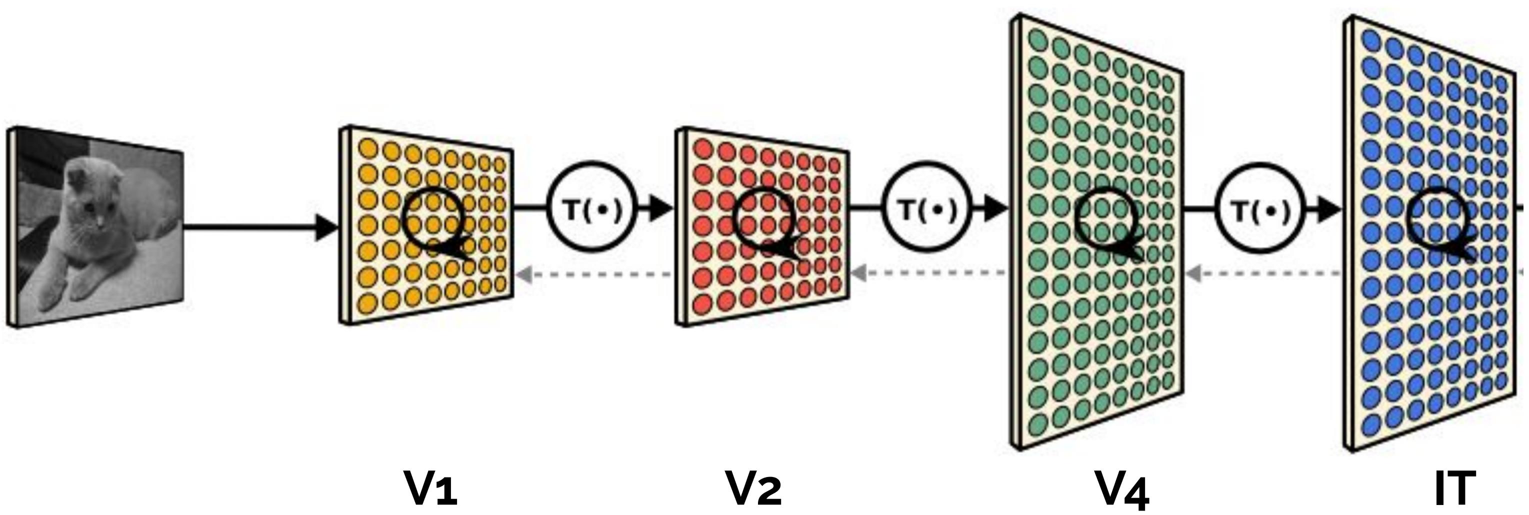
See Serre Ann Rev of Vision Science 2019 for a review

Computational models



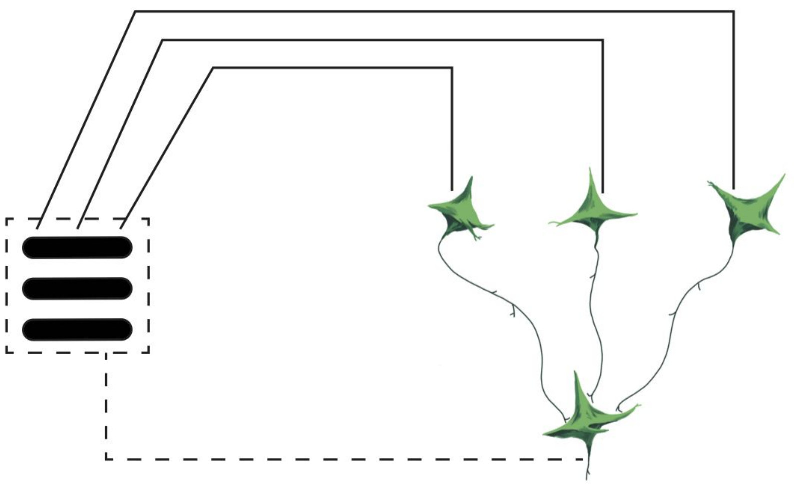


- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)

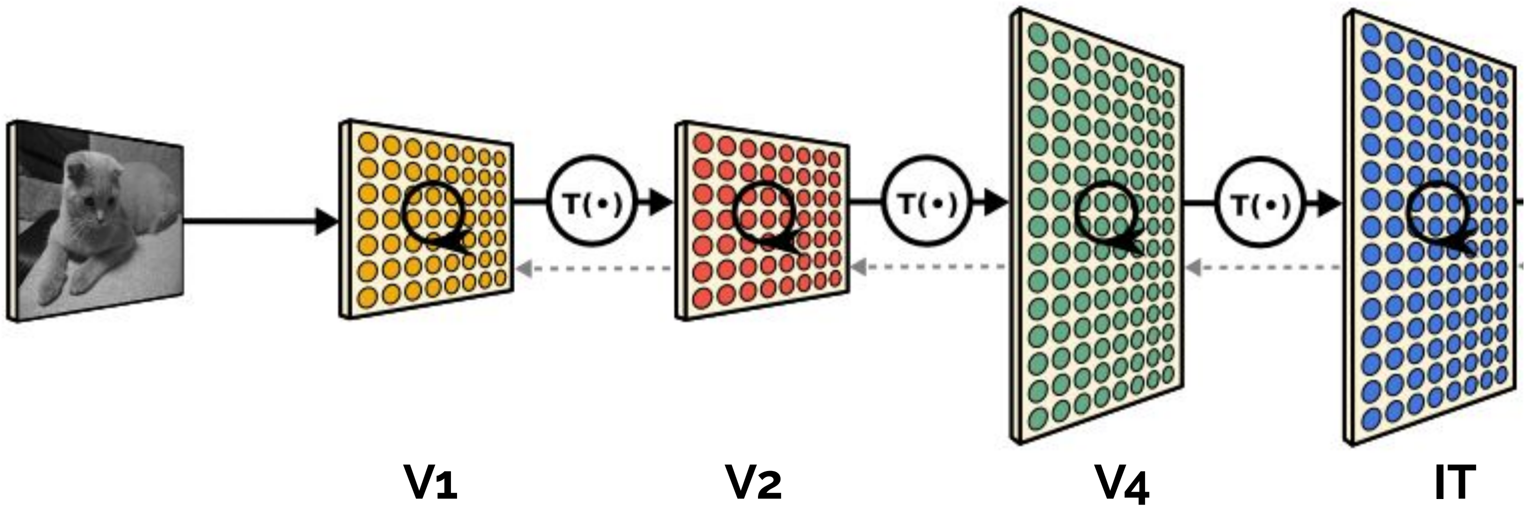


See Serre Ann Rev of Vision Science 2019 for a review

Computational models



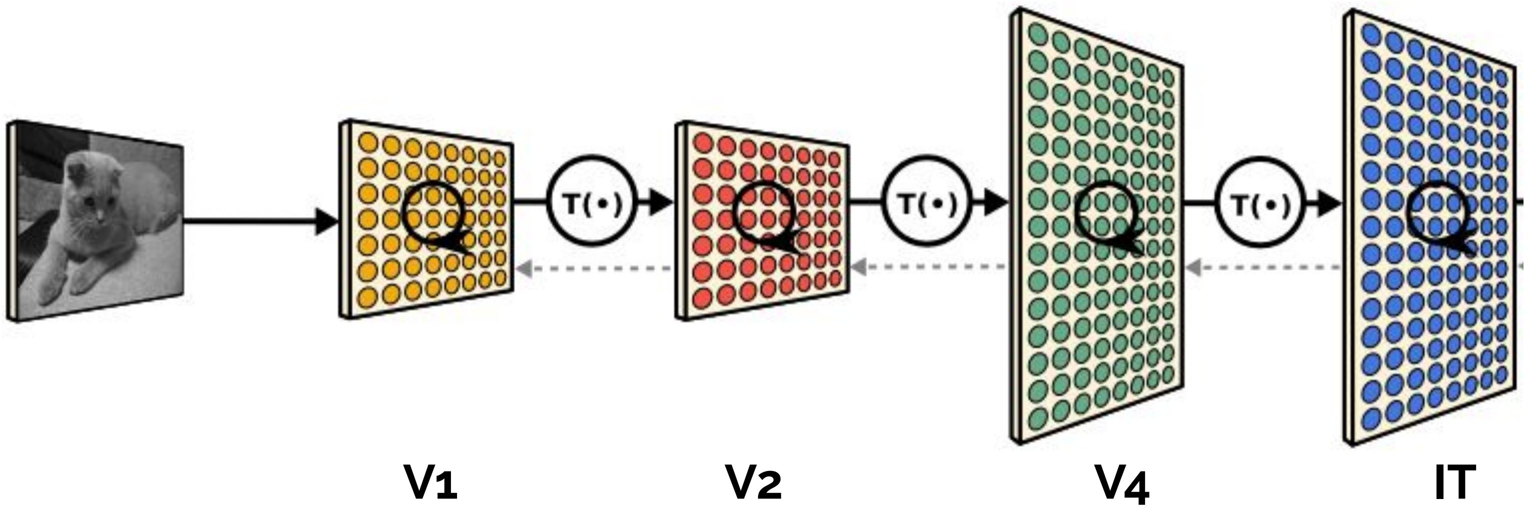
- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)



See Serre Ann Rev of Vision Science 2019 for a review

Computational models

- Consistent w|anatomy and physiology
- Consistent with H and non-H primate decisions during rapid categorization



See Serre Ann Rev of Vision Science 2019 for a review

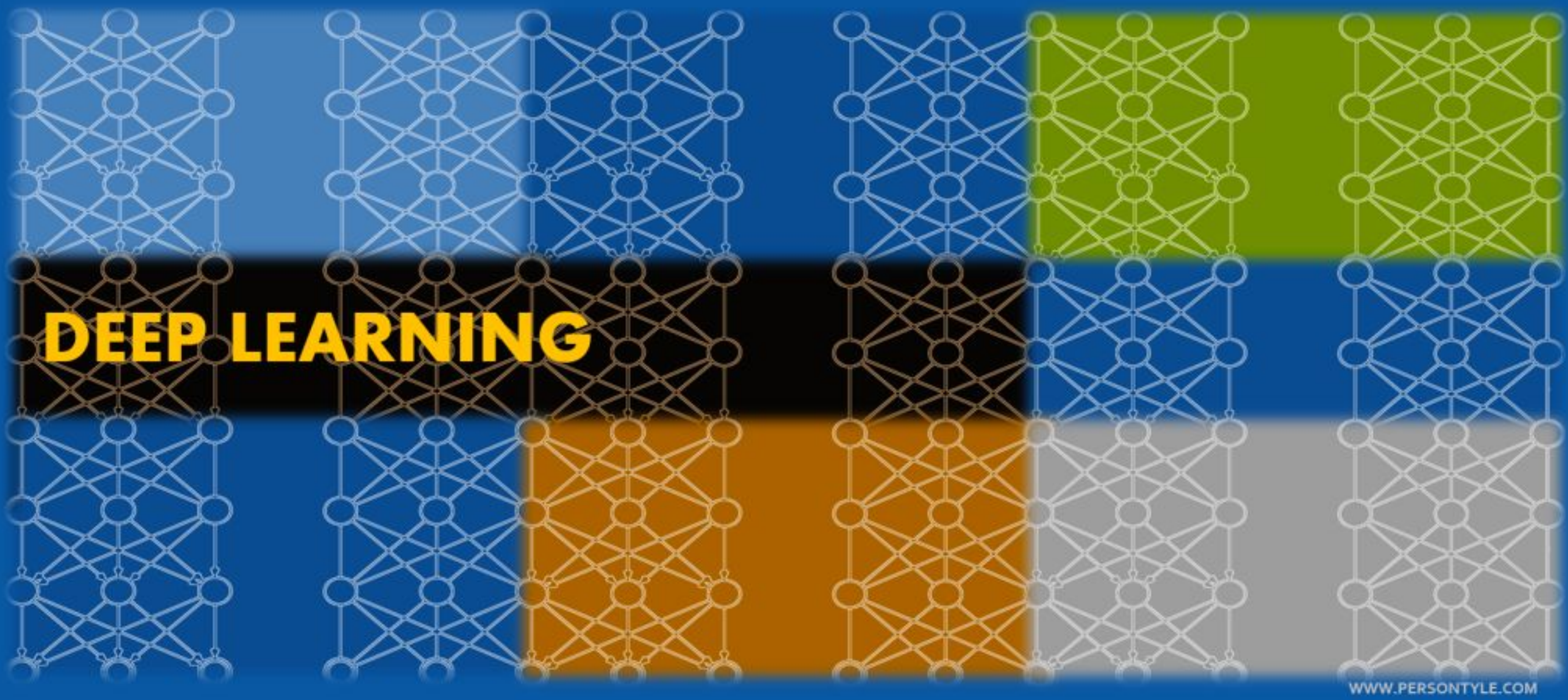


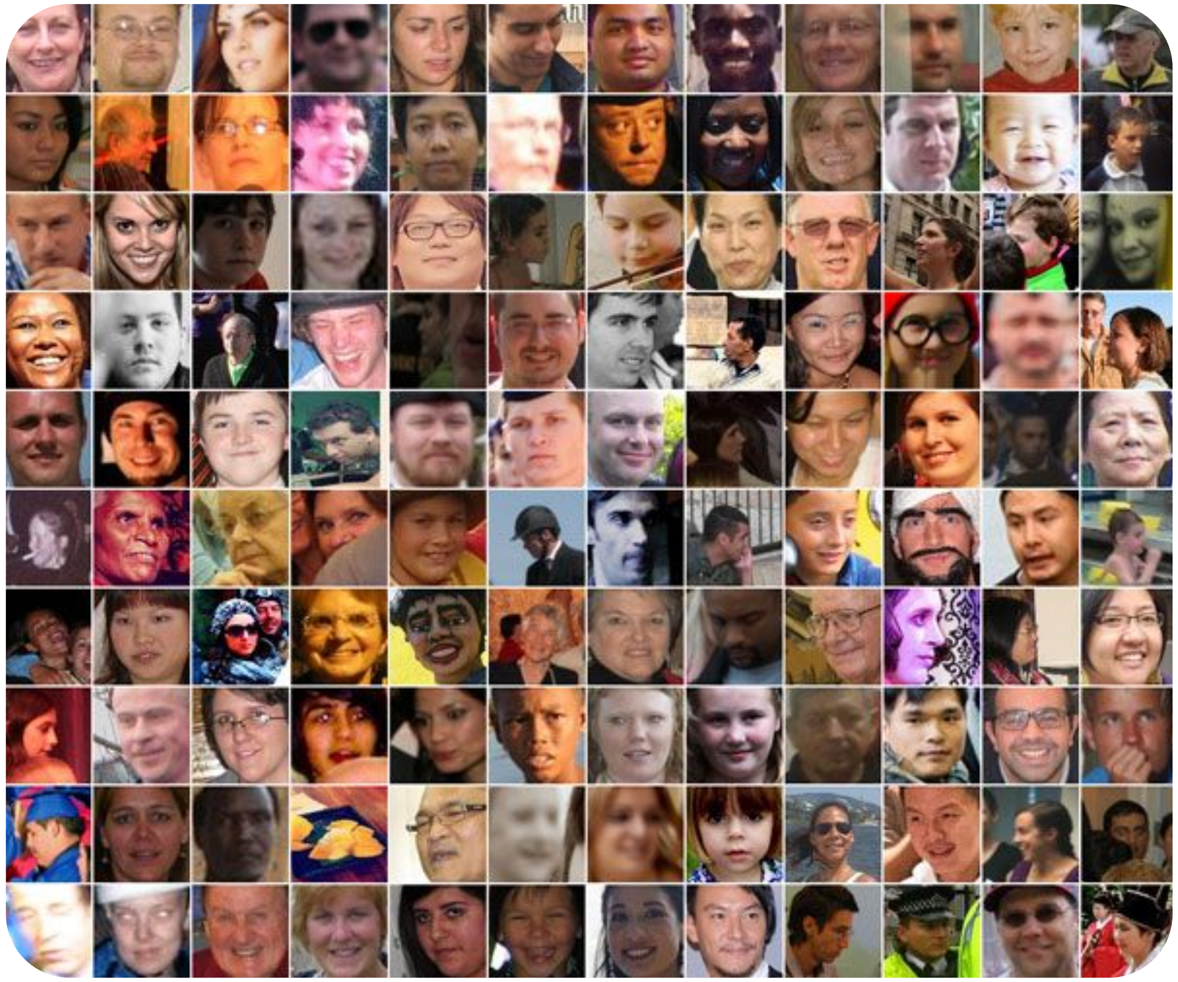
image source: http://www.persontyle.com/deep-learning-making-ai-intelligent-smarter

The good

The deep net
revolution

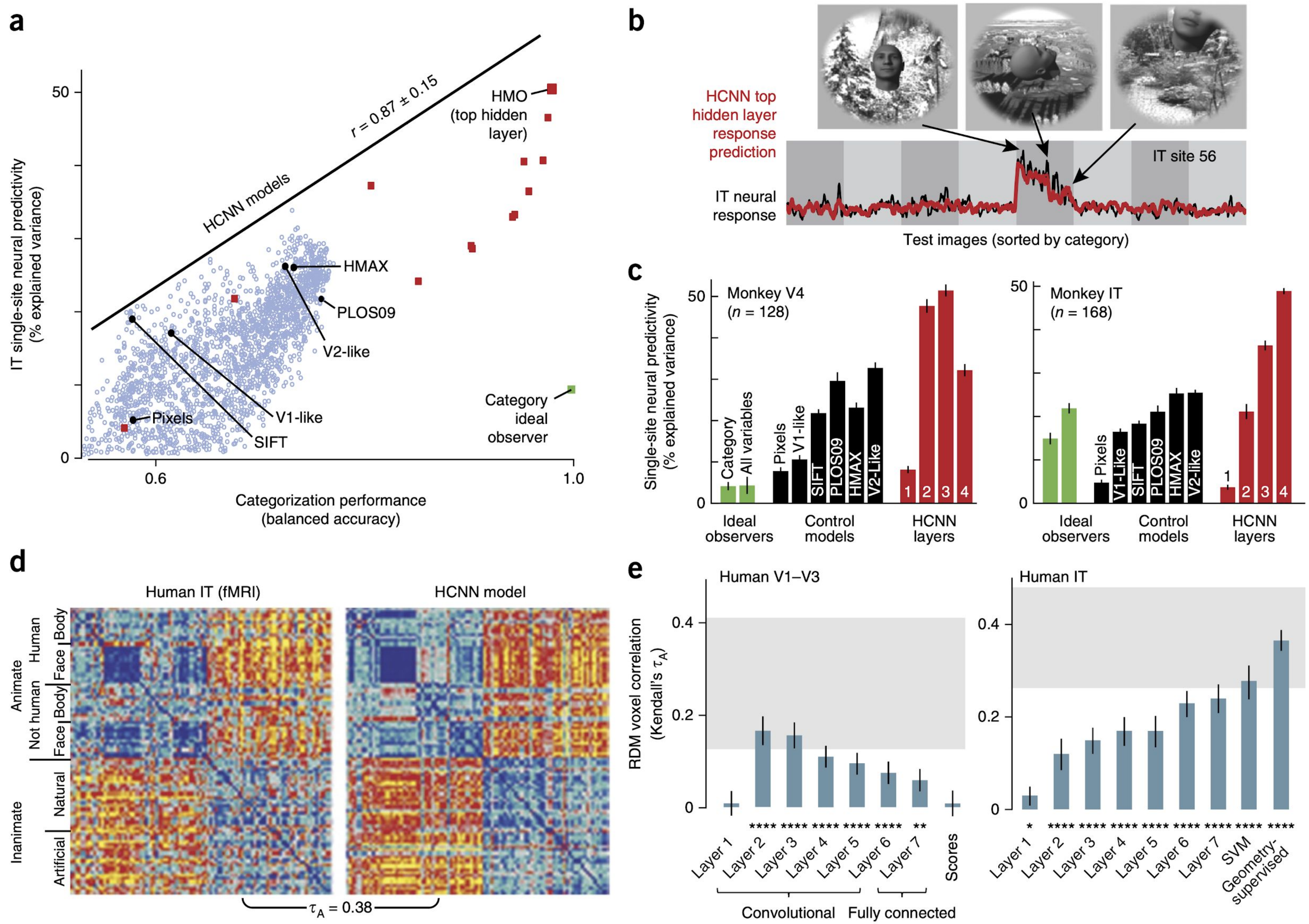
Object recognition (ImageNet)

face recognition (MegaFace)



See Serre Ann Rev of Vision Science 2019 for a review

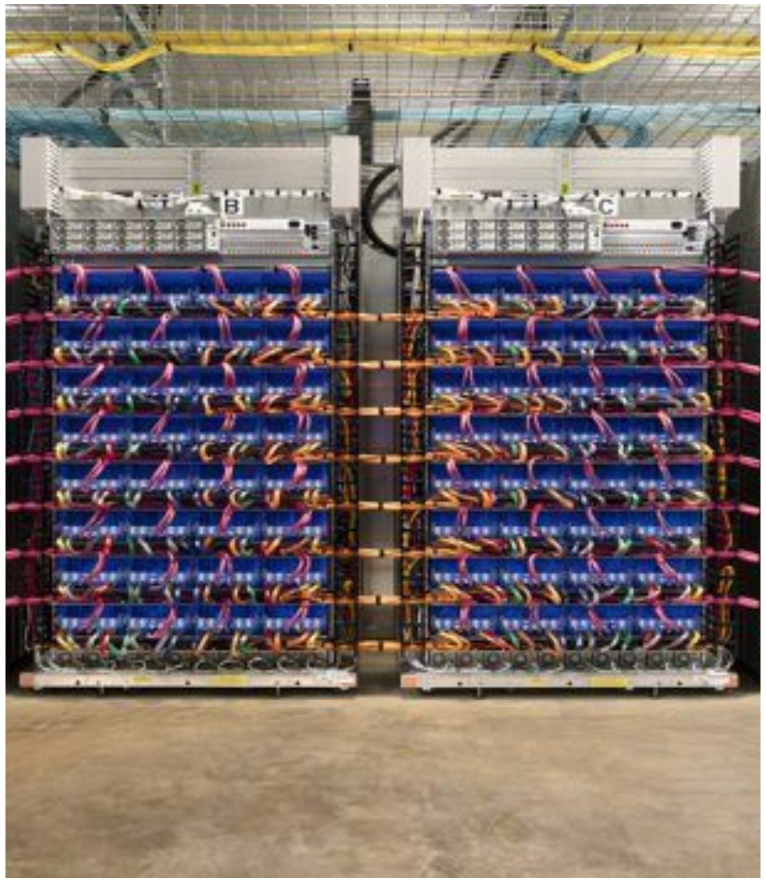
The deep net revolution

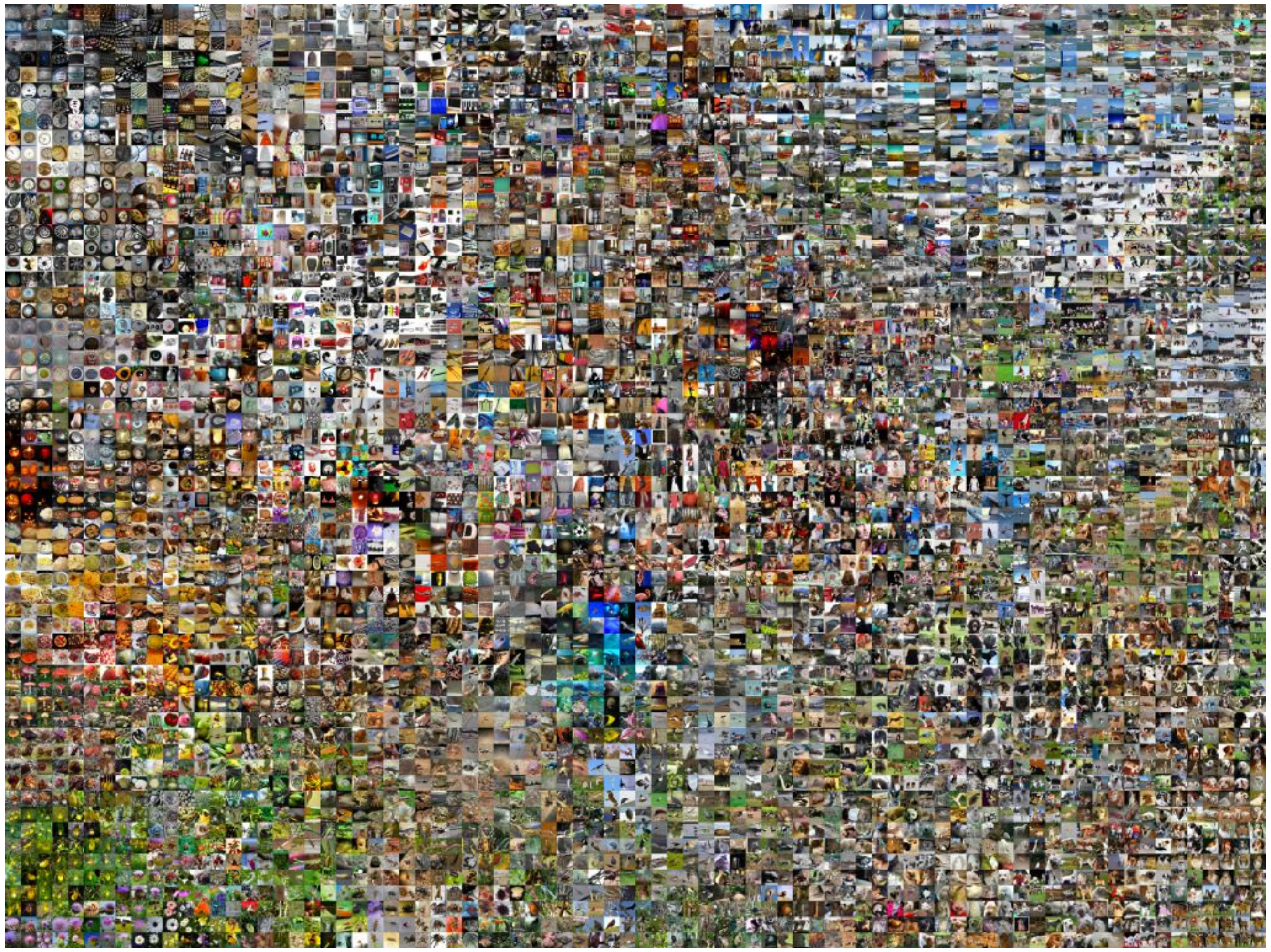


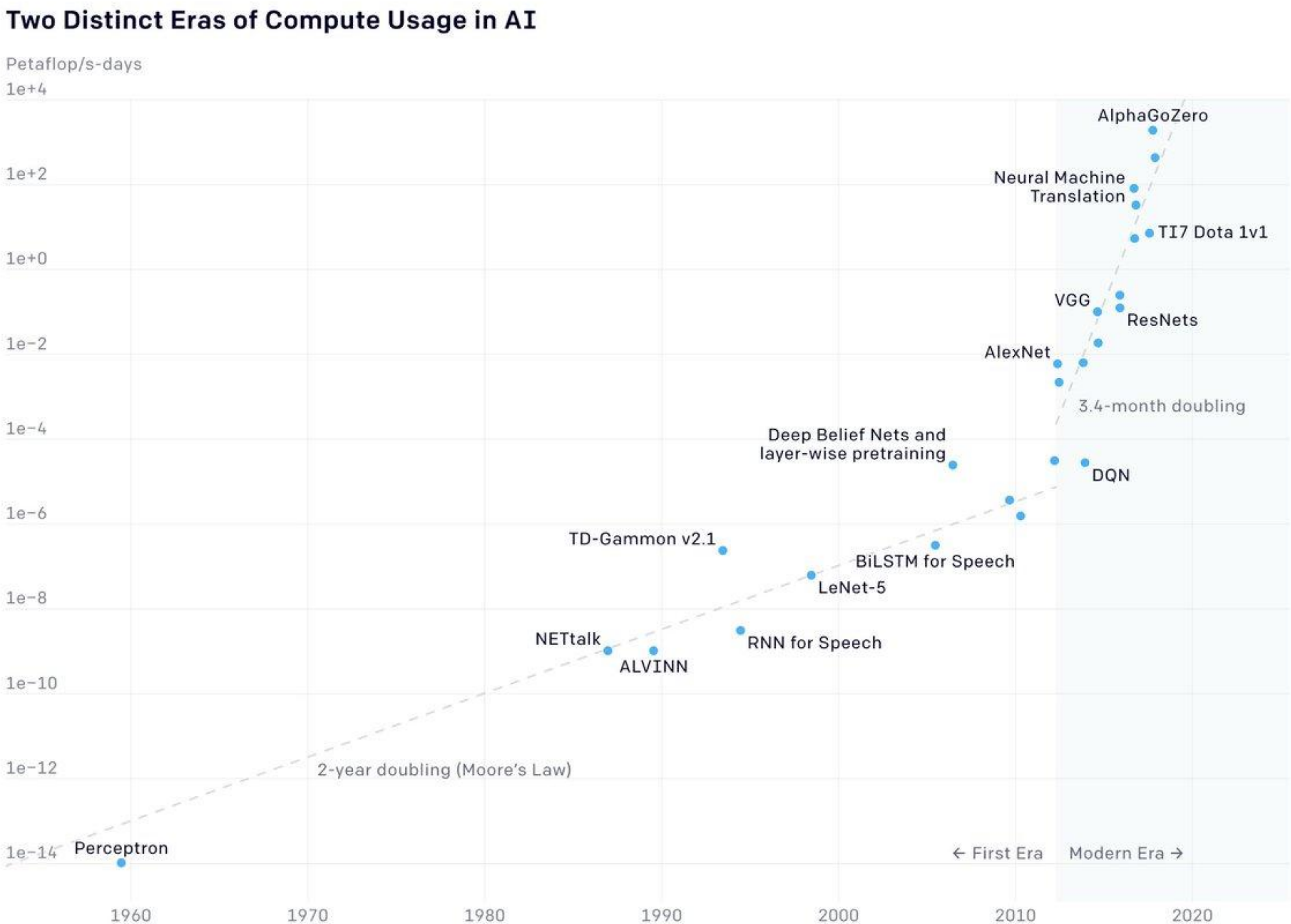
Source: Yamins & DiCarlo 2016

The bad

The deep net revolution



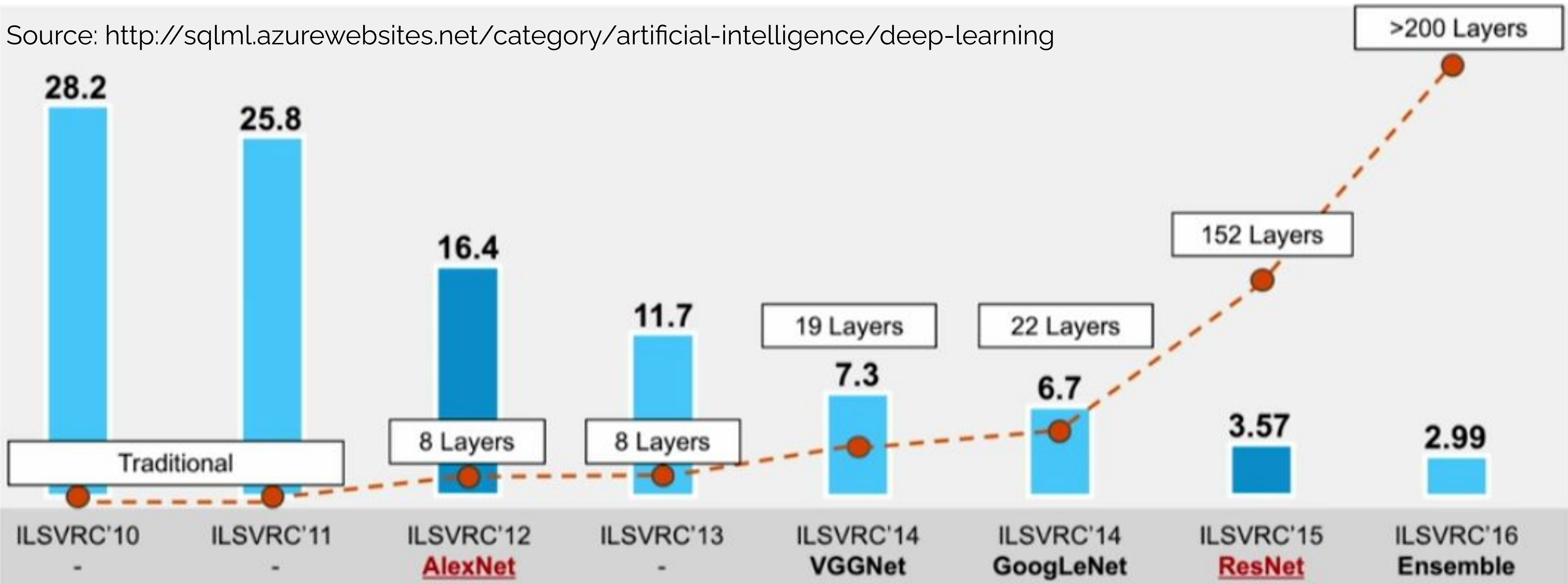




Source: openAI

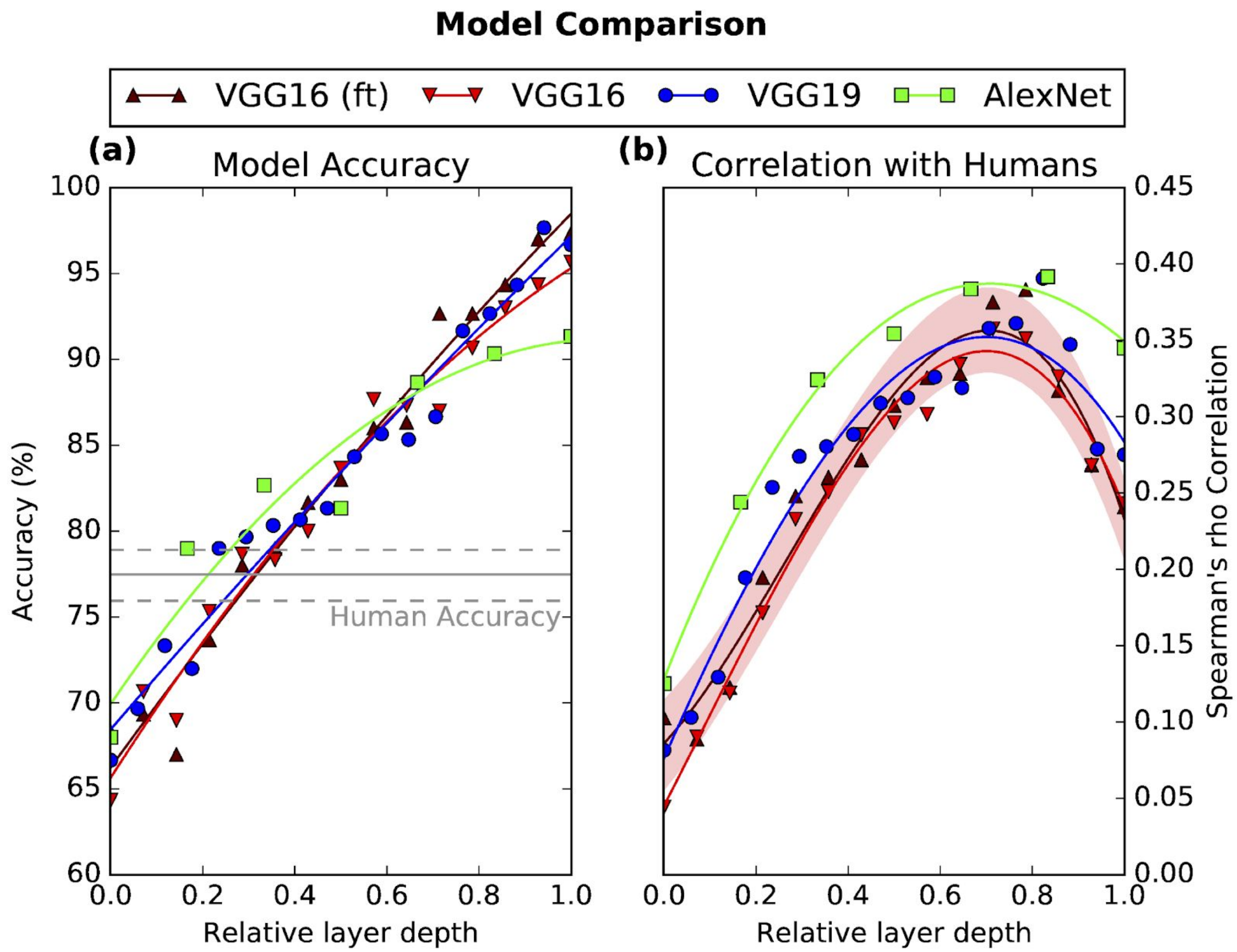
** A petaflop/s-day (pfs-day) consists of performing 10^{15} neural net operations per second for one day, or a total of about 10^{20} operations.*

The deep net
revolution



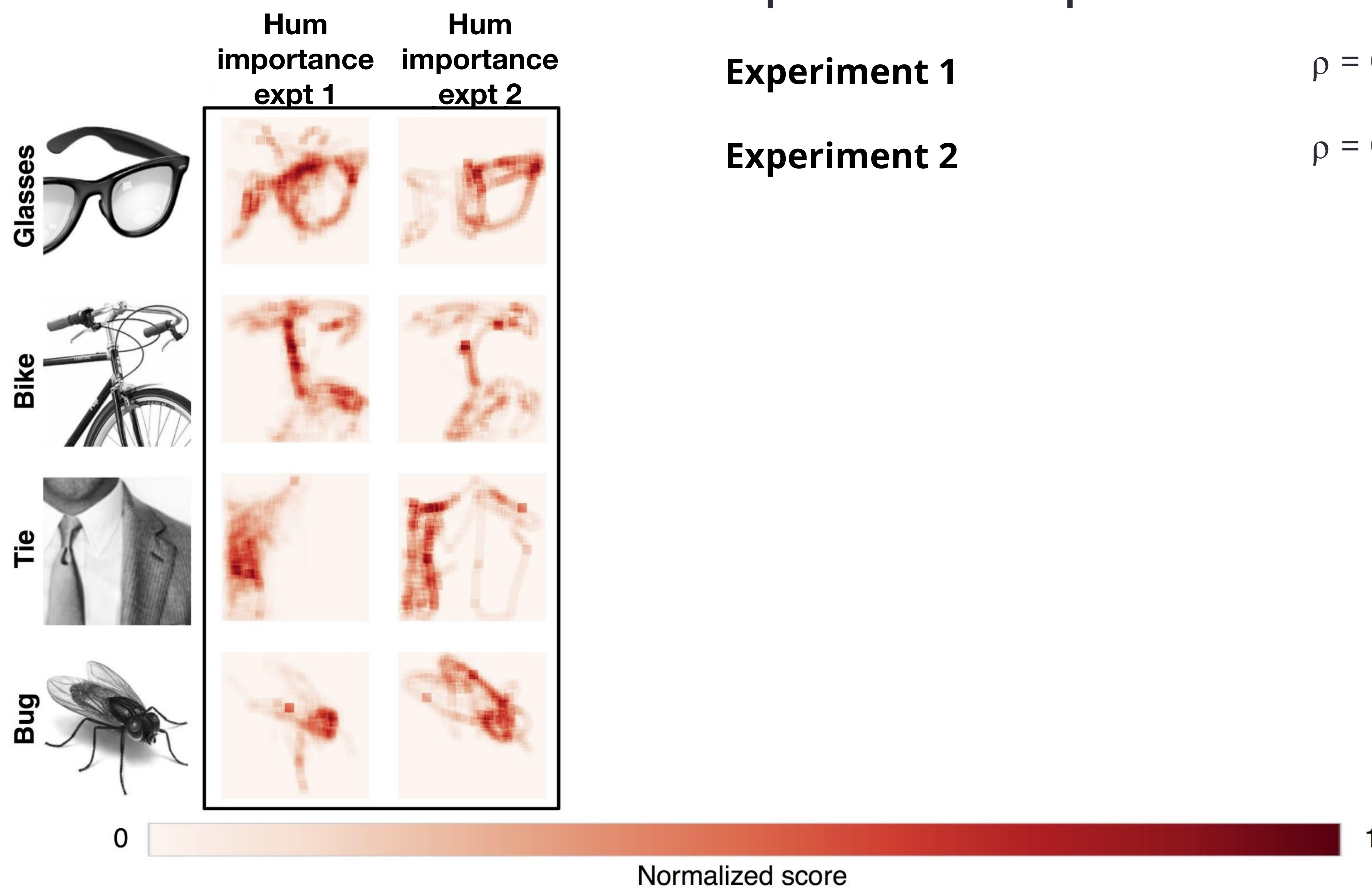
Deep net
limitations

Unsustainable growth



Eberhardt Cader & Serre NeurIPS 2016

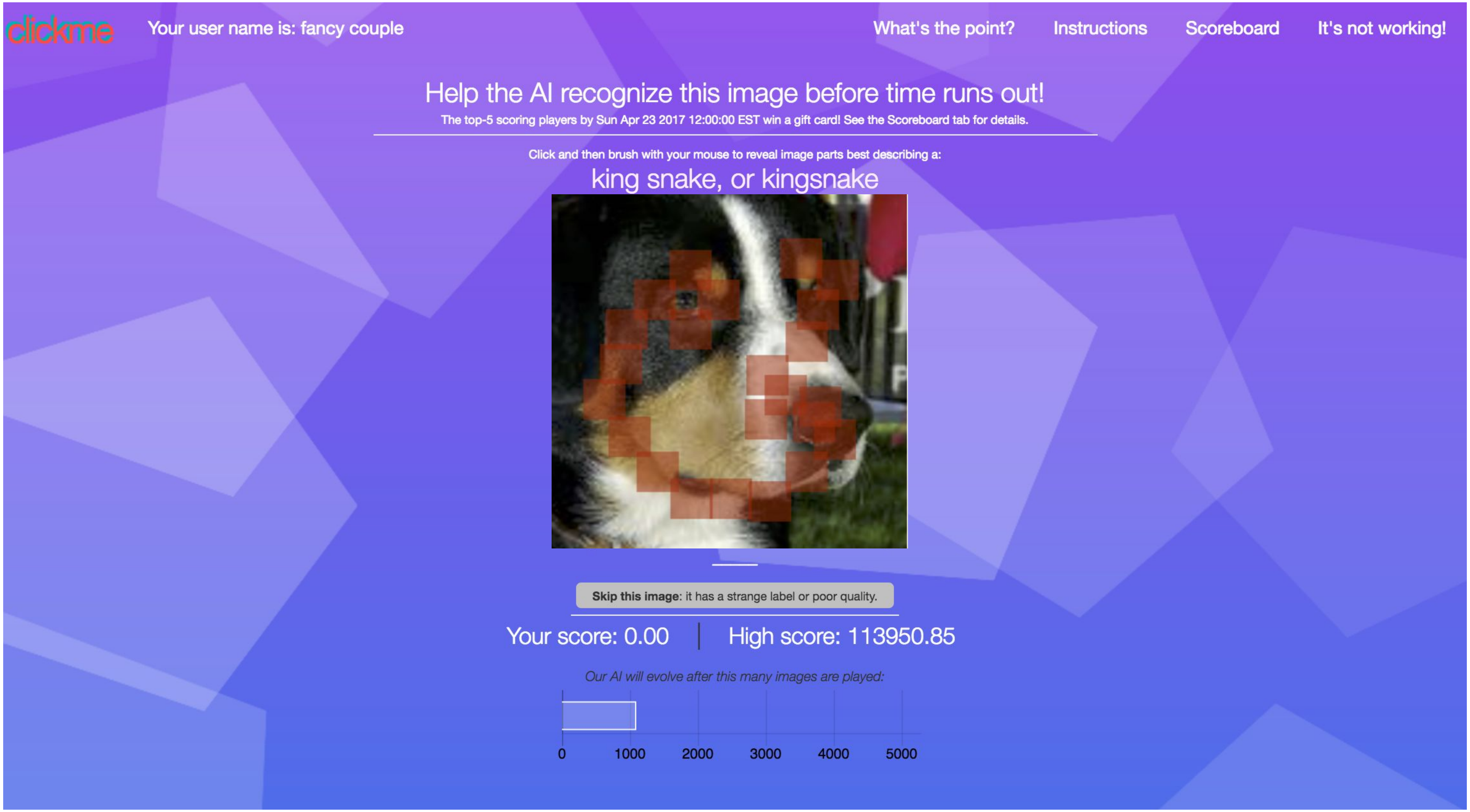
Deep net
limitations



	Reliability
Experiment 1 vs. Experiment 2	ρ = 0.55***
Experiment 1	ρ = 0.79***
Experiment 2	ρ = 0.60***

Linsley Eberhardt Sharma Gupta & Serre ICCVw 2017

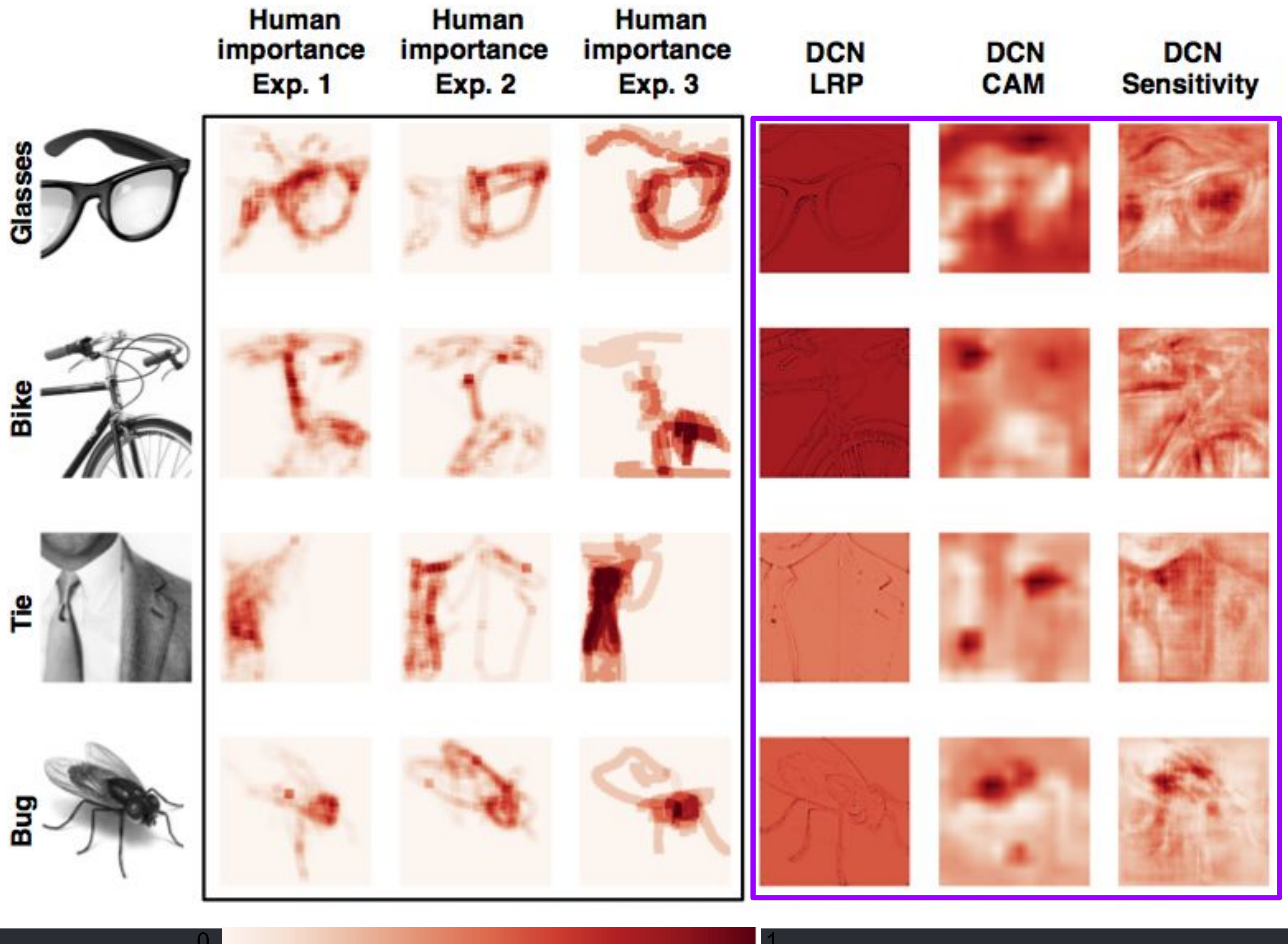
ClickMe.ai



http://www.clickme.ai; see also von Ahn et al 2006, Deng et al 2013

Deep net
limitations

	Reliability
LRP	$\rho = 0.33^{**}$
CAM	$\rho = 0.1$, ns
Sensitivity	$\rho = 0.03$, ns

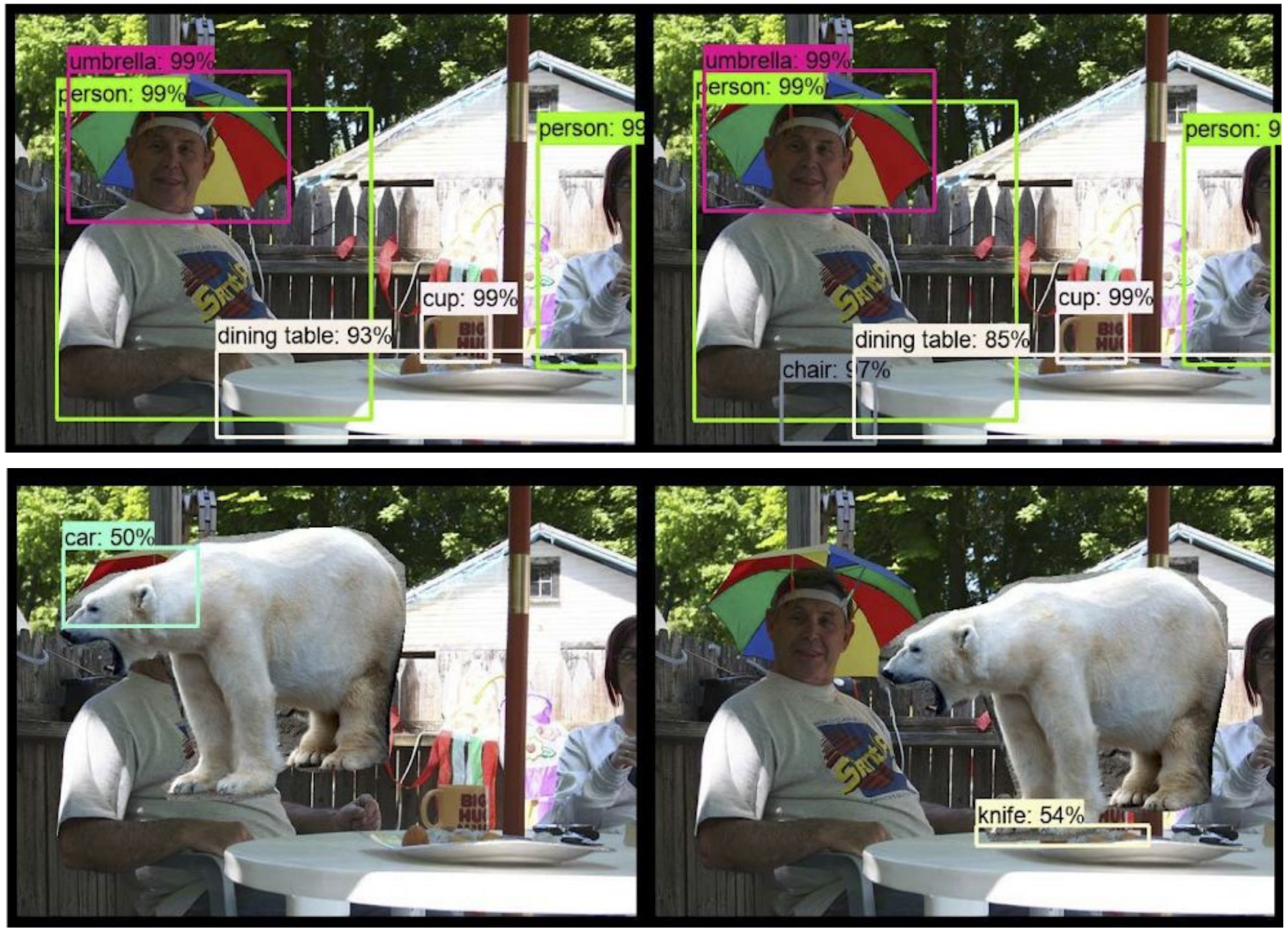


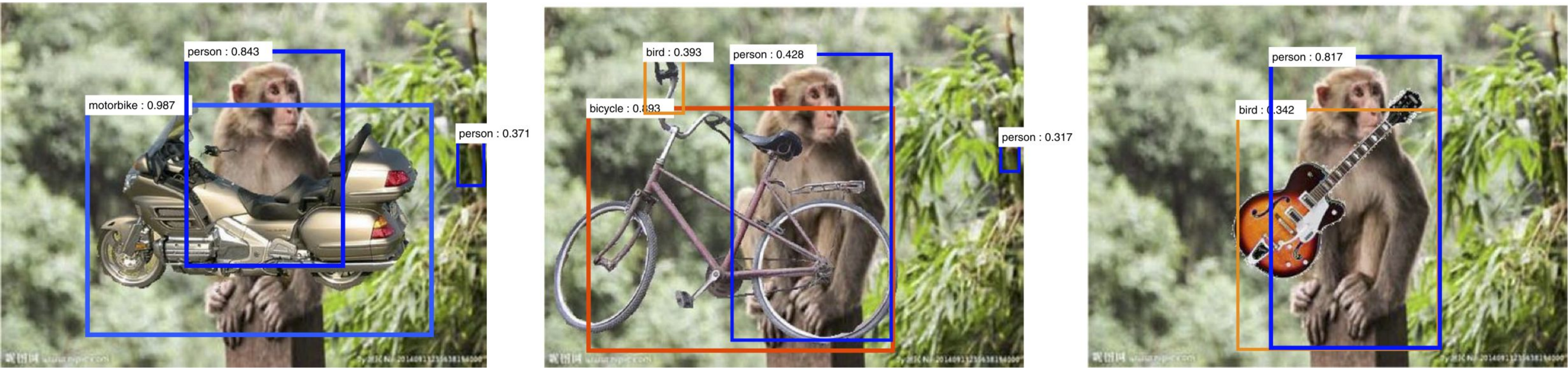
Linsley Eberhardt Sharma Gupta & Serre ICCVw 2017

The uggly

DNN limitations

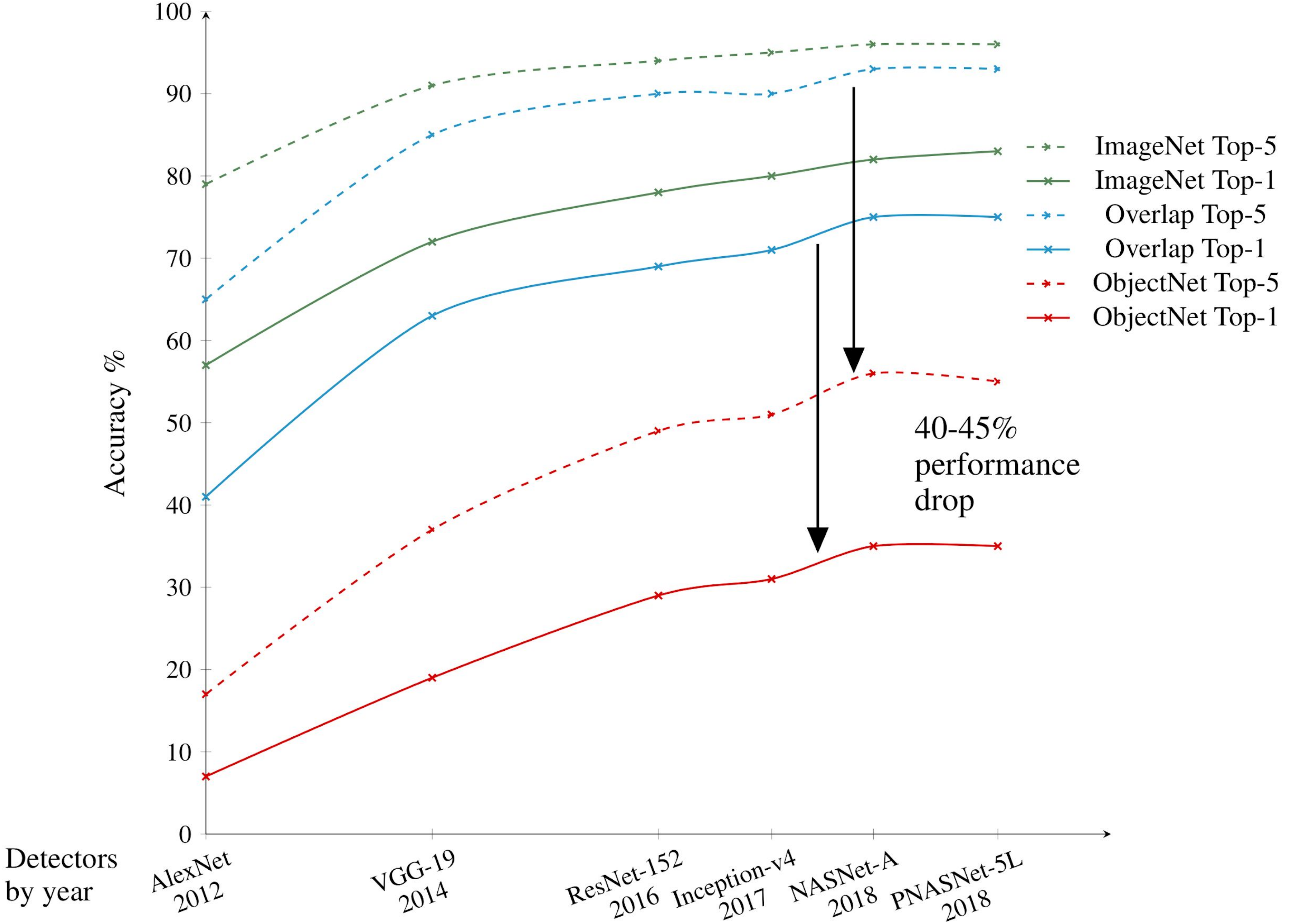
"Object transplanting"





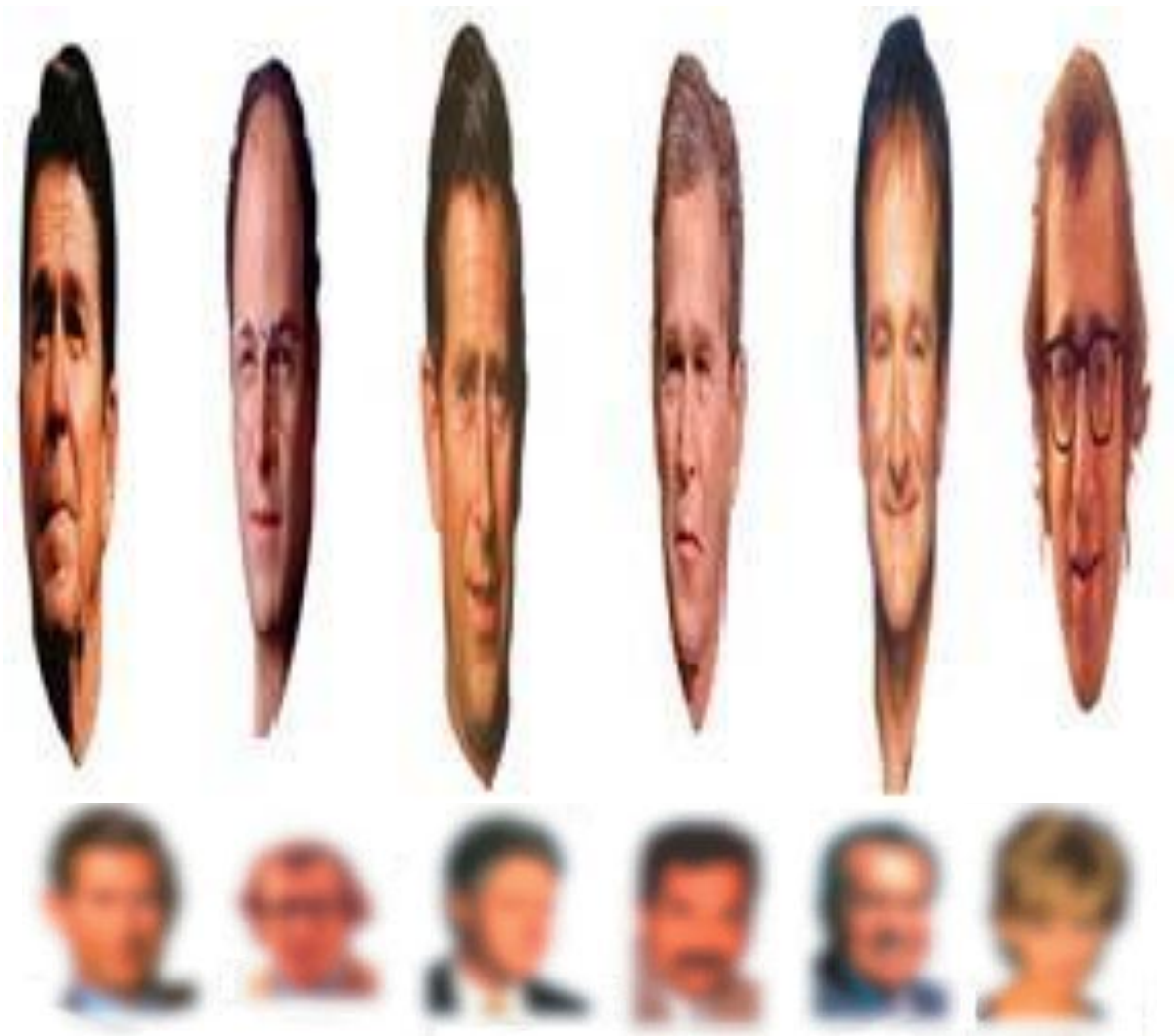
Wang et al 2018; Rosenfeld Zemel Tsotsos ArXiv 2018

Limited generalization



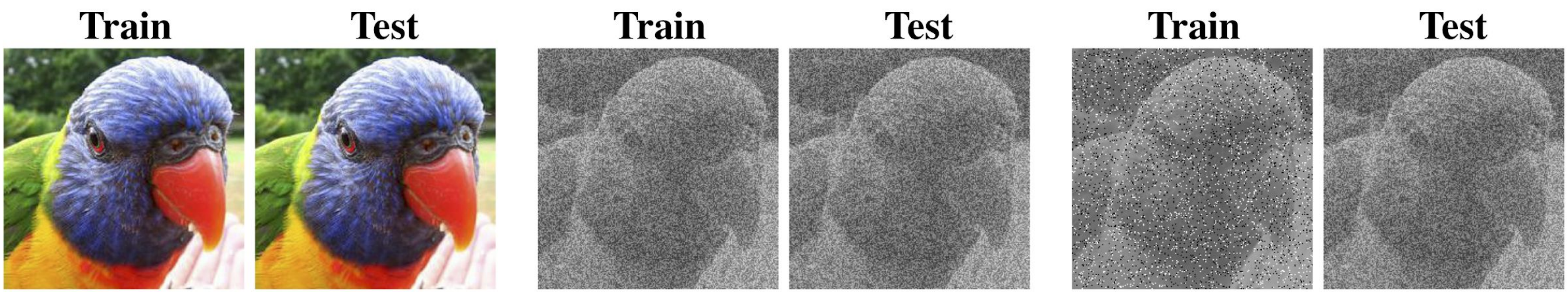
Barbu et al NeurIPS 2019; see also Recht et al ArXiv 2019

Generalizing to unseen degradations



Sinha Nat Neuroscience 2002

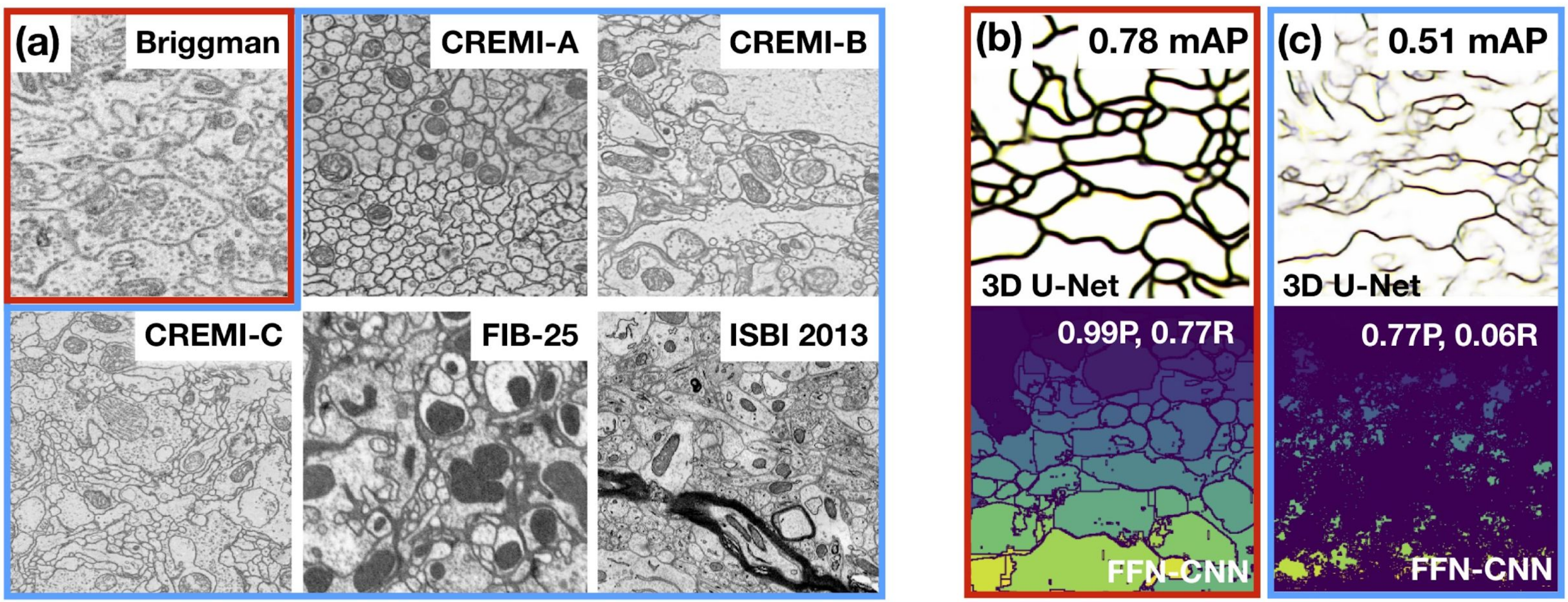
Limited
generalization



(a) Super-human performance (b) Super-human performance (c) Chance level performance

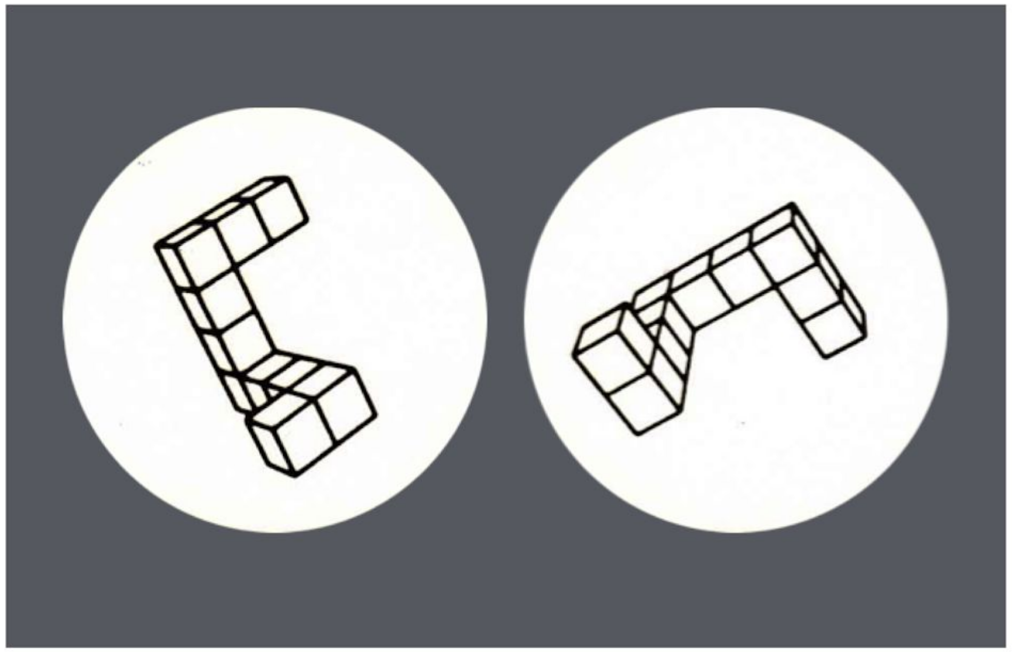
Gheiros et al NeurIPS 2018

Limited
generalization

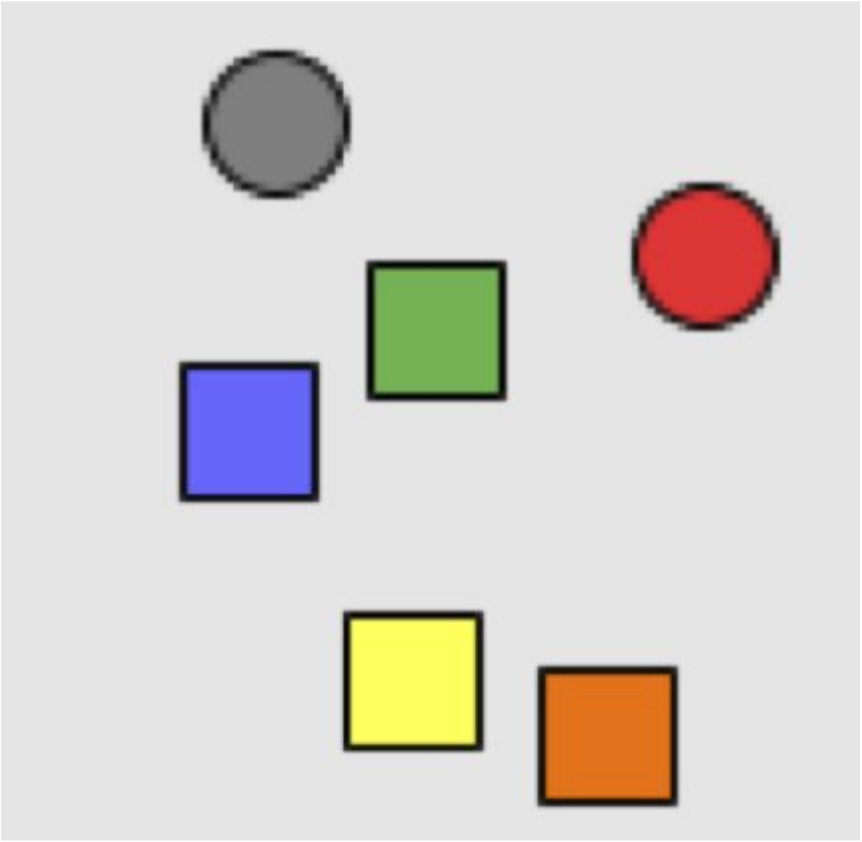


Linsley Kim Serre ArXiv 2018

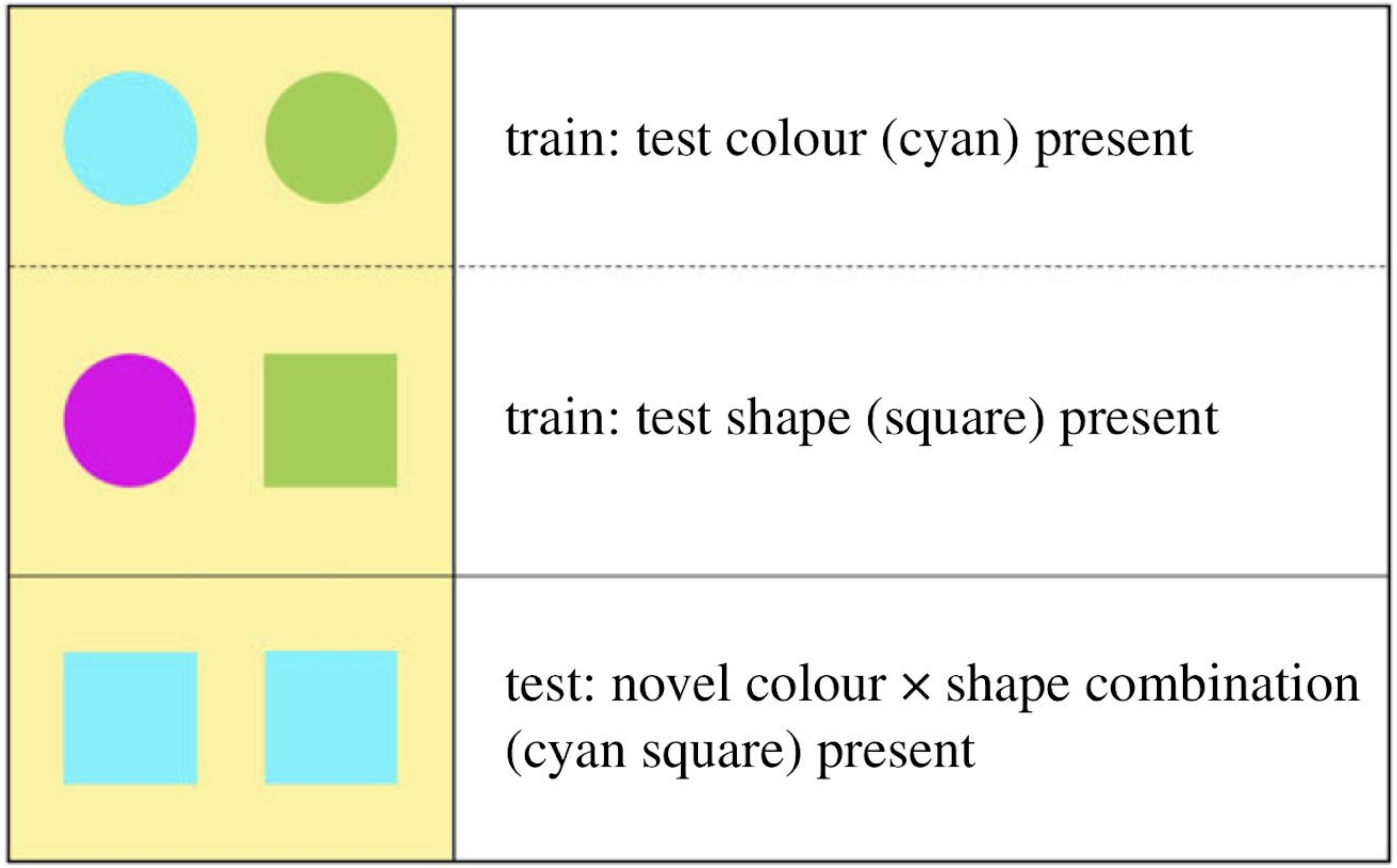
Limited generalization



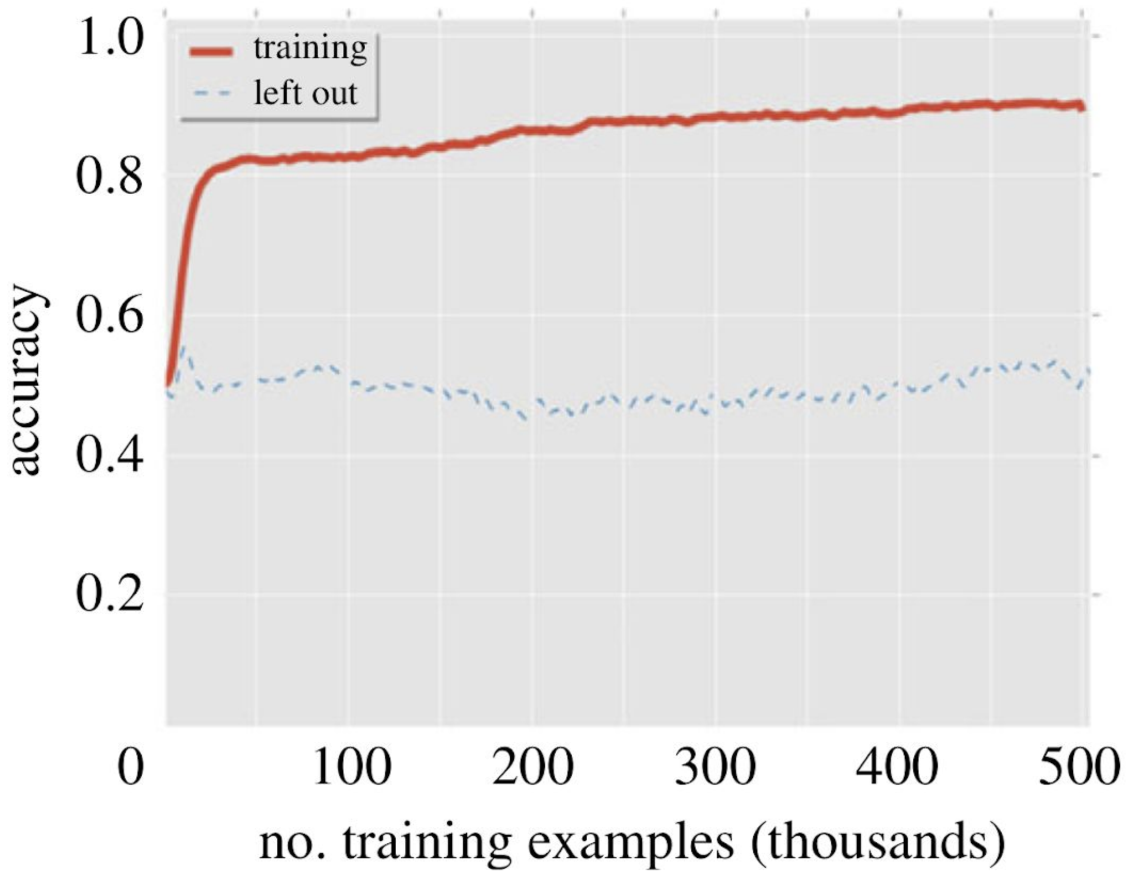
Sort-of-CLEVR dataset
(Santoro et al 2017)



Limited generalization



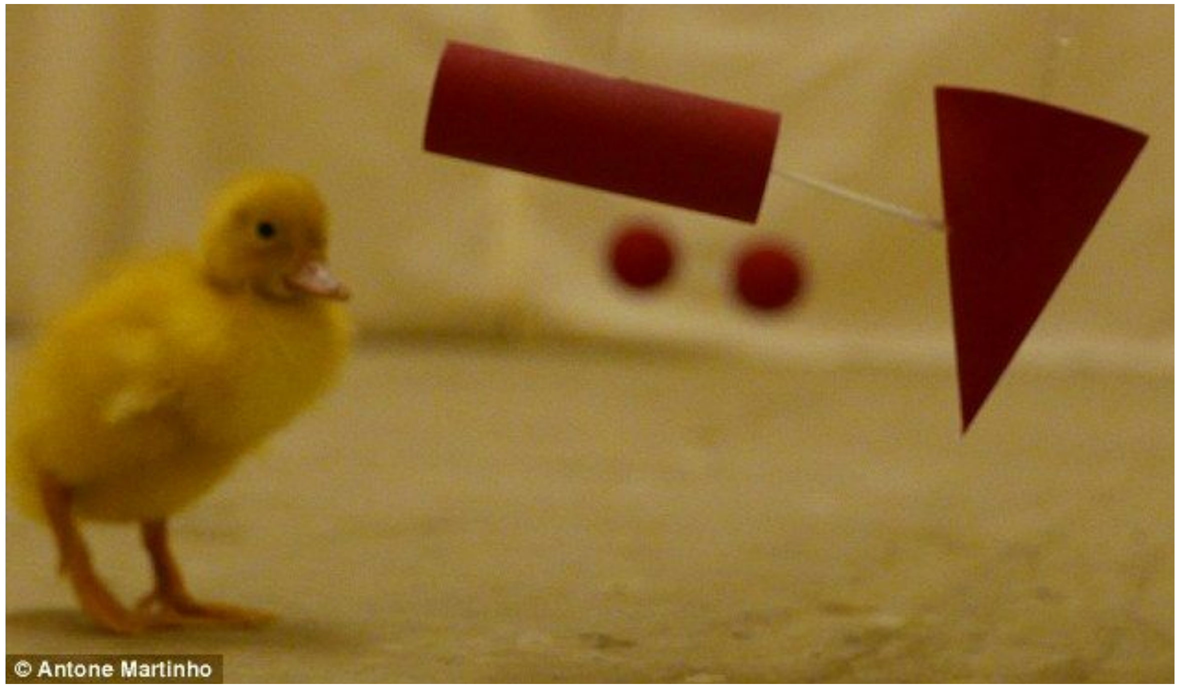
"Relational network"

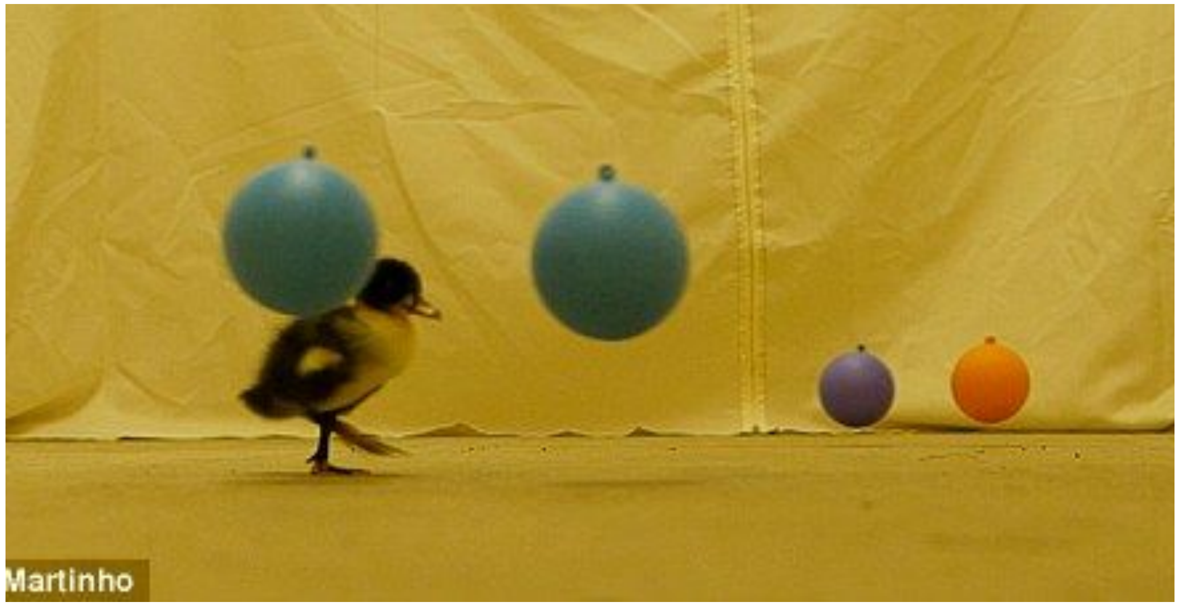


Kim Ricci & Serre Royal Soc Interface Focus 2018

One-shot learning
In newborn ducklings

- Computational evidence that same-different recognition task trivially solved w| perceptual grouping mechanisms

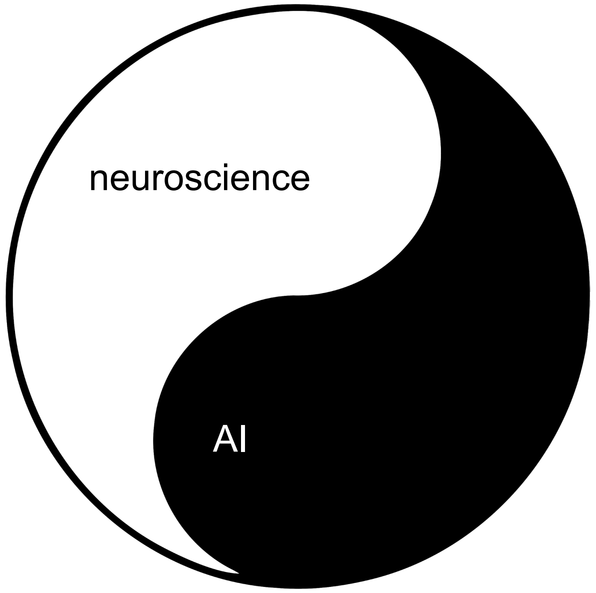




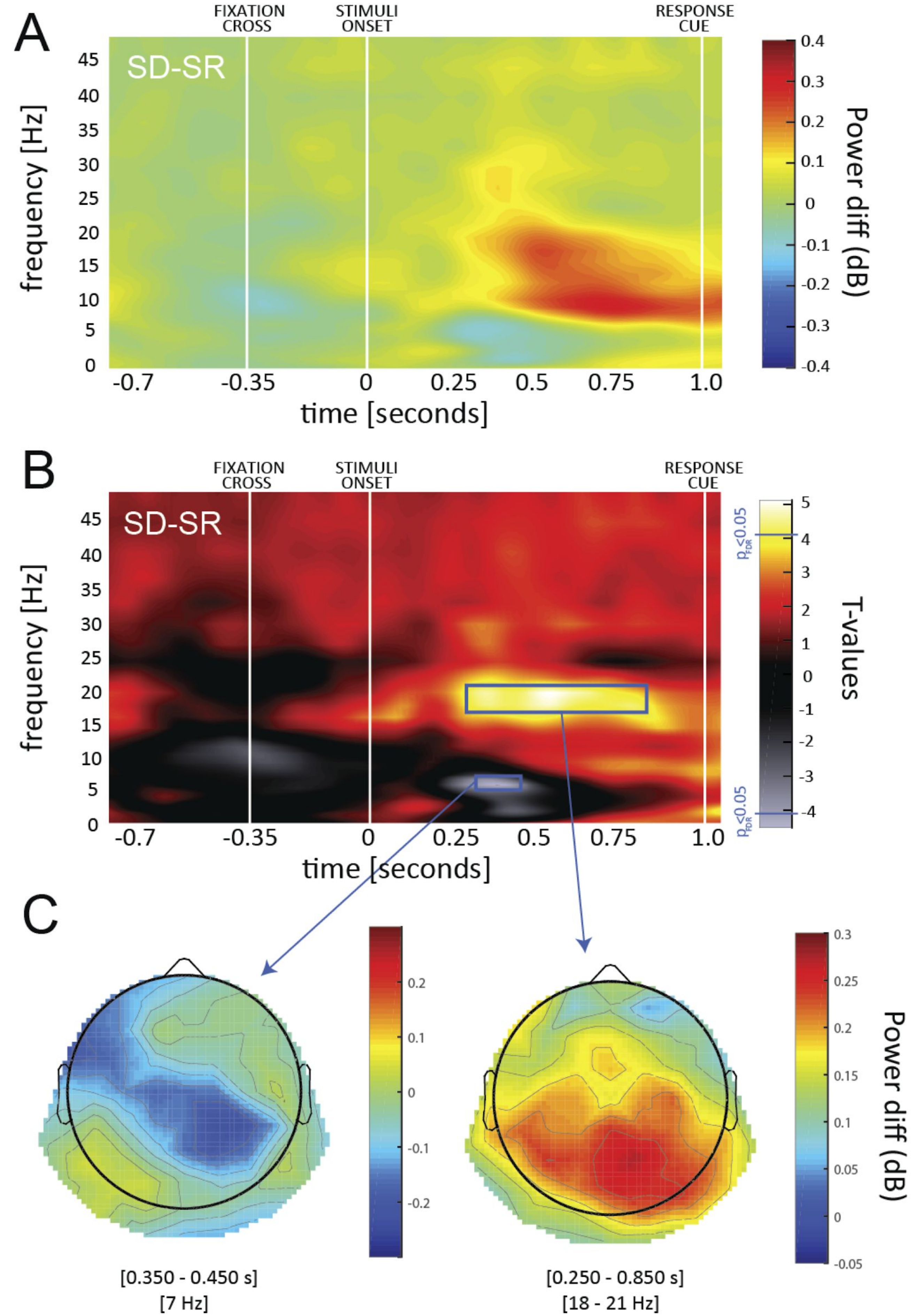
Kim Ricci & Serre Royal Soc Interface Focus 2018

Martinho III & Kacelnik Science 2016

The future

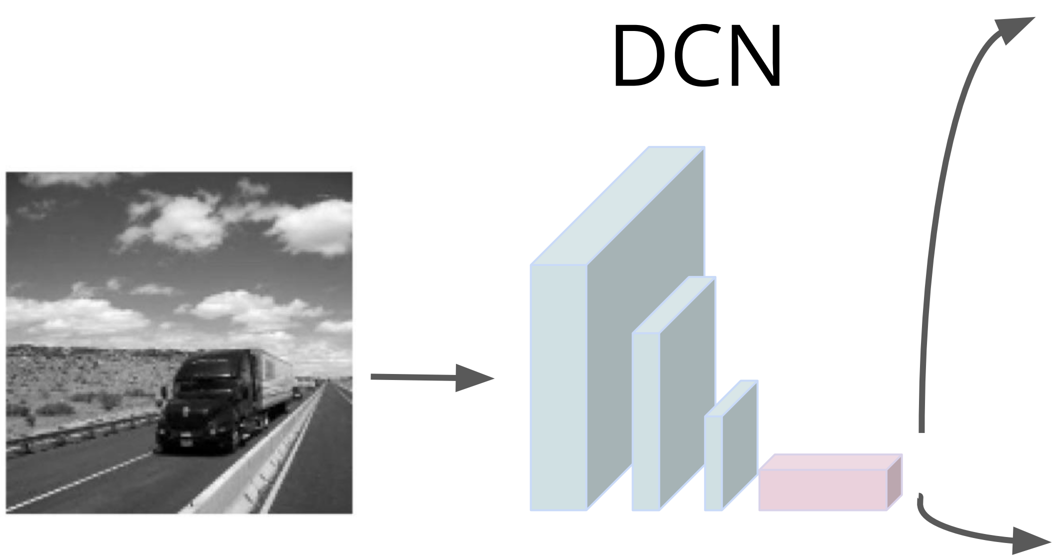


Attention and cortical feedback



Alamia Luo Kim Ricci & Serre VanRullen in prep

Driving DCNs towards
human-like recognition

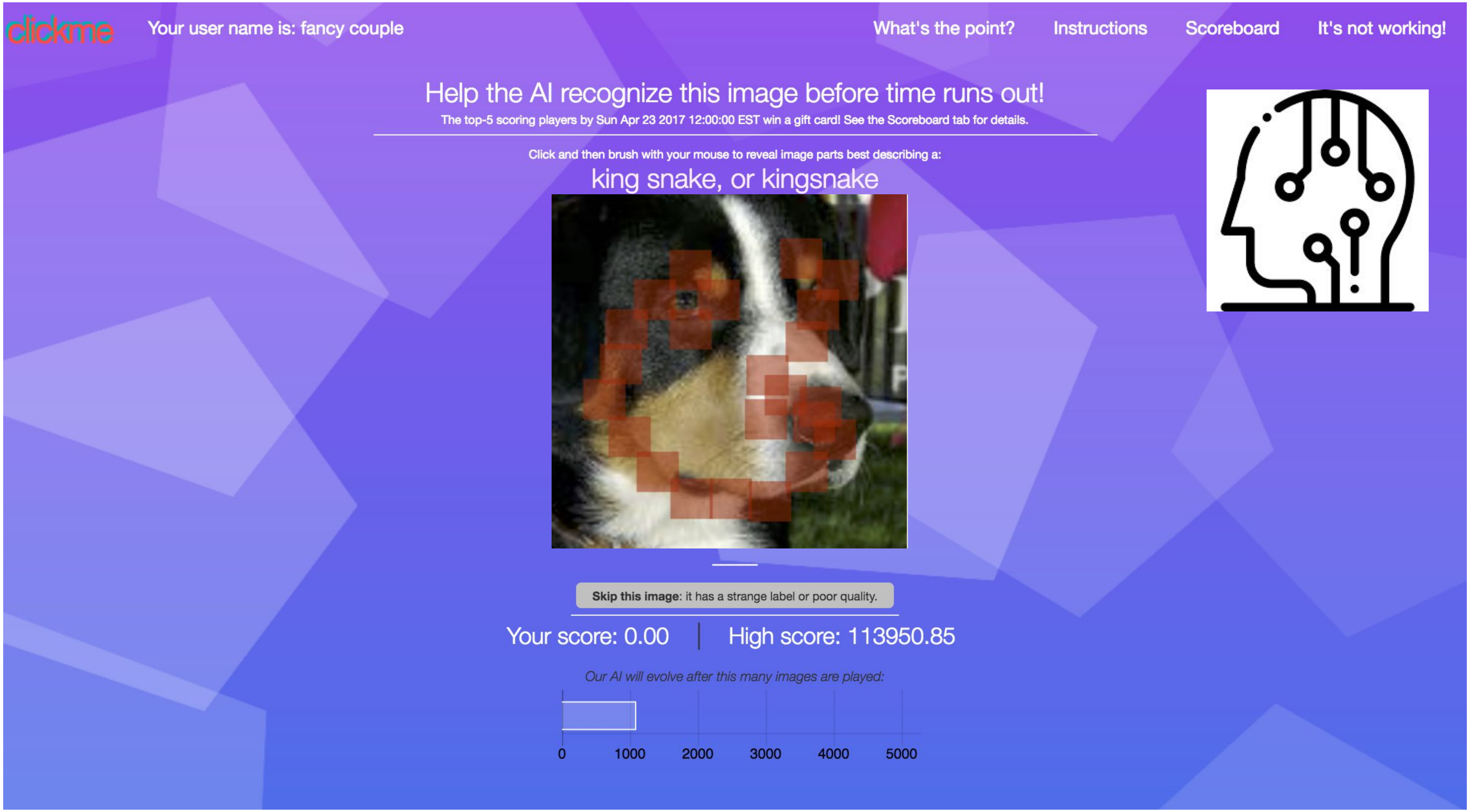


Objective 1: Object classification:
This is a truck

Objective 2: Learn attention mask
(feature-based and spatial)

Linsley Schiebler Eberhardt Serre ICLR 2019

ClickMe.ai v2

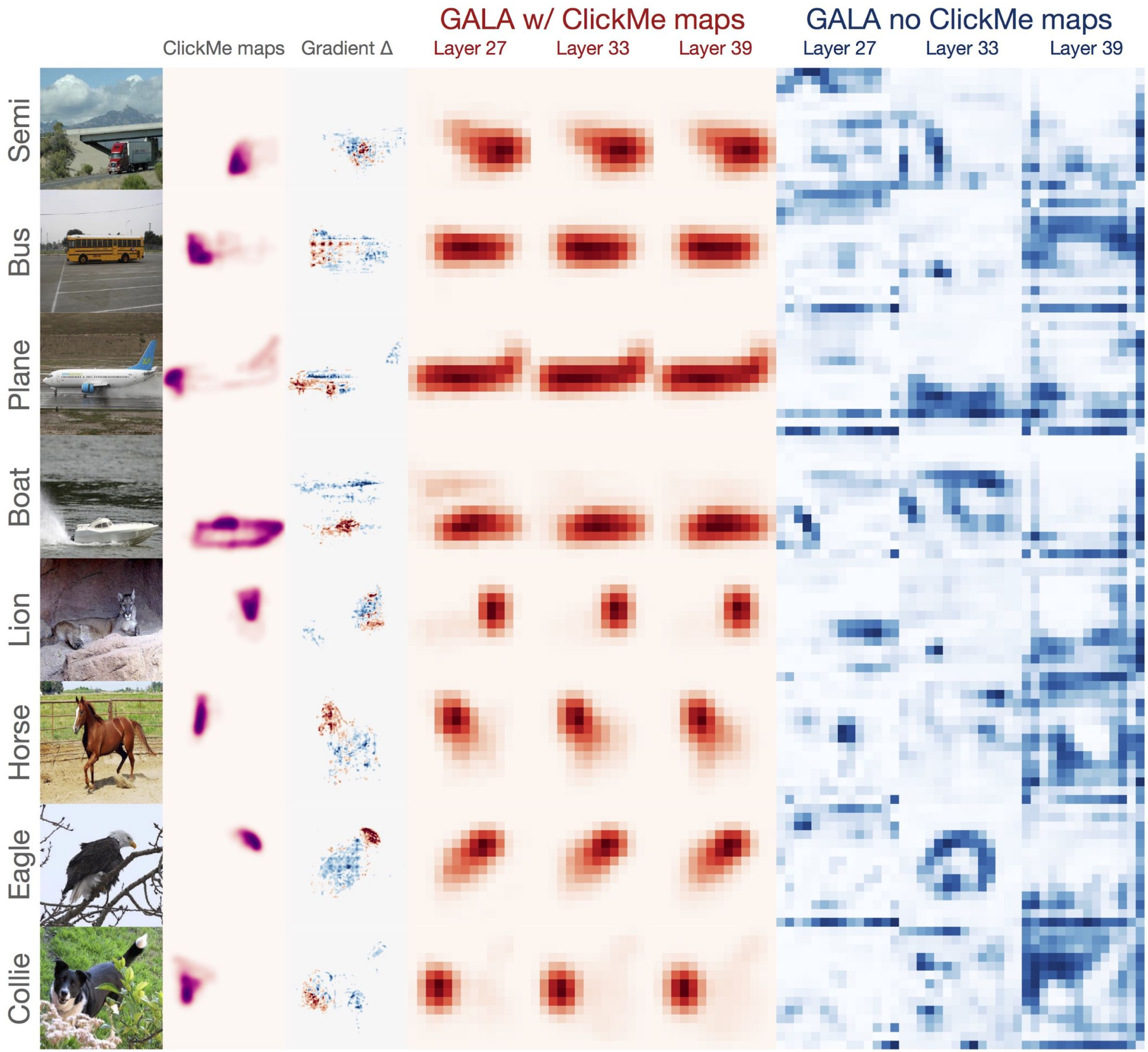


ClickMe.ai

Driving DCNs towards
human-like recognition

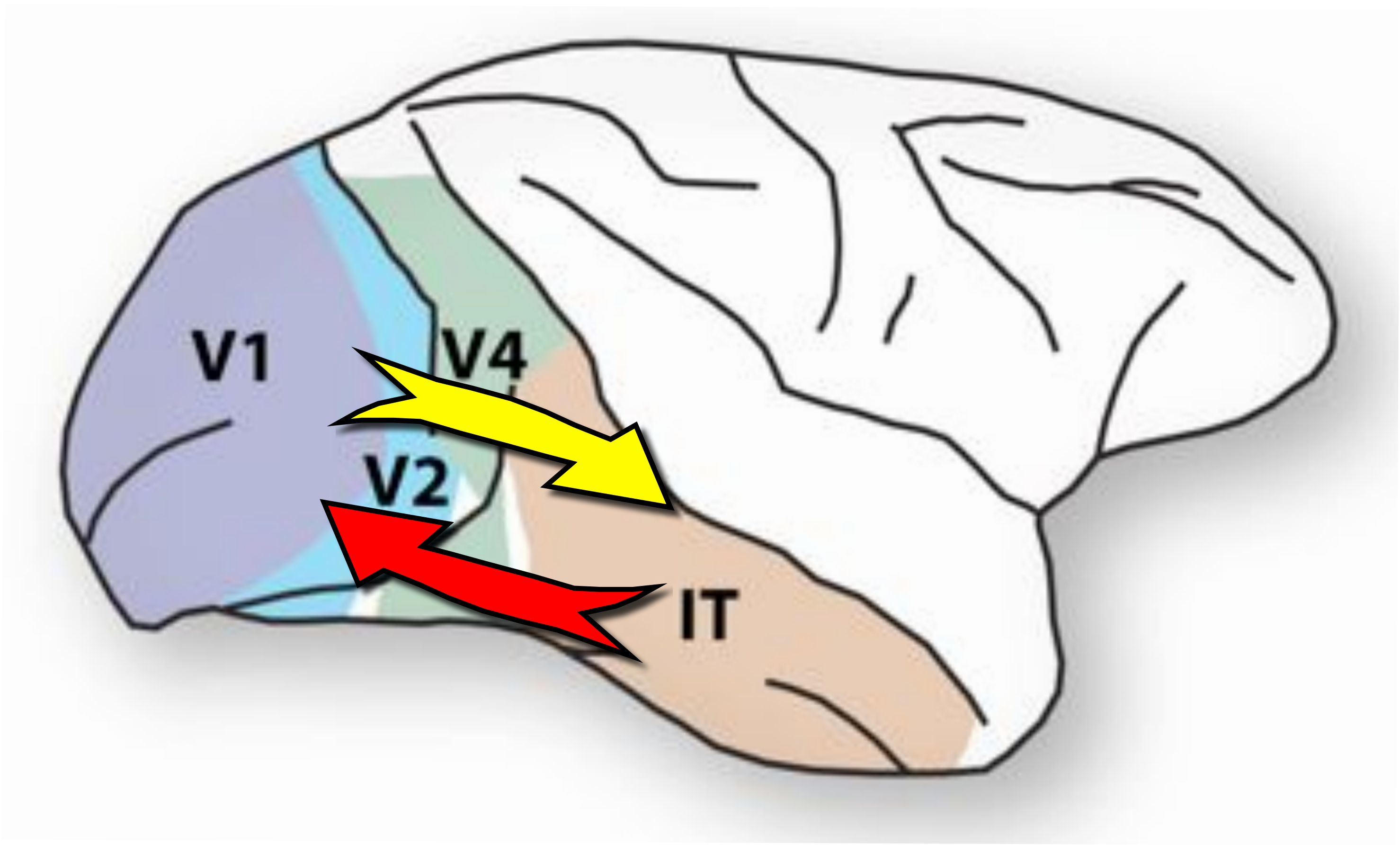
	top-1 err	top-5 err	maps
SE-ResNet-50	66.17	42.48	64.36**
ResNet-50	63.68	40.65	43.61
GALA-ResNet-50 no ClickMe	53.90	31.04	64.21**
GALA-ResNet-50 w/ ClickMe	**49.29**	**27.73**	**88.56****

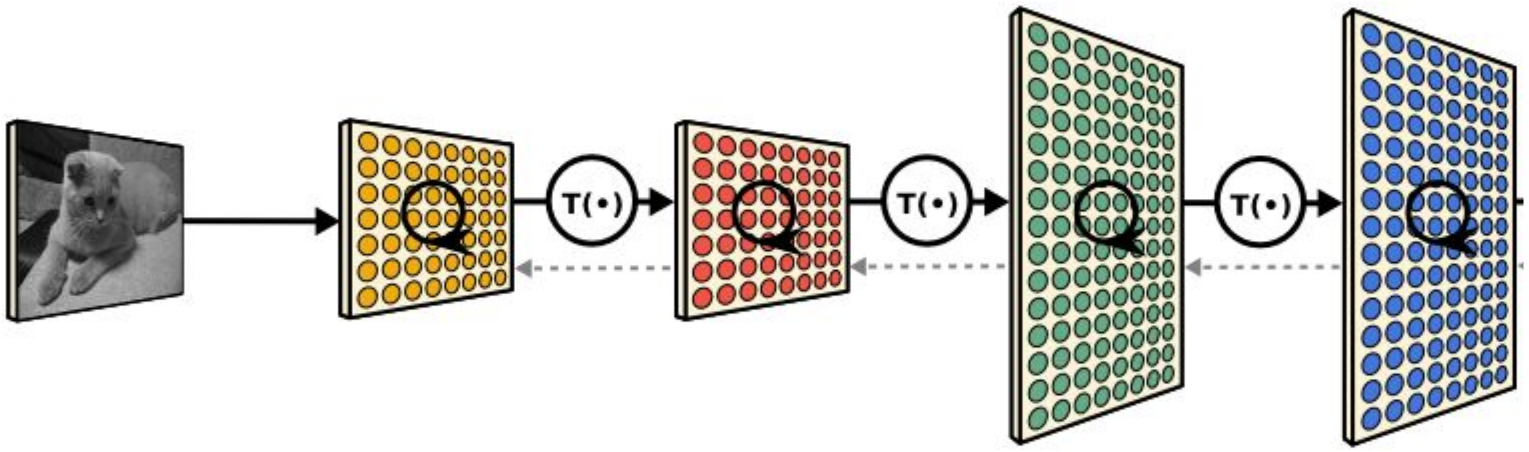
Linsley Schiebler Eberhardt Serre ICLR 2019



Linsley Schiebler Eberhardt Serre ICLR 2019

Feedback mechanisms

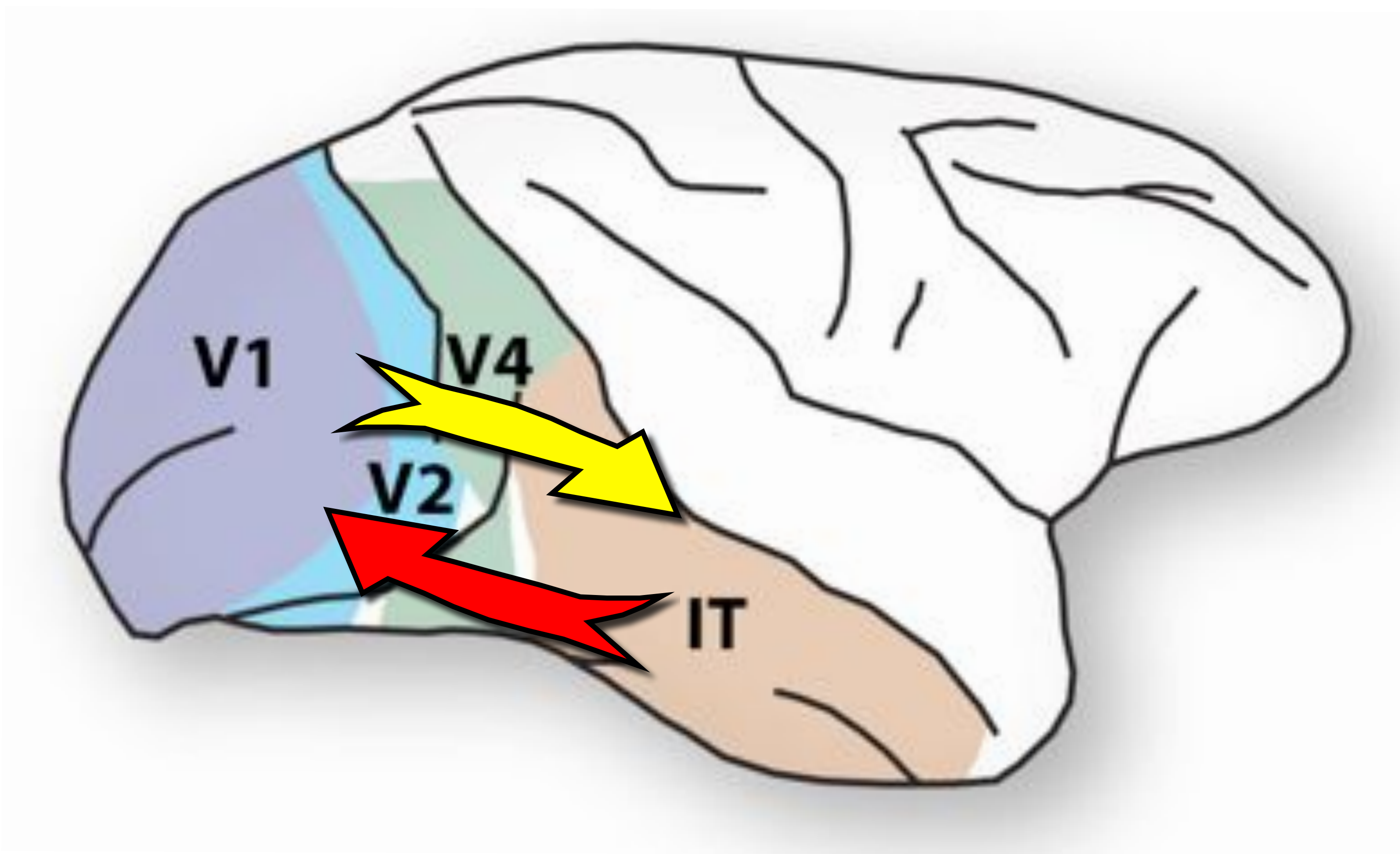


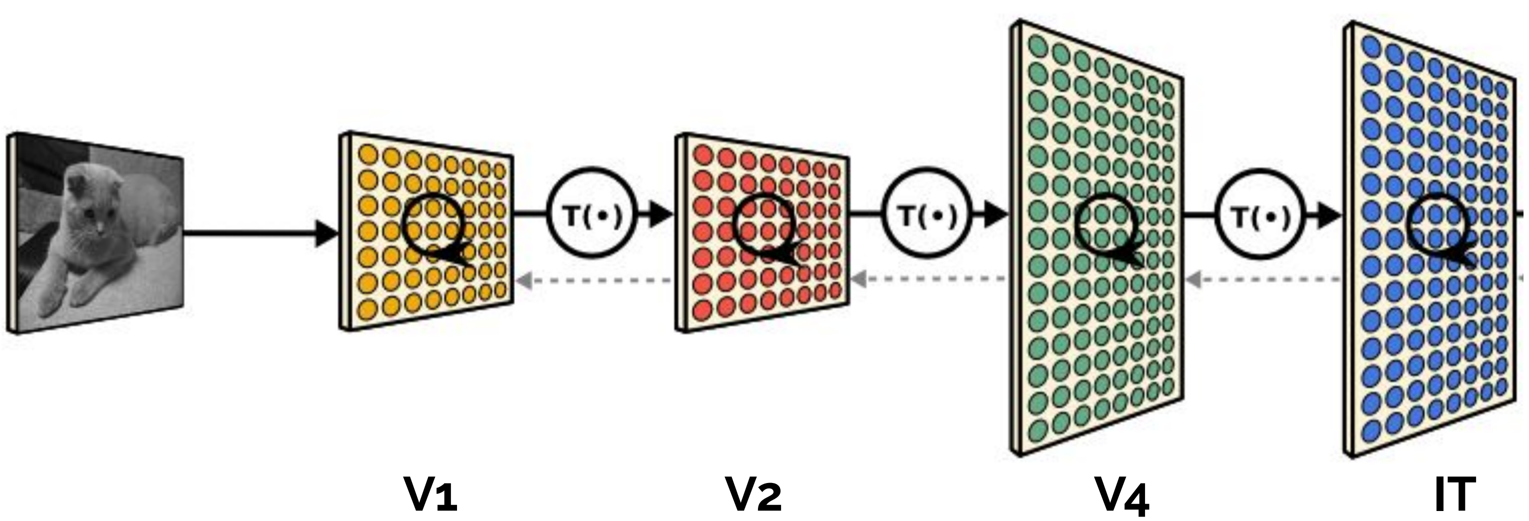


- Deep feedforward nets (CNNs) have become orders of magnitude deeper than our visual system
- Feedback plays key role in visual perception beyond feedforward sweep

Kreiman & Serre Year in Cognitive Neurosci. 2020

Feedback
mechanisms



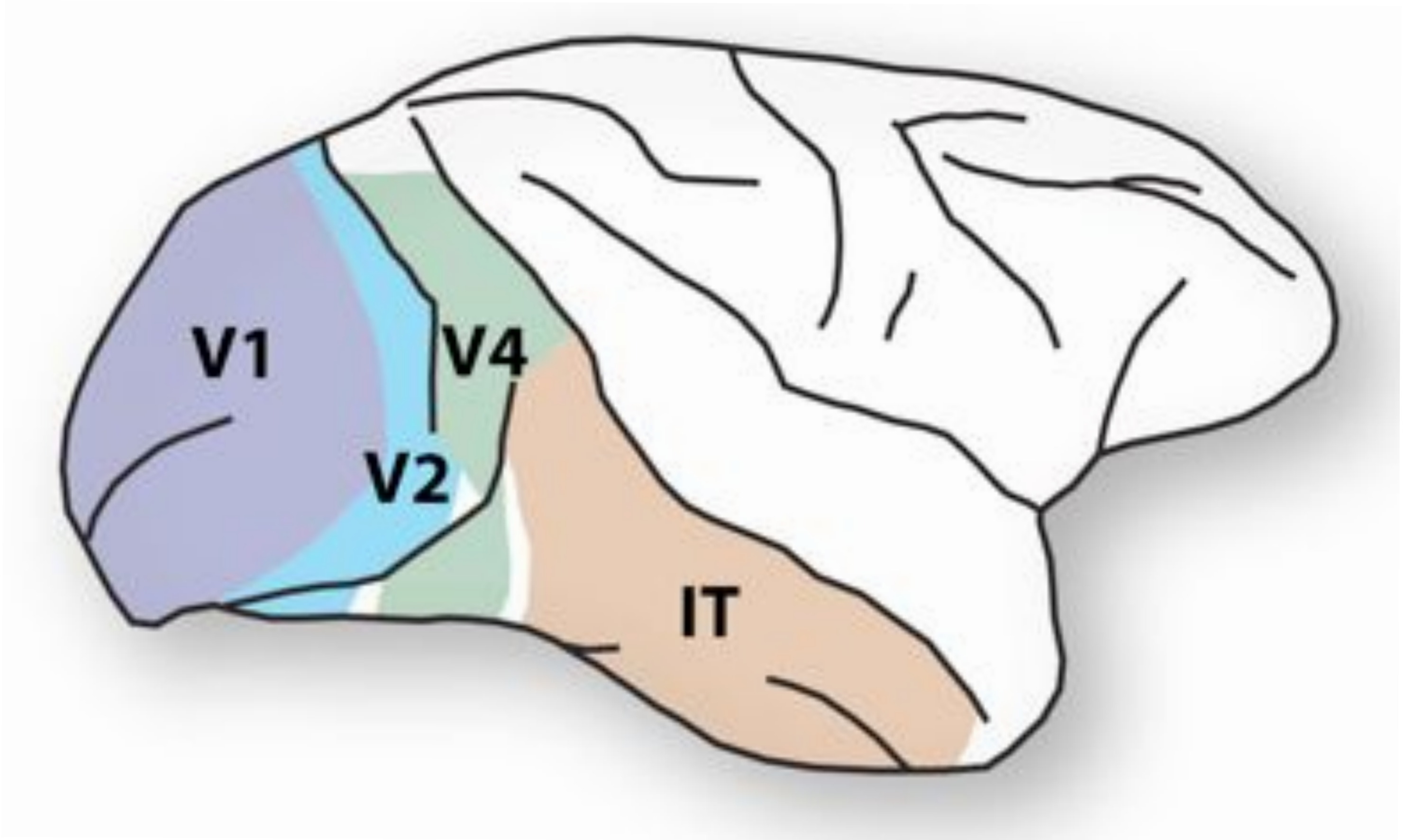


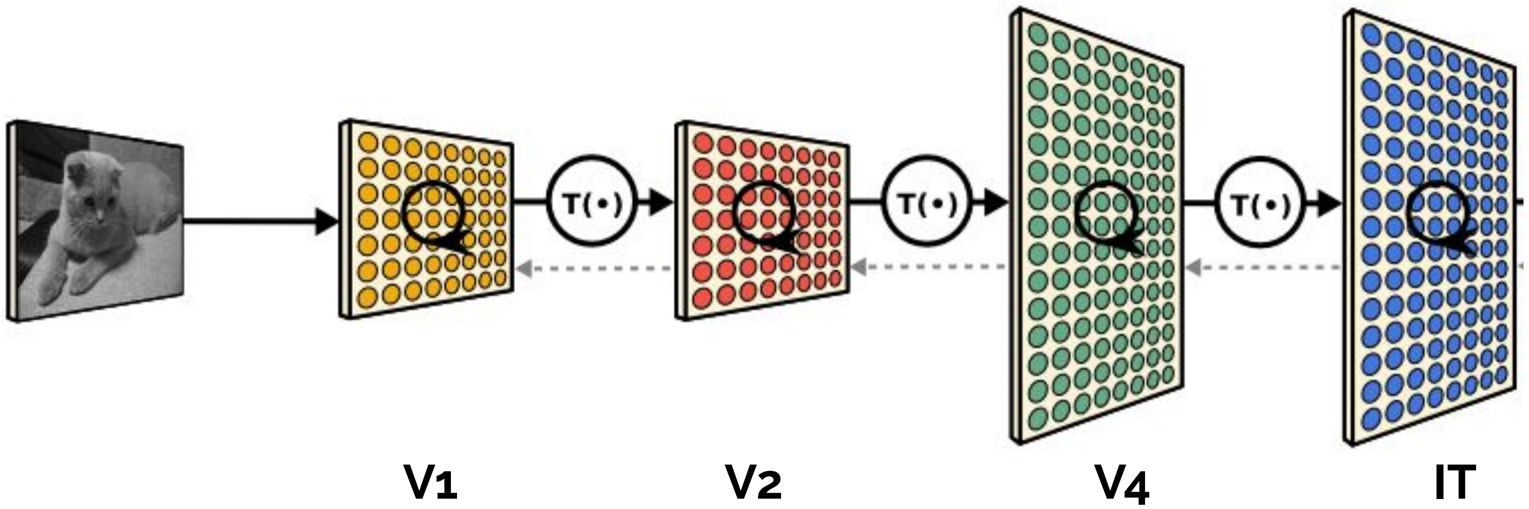
- Visual processing builds "depth" dynamically through recurrent connections

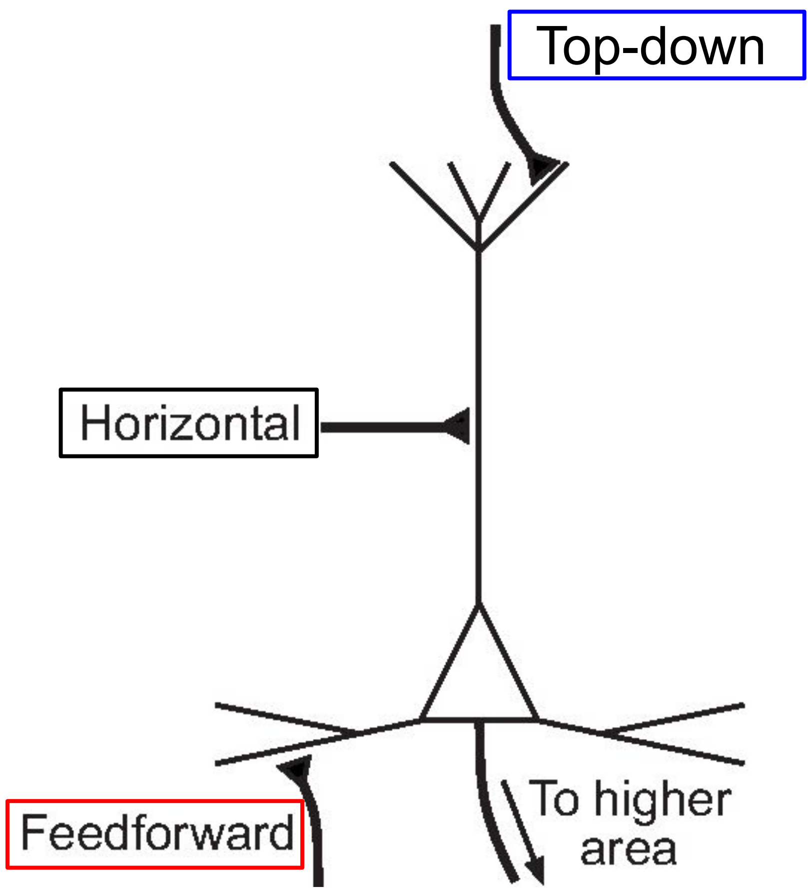
Kreiman & Serre Year in Cognitive Neurosci. 2020

Feedback mechanisms

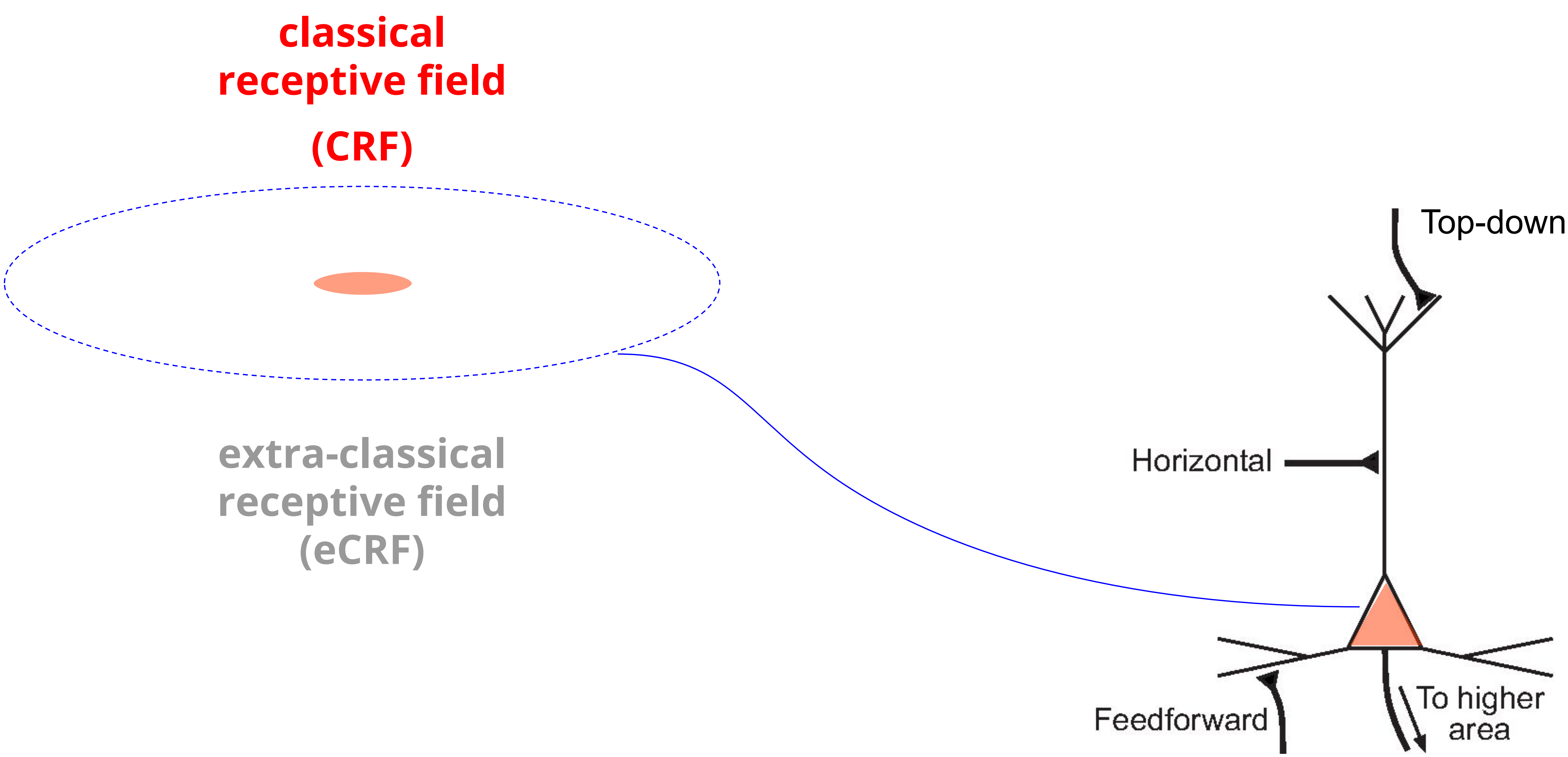
- Visual processing builds “depth” dynamically through recurrent connections



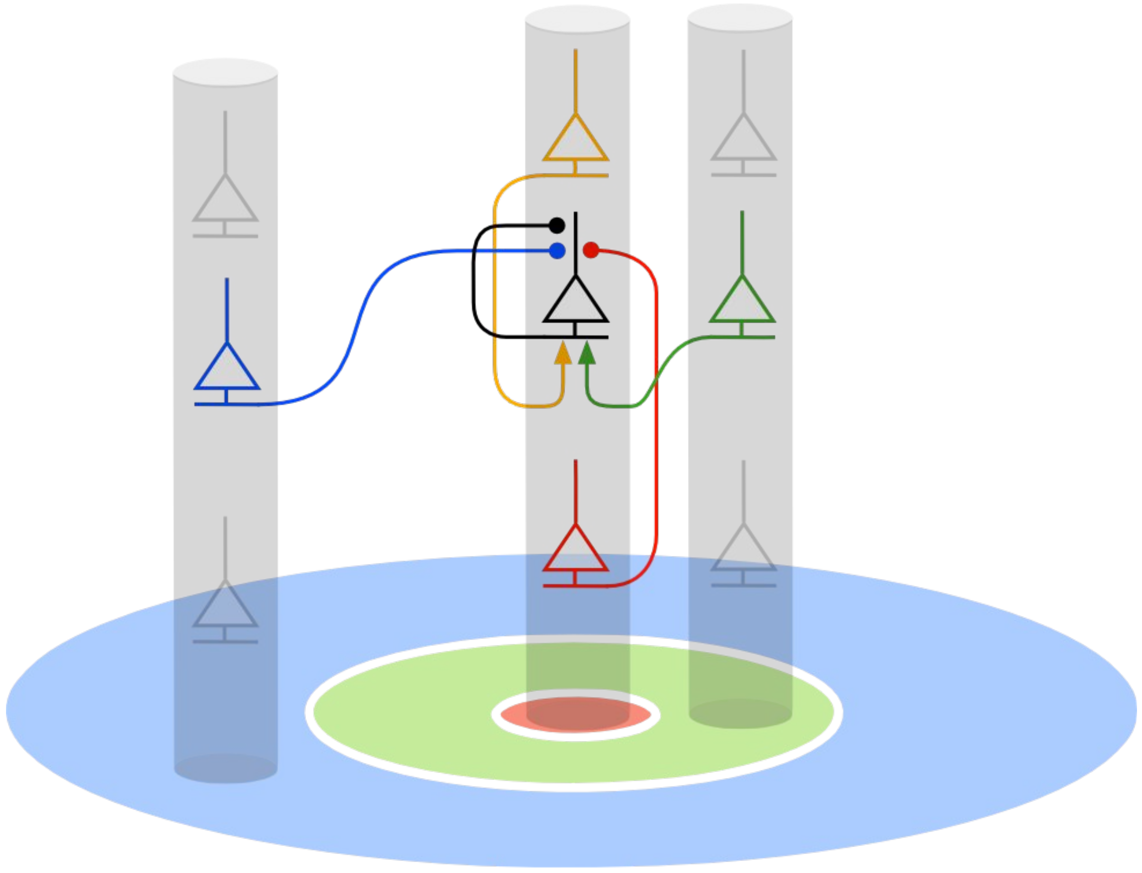




Feedback mechanisms



Recurrent neural circuit model



$$C^{(1)}_{xyk} = (\mathbf{W}^I * \mathbf{H}^{(2)})_{xyk}$$

$$C^{(2)}_{xyk} = (\mathbf{W}^E * \mathbf{H}^{(1)})_{xyk}$$

$$\eta \dot{H}^{(1)}_{xyk} + \epsilon^2 H^{(1)}_{xyk} = \left[\xi X_{xyk} - \left(\alpha H^{(1)}_{xyk} + \mu\right) C^{(1)}_{xyk}\right]_+$$

$$\tau \dot{H}^{(2)}_{xyk} + \sigma^2 H^{(2)}_{xyk} = \left[\gamma C^{(2)}_{xyk}\right]_+ .$$

Mely Linsley & Serre Psych Review 2018

CRF and eCRF
mechanisms

cross-orientation normalization

Busse, Wade & Carandini (2009)

surround suppression (multi-modality)

DeAngelis, Freeman & Ohzawa (1994)
Cavanaugh, Bair & Movshon (2002)
Bradley & Andersen (1998)
Solomon, Peirce & Lennie (2003)

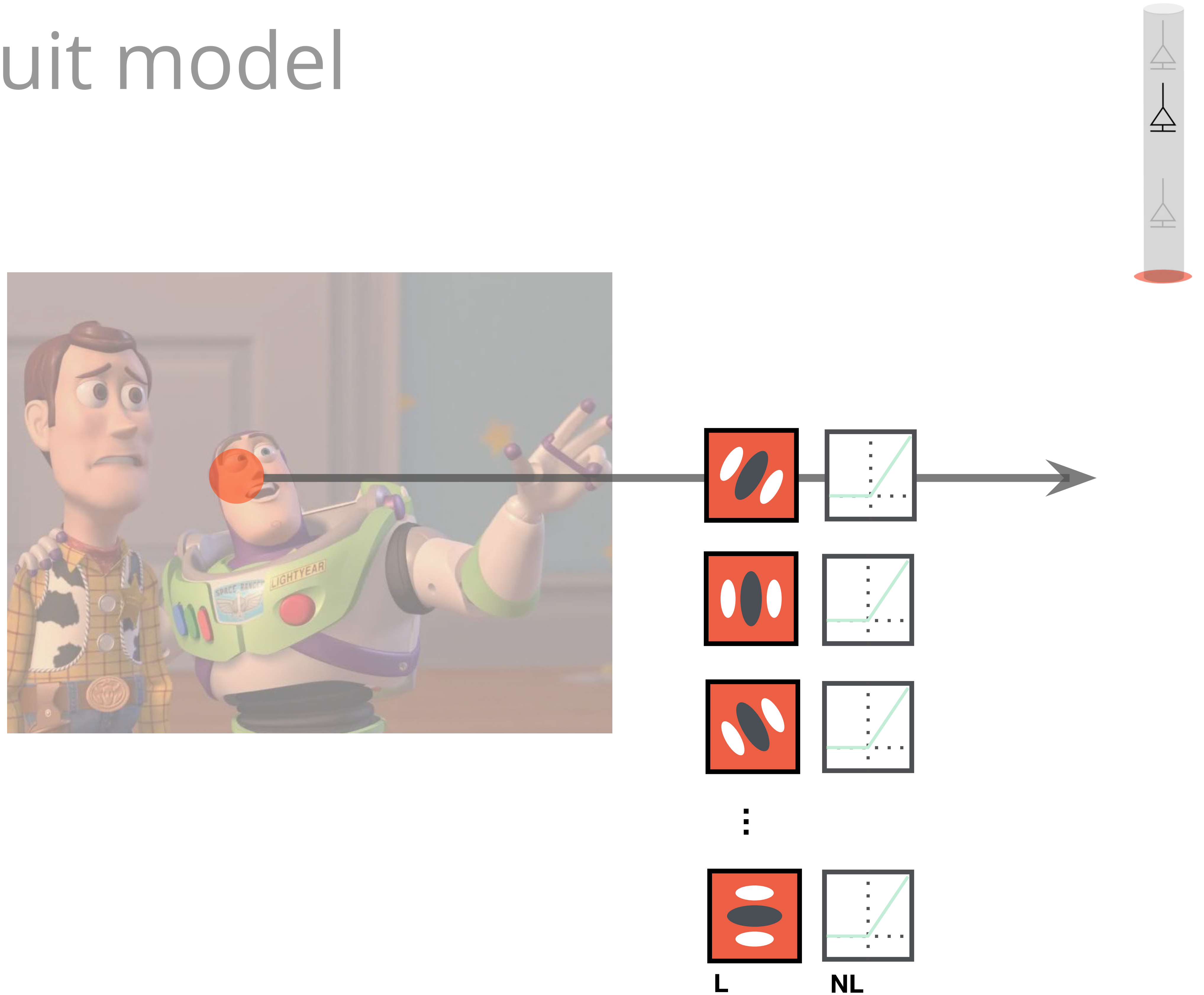
feature-selective surround suppression

Trott & Born (2015)

size tuning

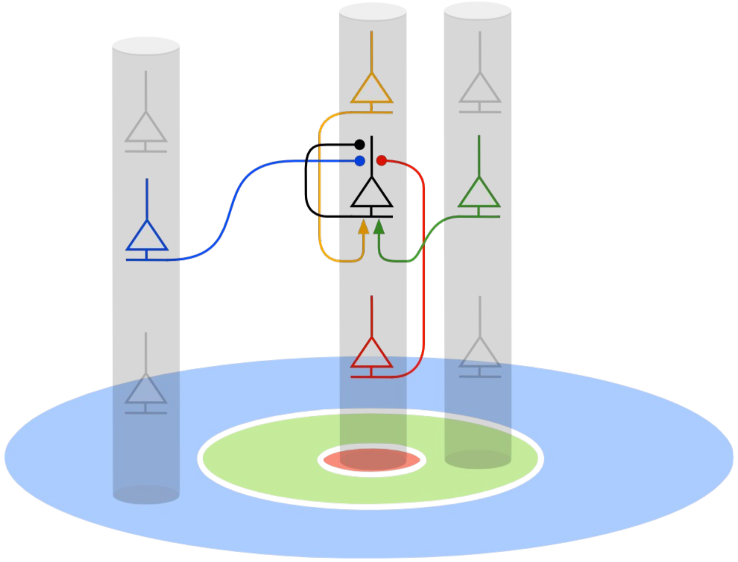
Xing, Shapley, Hawken, Ringach (2005)

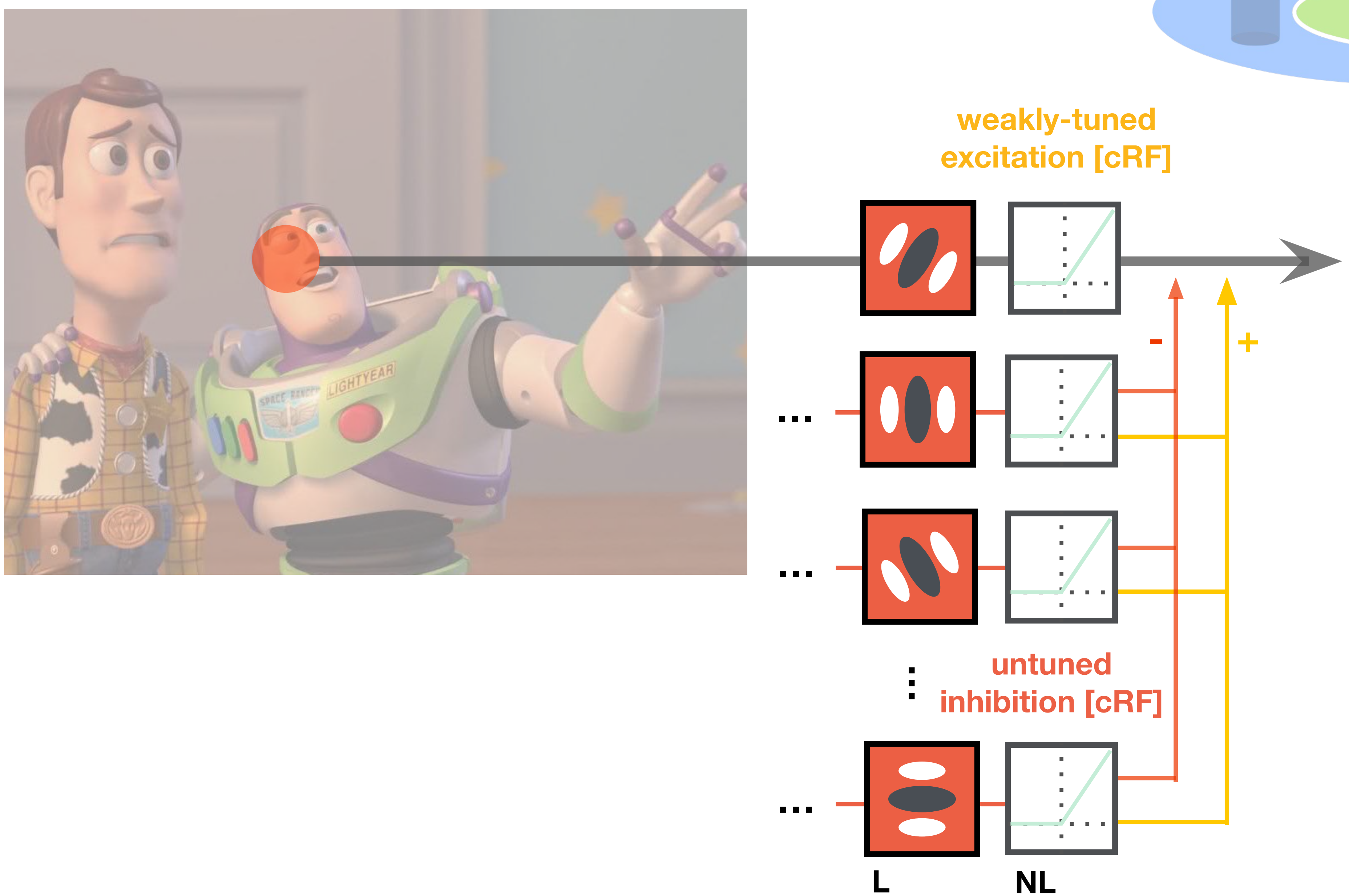
Recurrent neural
circuit model



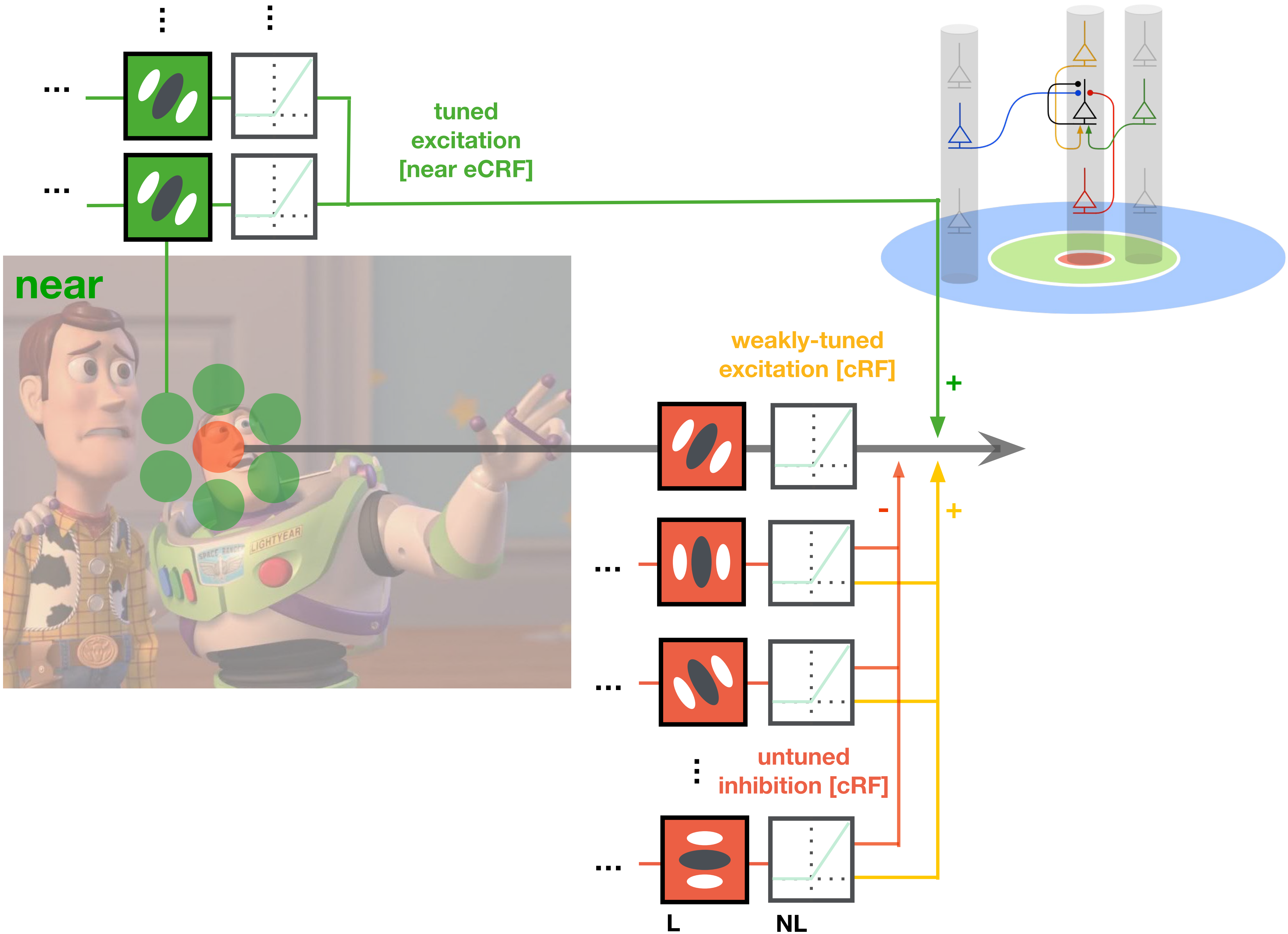
Mely Linsley & Serre Psych Review 2018

Recurrent neural
circuit model

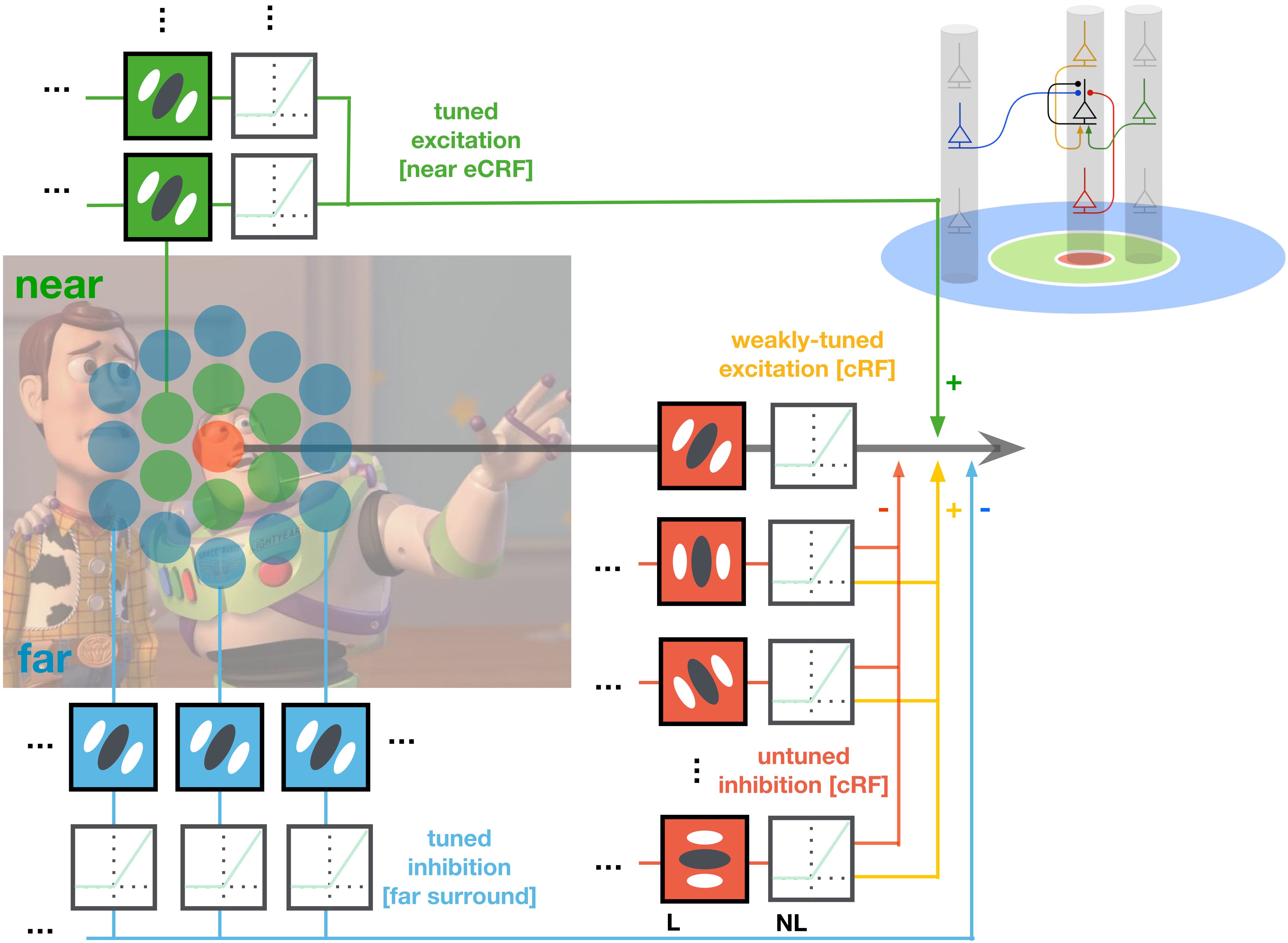




Mely Linsley & Serre Psych Review 2018



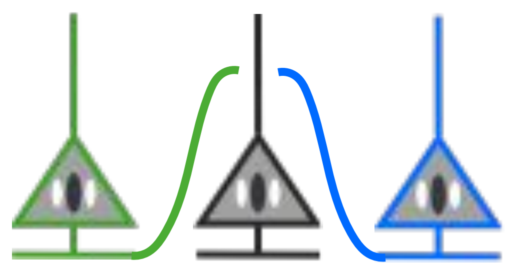
Mely Linsley & Serre Psych Review 2018


tuned
excitation
[near eCRF]
near
far
weakly-tuned
excitation [cRF]
+
-
+
-
untuned
inhibition [cRF]
tuned
inhibition
[far surround]
L
NL

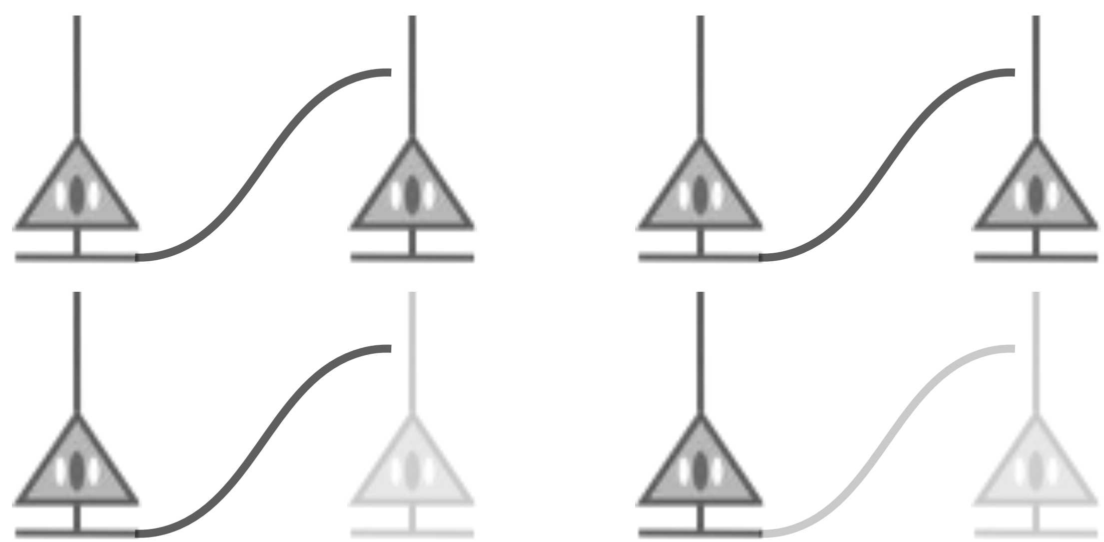
Recurrent
network model

2 opponent surround regions
near = excitation
far = inhibition

tuned connections from surround

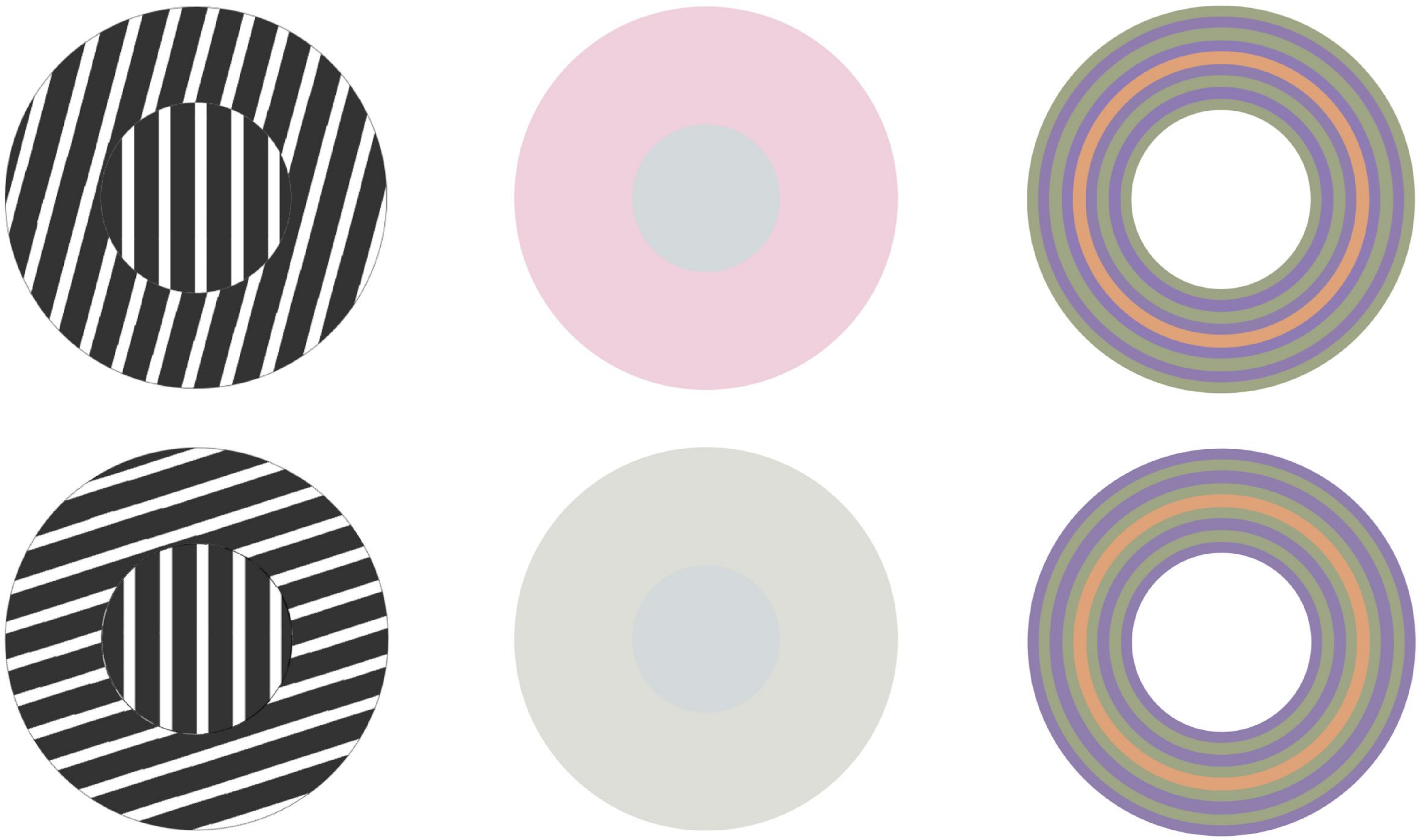


asymmetry of excitation and inhibition



Mely Linsley & Serre Psych Review 2018

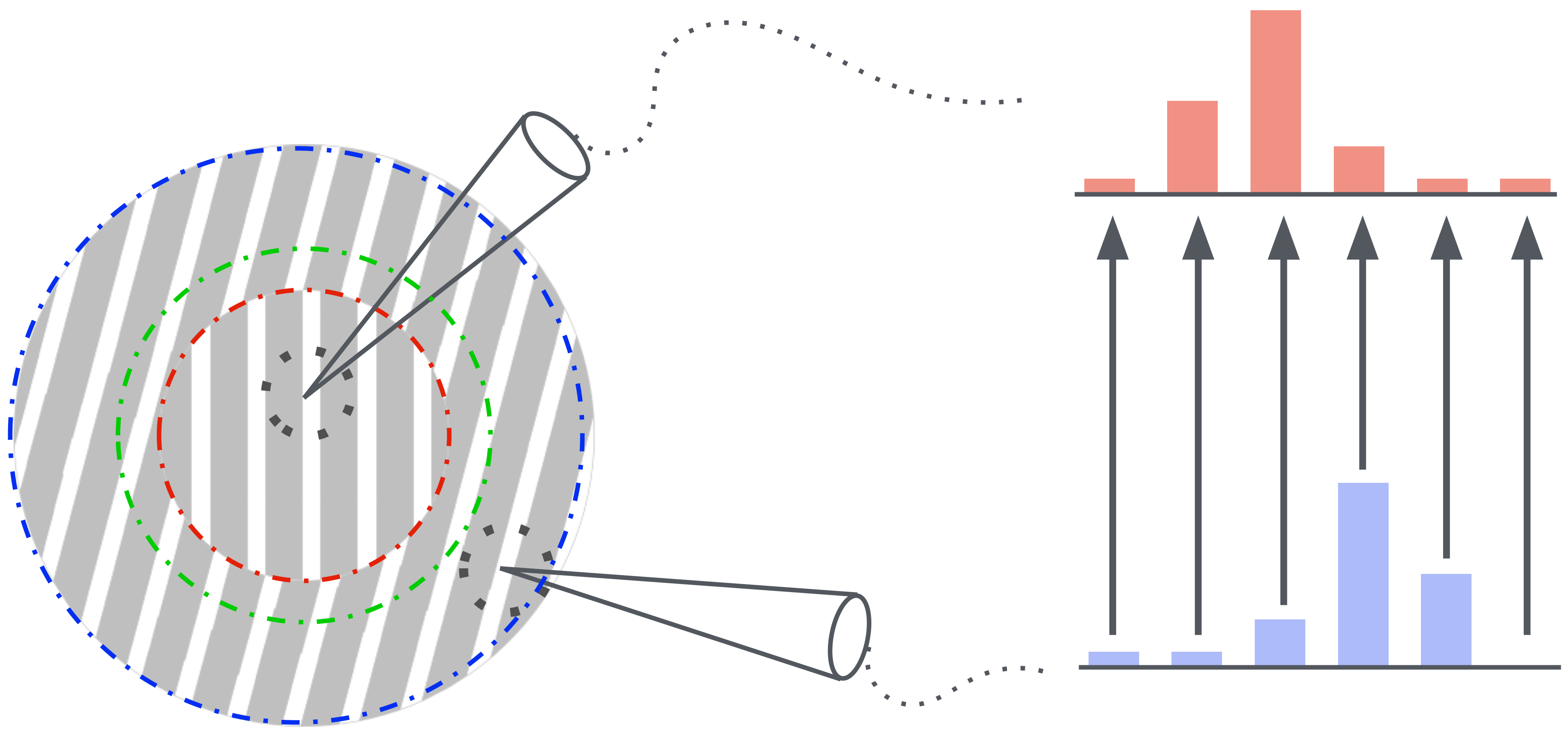
Recurrent neural
circuit model



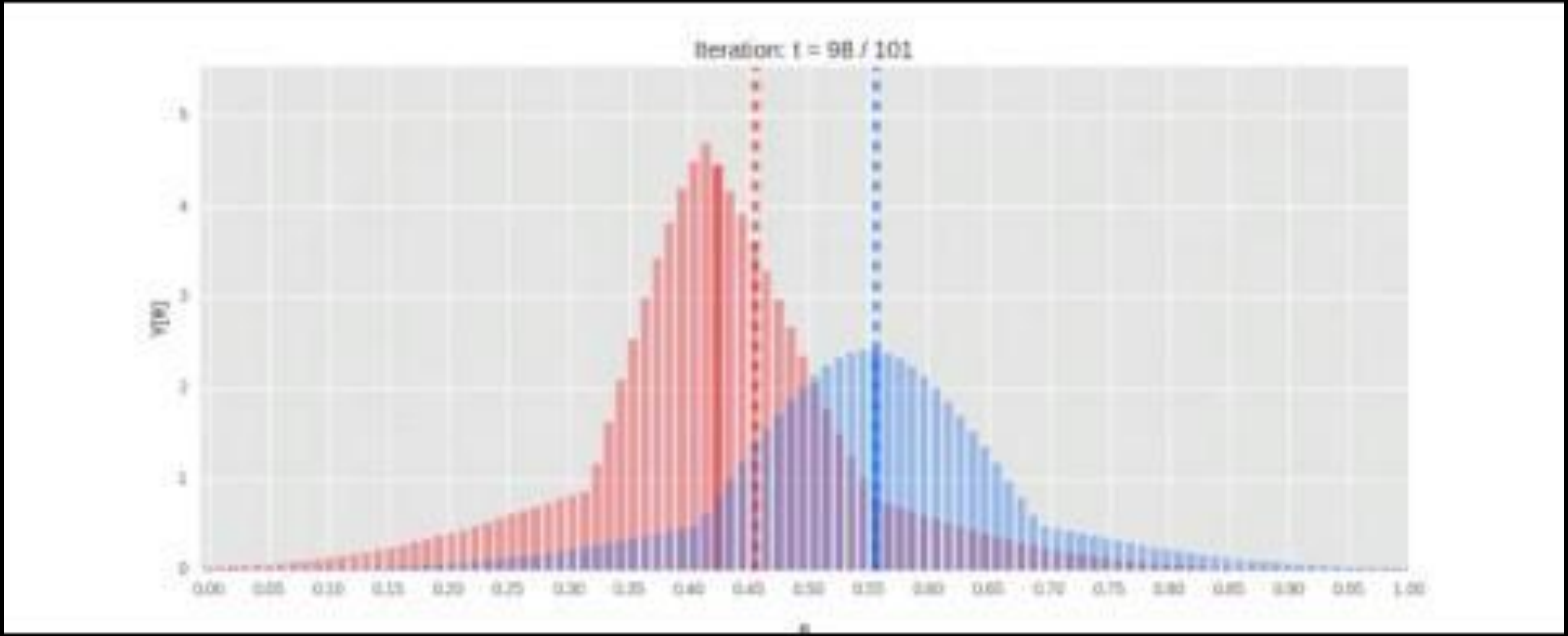
Mely Linsley & Serre Psych Review 2018

Recurrent neural circuit model

e.g., repulsion in tilt illusion

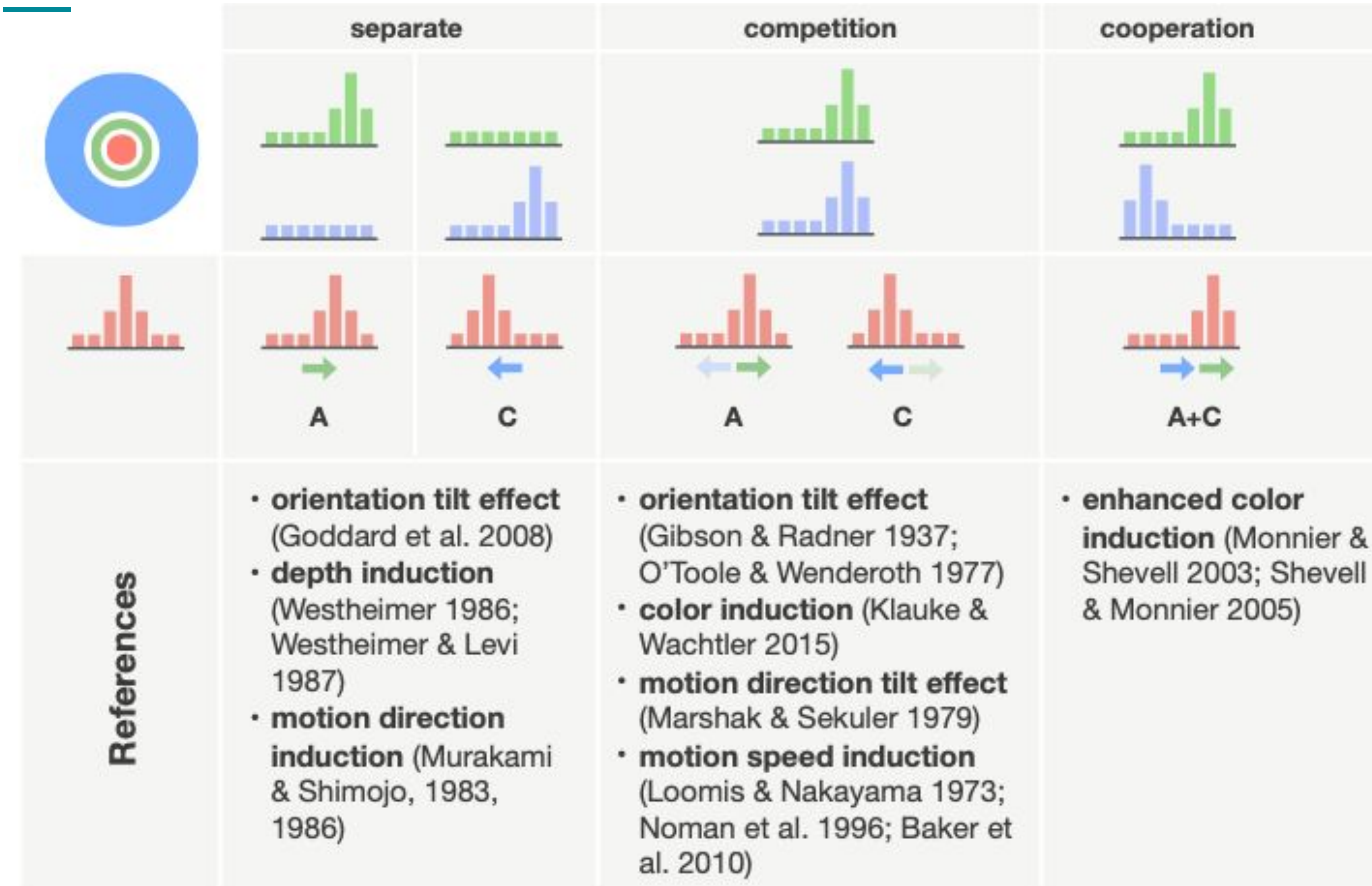


Neurophysiology: Gilbert & Wiesel 1990; Thomas et al 2002; Kusunoki et al 2006; Li et al 1999
Modeling: Grossberg et al 2011



Iteration: t = 98 / 101

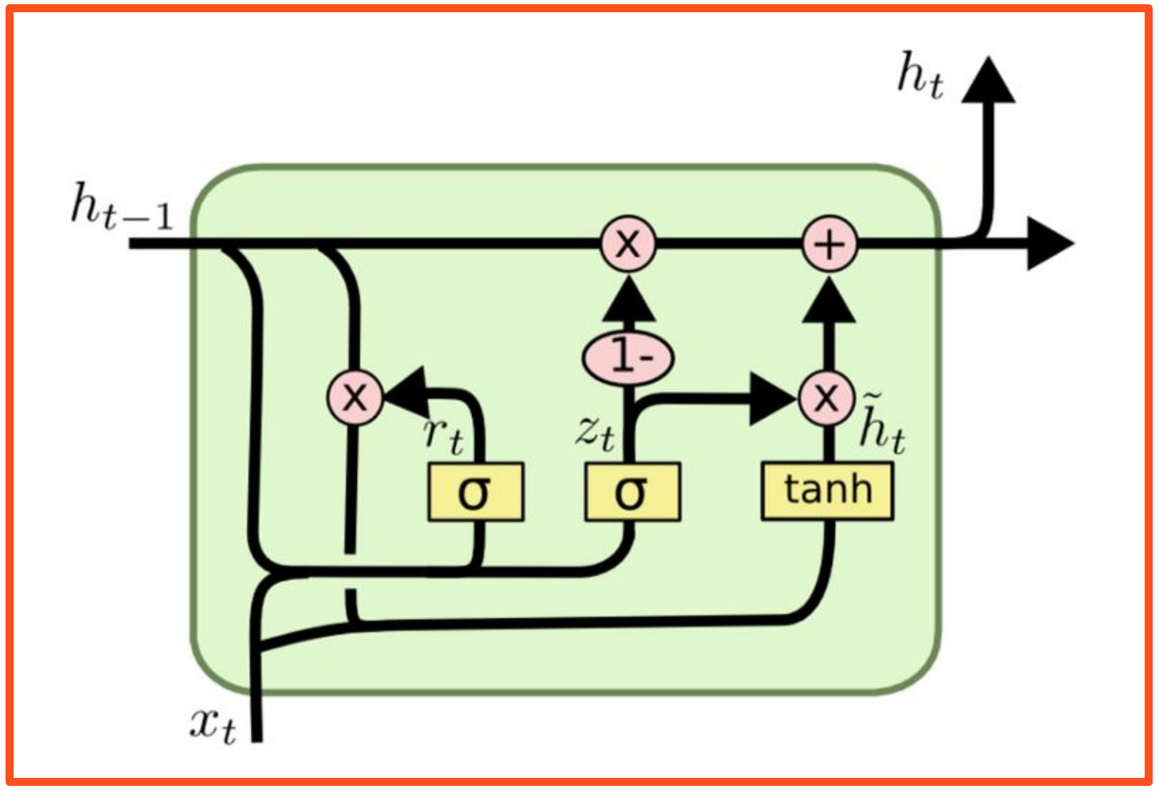
A new taxonomy
of contextual phenomena

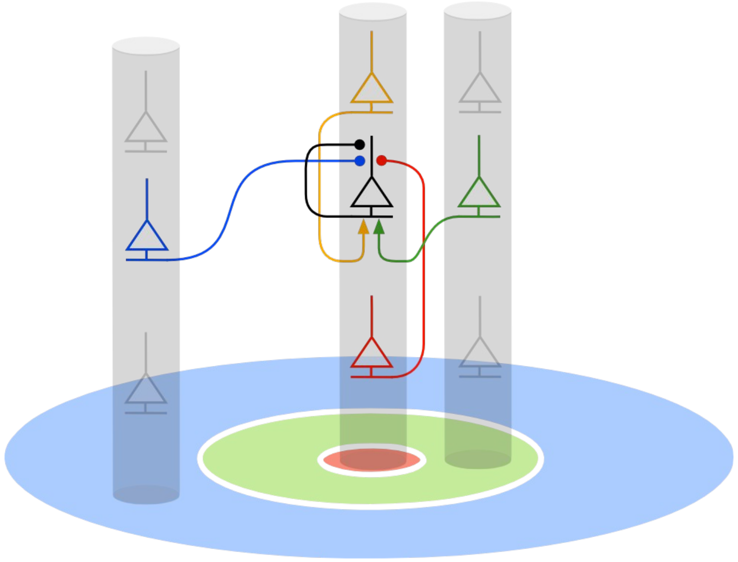


Mely Linsley & Serre Psych Review 2018

From a neural circuit...

... to an ML module





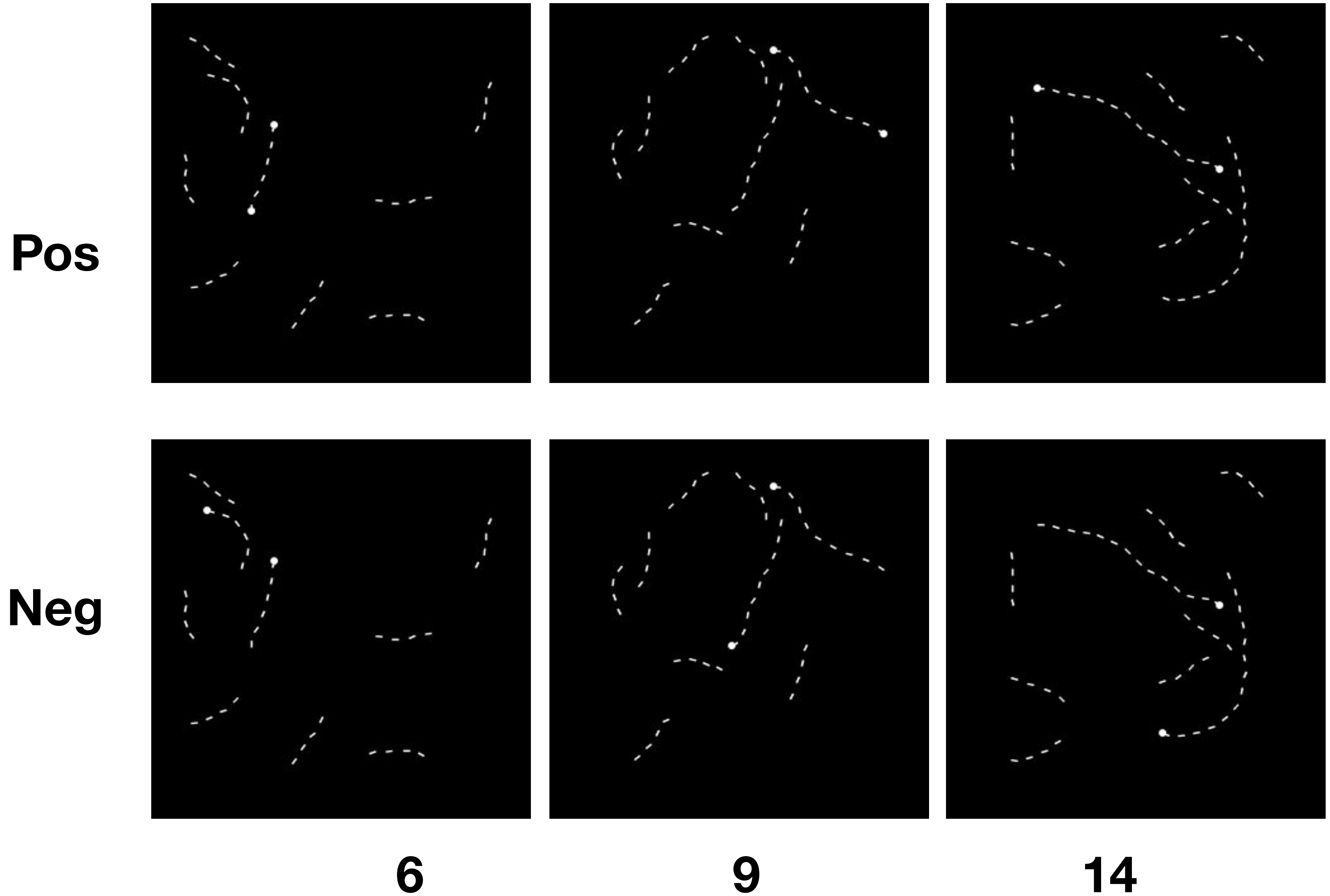
$$C^{(1)}_{xyk} = (\mathbf{W}^I * \mathbf{H}^{(2)})_{xyk}$$

$$C^{(2)}_{xyk} = (\mathbf{W}^E * \mathbf{H}^{(1)})_{xyk}$$

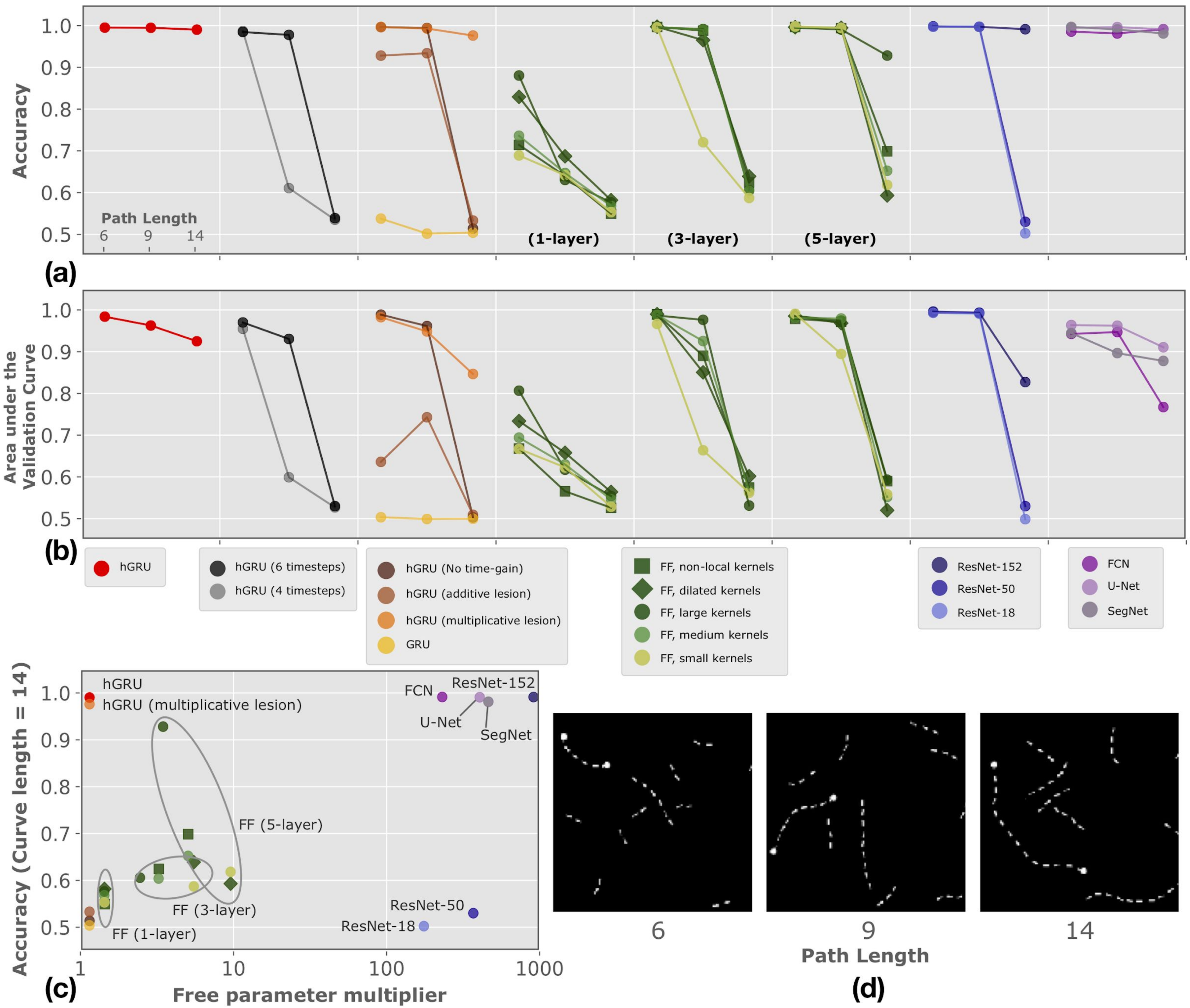
$$\eta \dot{H}^{(1)}_{xyk} + \epsilon^2 H^{(1)}_{xyk} = \left[\xi X_{xyk} - (\alpha H^{(1)}_{xyk} + \mu)\, C^{(1)}_{xyk}\right]_+$$

$$\tau \dot{H}^{(2)}_{xyk} + \sigma^2 H^{(2)}_{xyk} = \left[\gamma C^{(2)}_{xyk}\right]_+ .$$

Horizontal GRU
(hGRU)

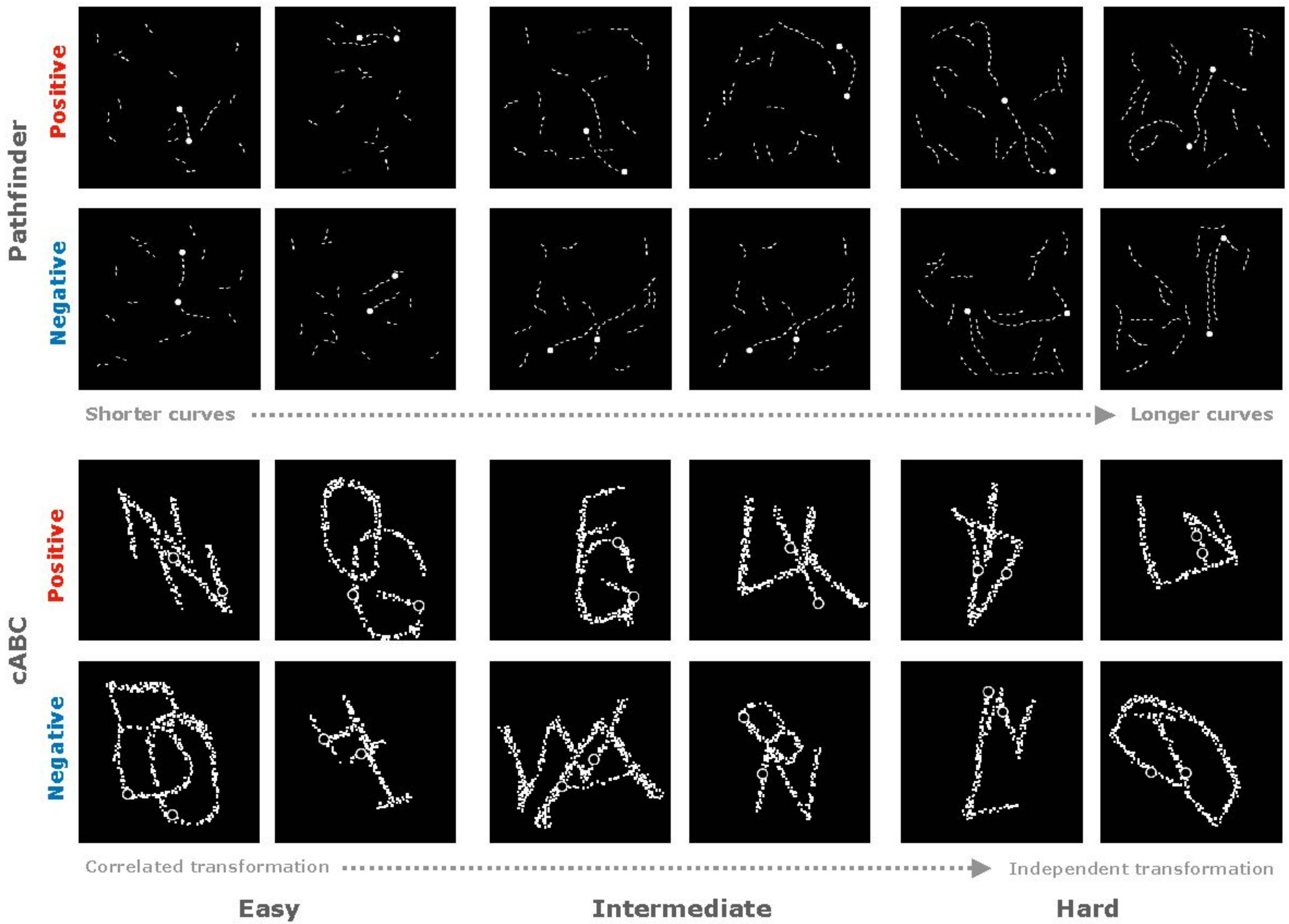


Linsley Kim Veerabadran Windolf & Serre NeurIPS 2018



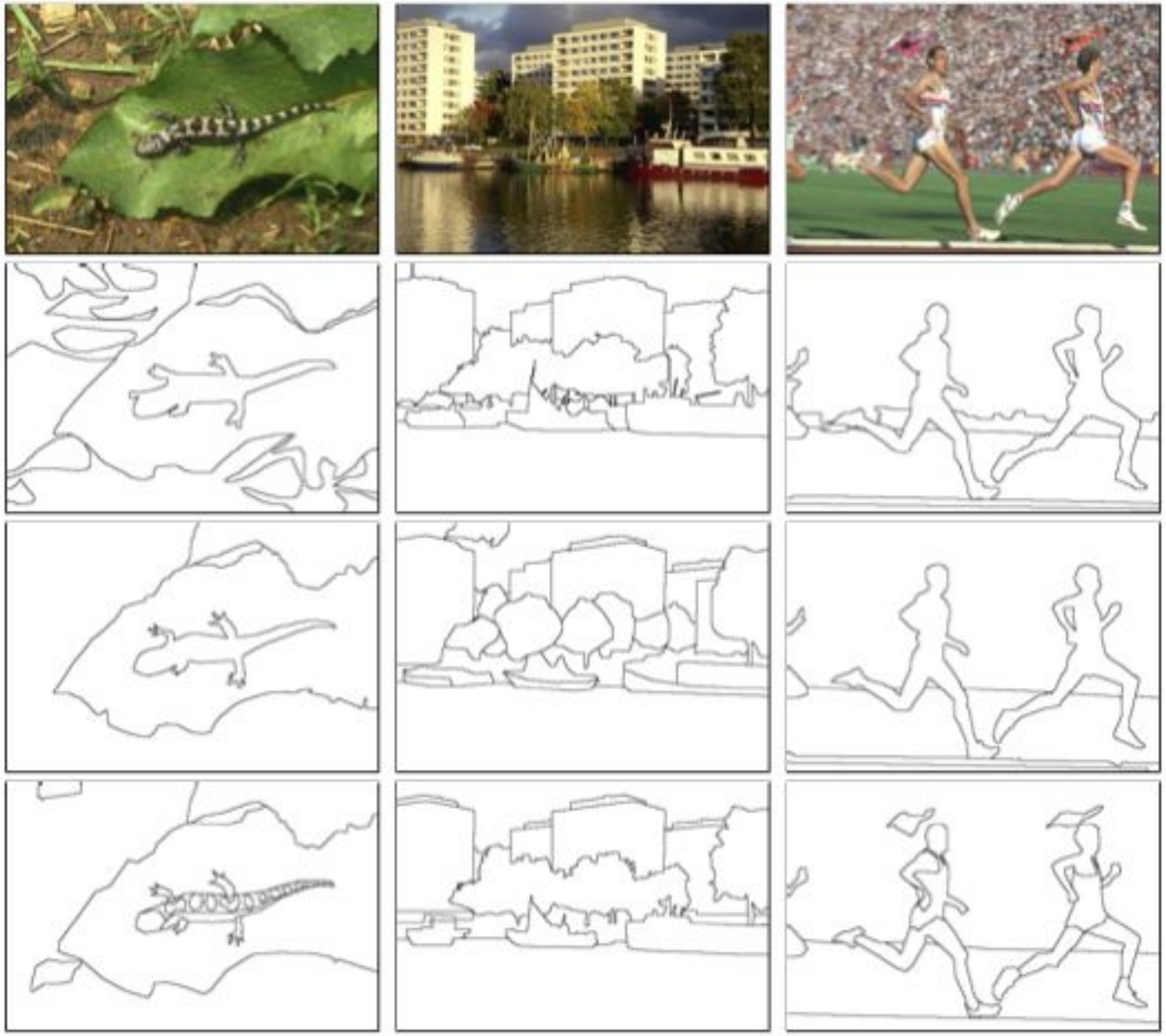
Linsley Kim Veerabadran Windolf & Serre NeurIPS 2018

Disentangling neural mechanisms
for perceptual grouping



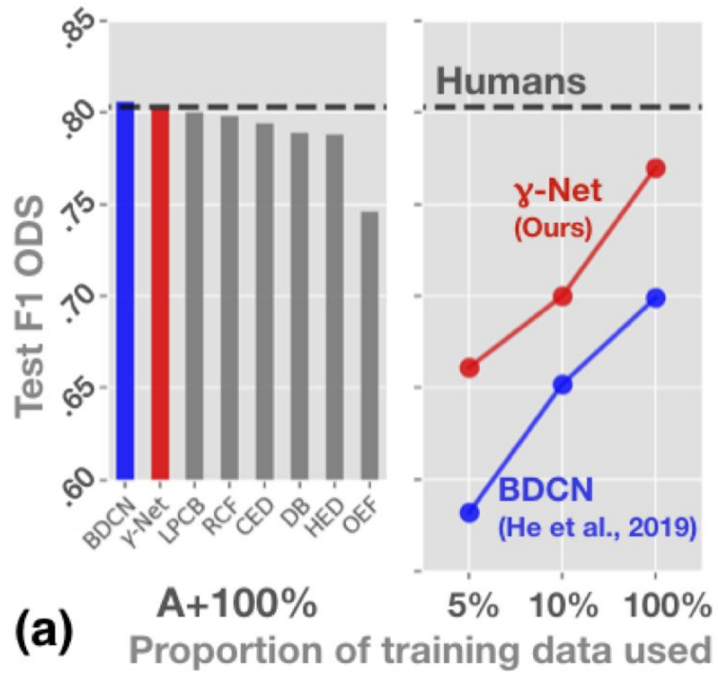
Kim Linsley* Thakkar & Serre ICLR 2020*

Berkeley image
segmentation dataset



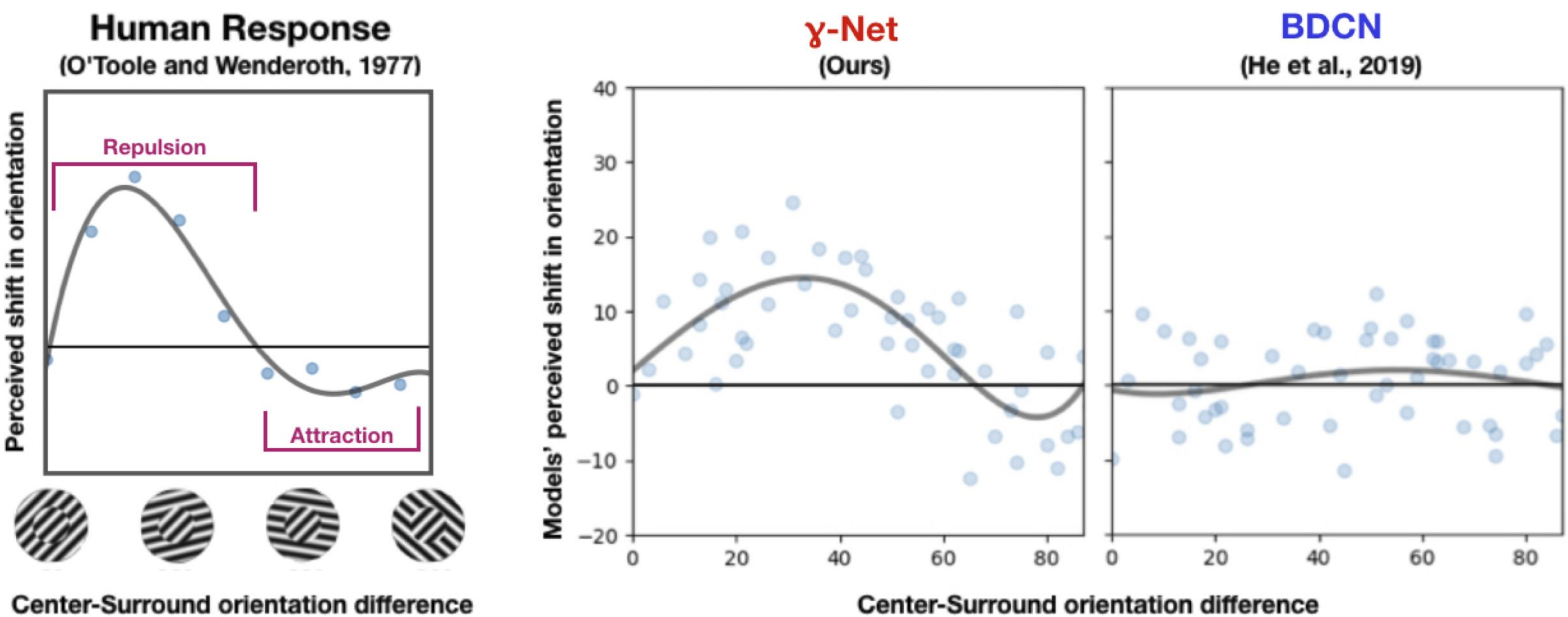
Martin Fowlkes Tal & Malik 2001

Recurrent neural circuits
for contours detection



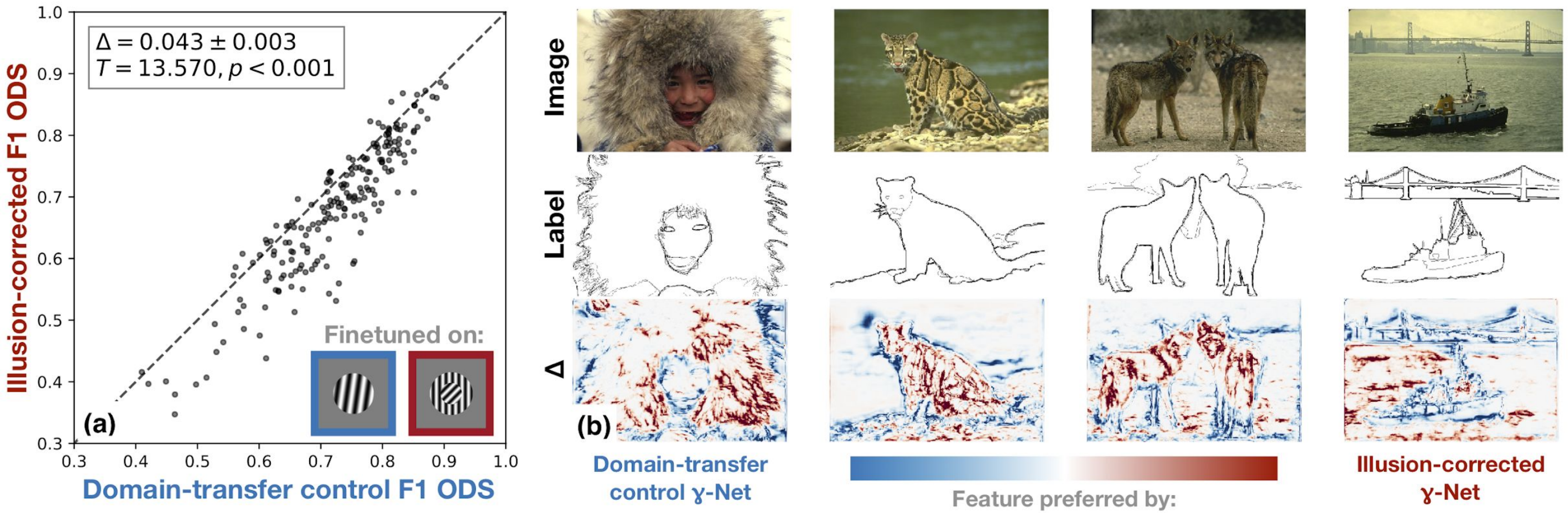
Linsley* Kim* Ashok & Serre ICLR 2020

Recurrent neural circuits
for contours detection



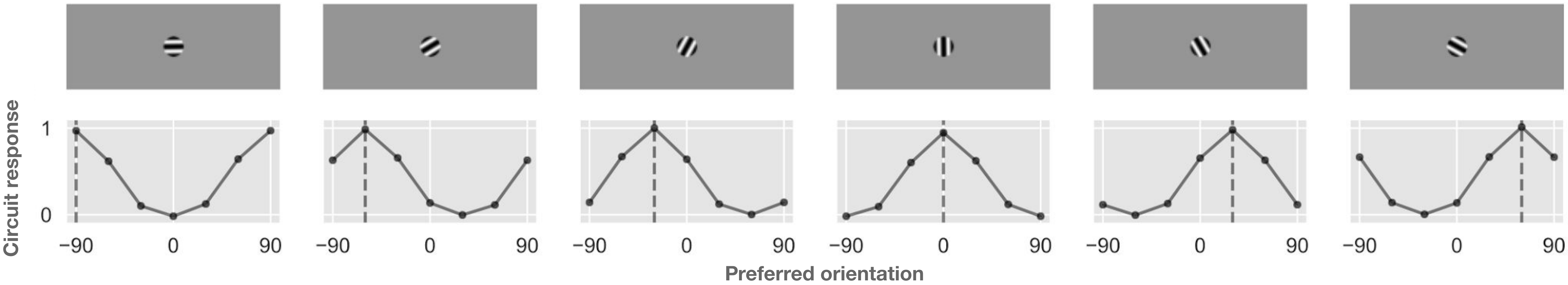
Linsley* Kim* Ashok & Serre ICLR 2020

Recurrent neural circuits
for contours detection

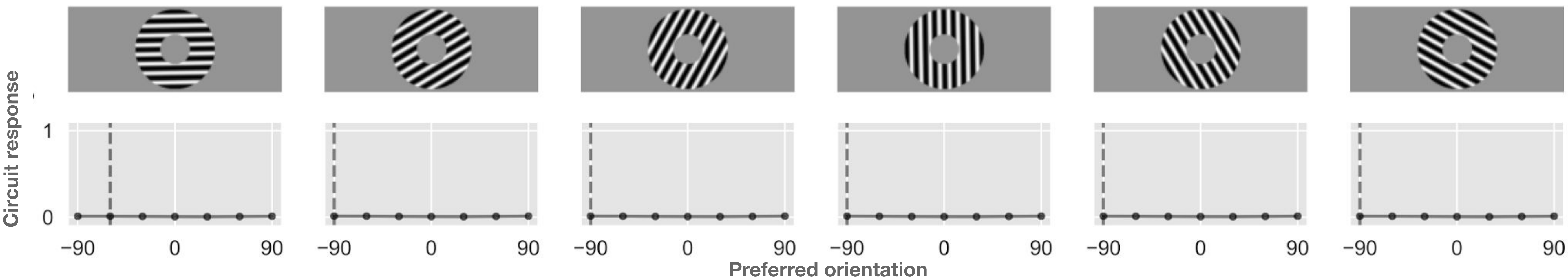


Linsley* Kim* Ashok & Serre ICLR 2020

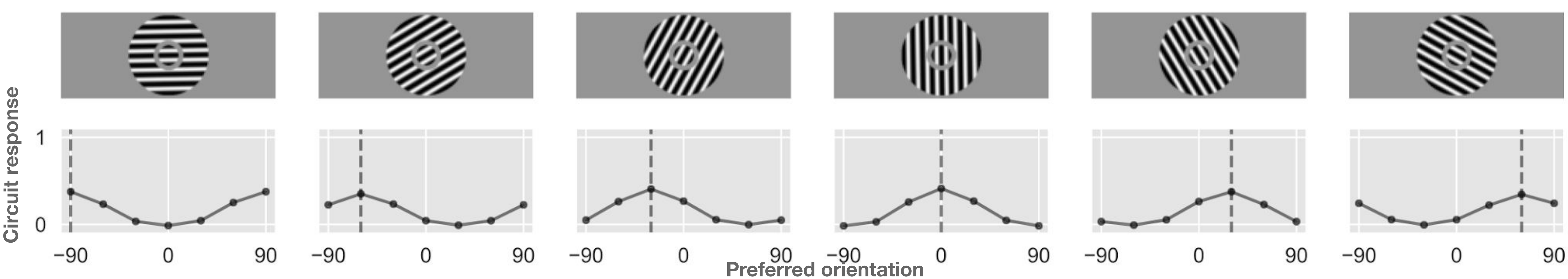
Model has learned orientation-tuned CRFs matched in size to V1



...and purely modulatory extra-classical receptive fields...

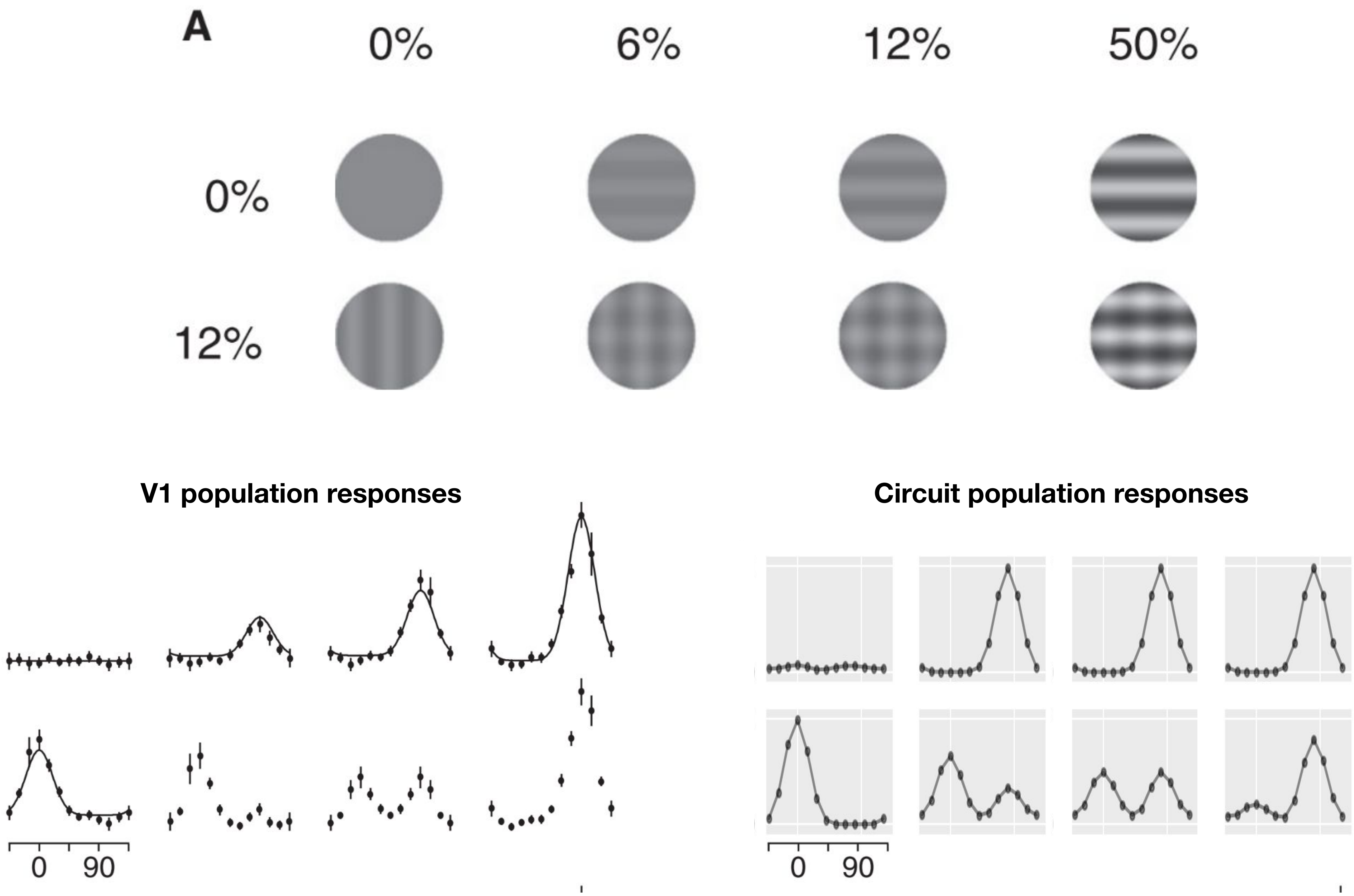


...which combine to contribute surround suppression

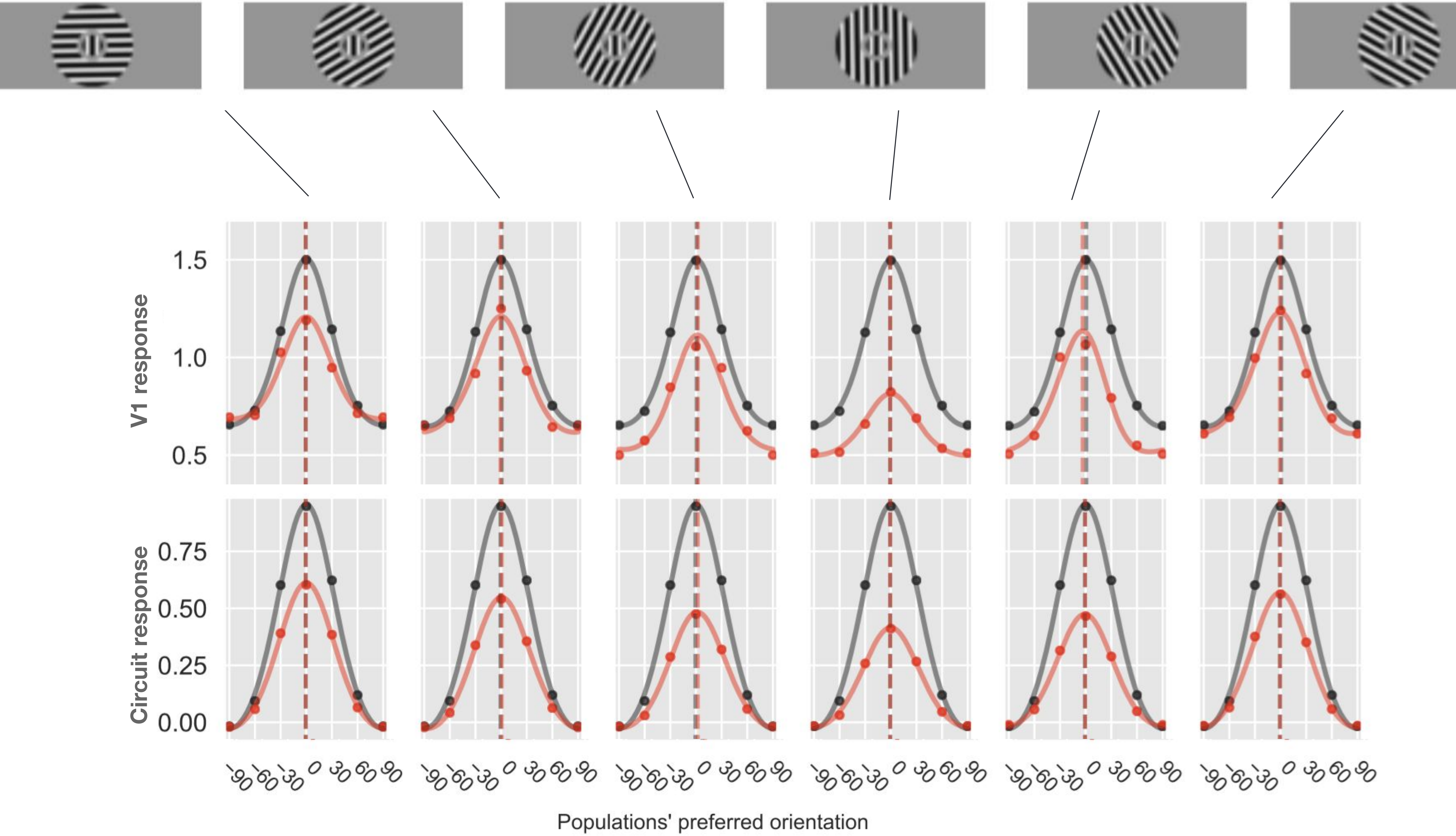


Linsley Govindarajan Serre in prep

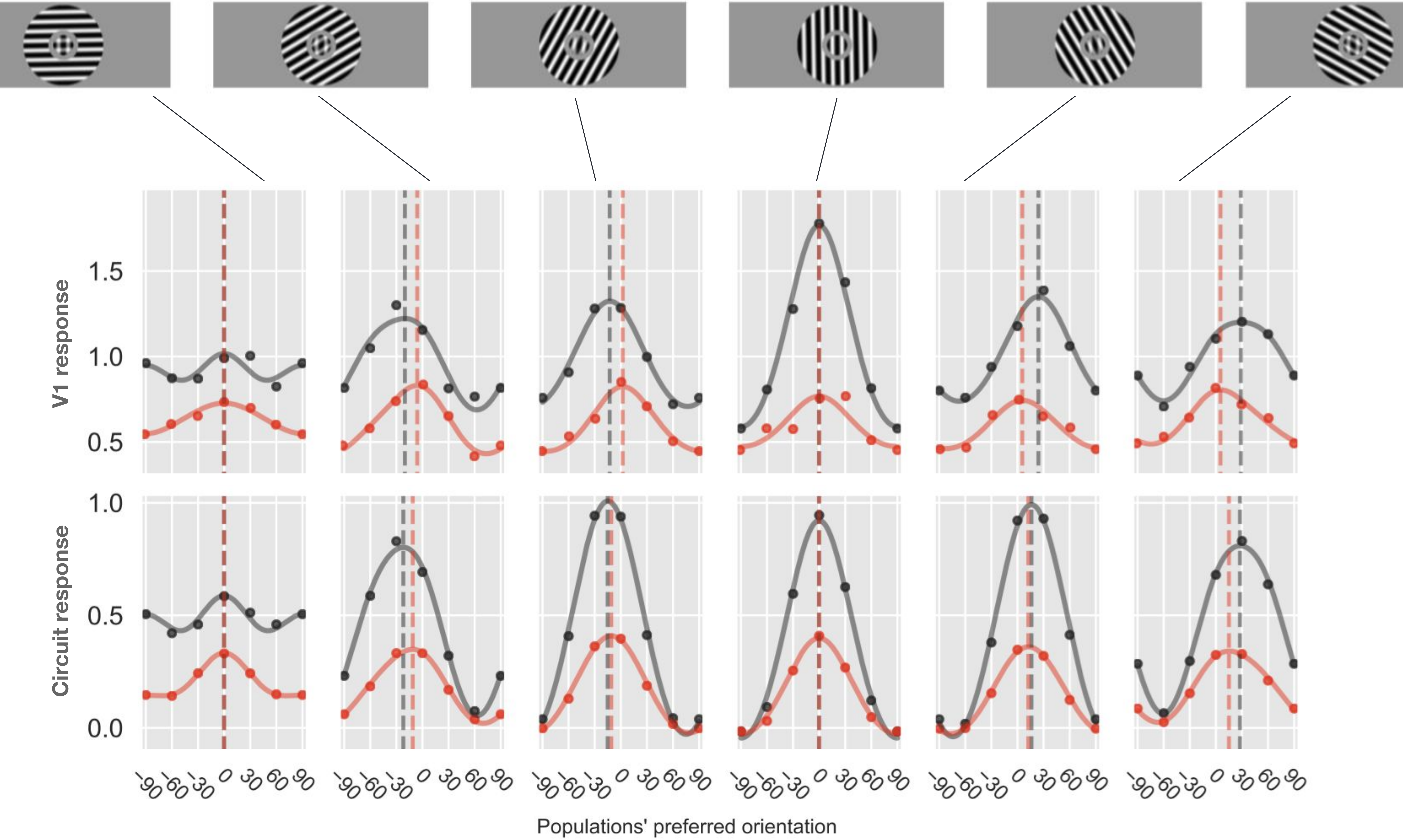
V1-like stimulus competition
(Busse, Wade, & Carandini 2009)



Tuned surround suppression
(Trott & Born, 2015)

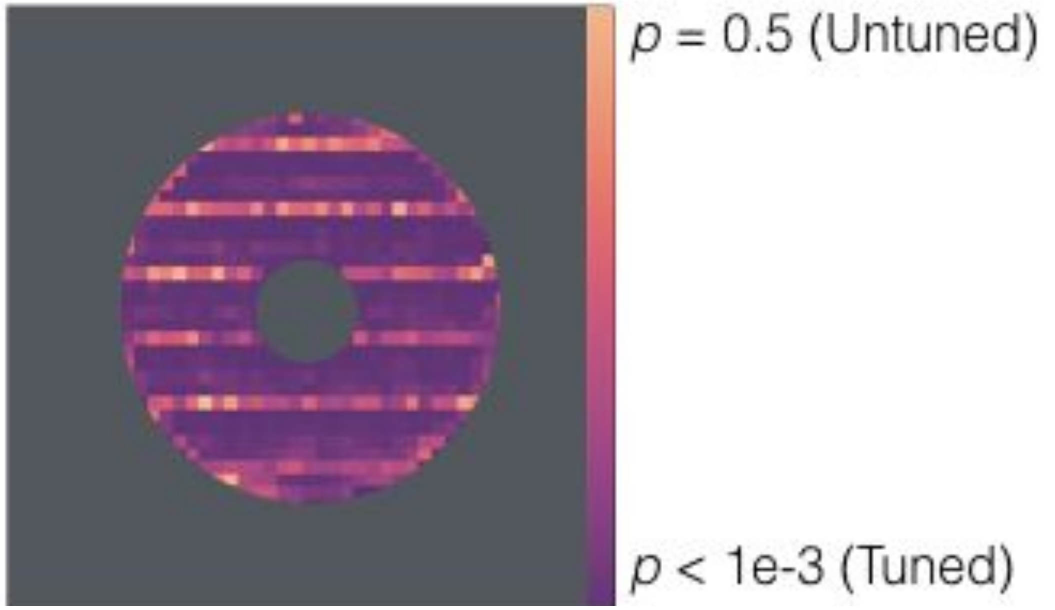


Feature-selective surround suppression
(Trott & Born, 2015)

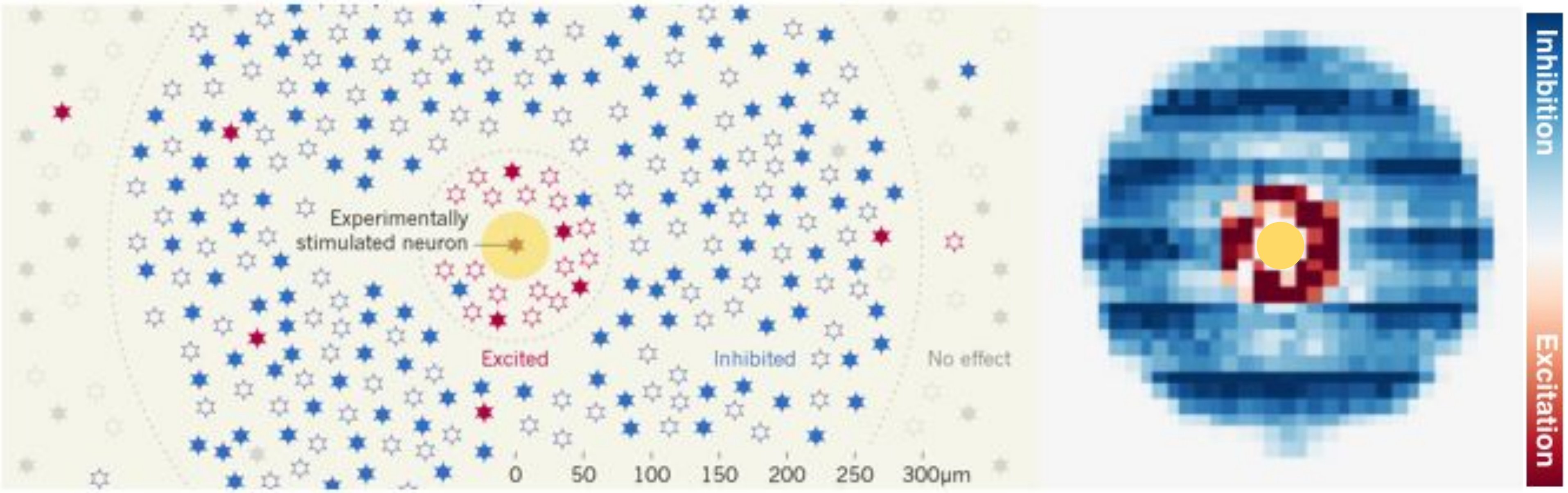


Optogenics
on the circuit model

eCRF connectivity is tuned



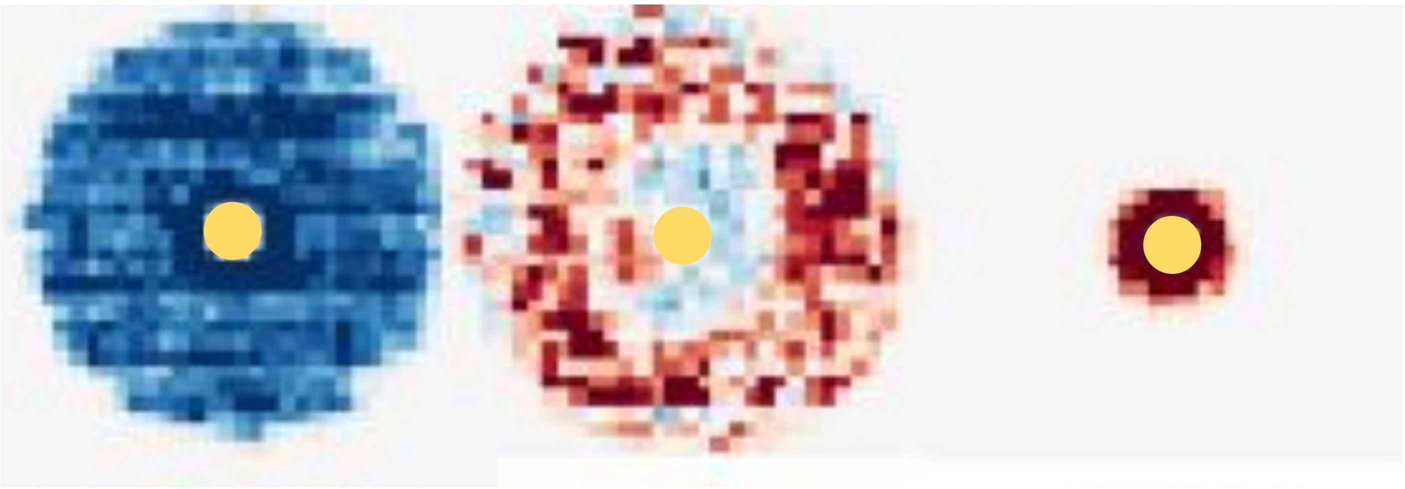
Chettih & Harvey 2019



Linsley Govindarajan Serre in prep

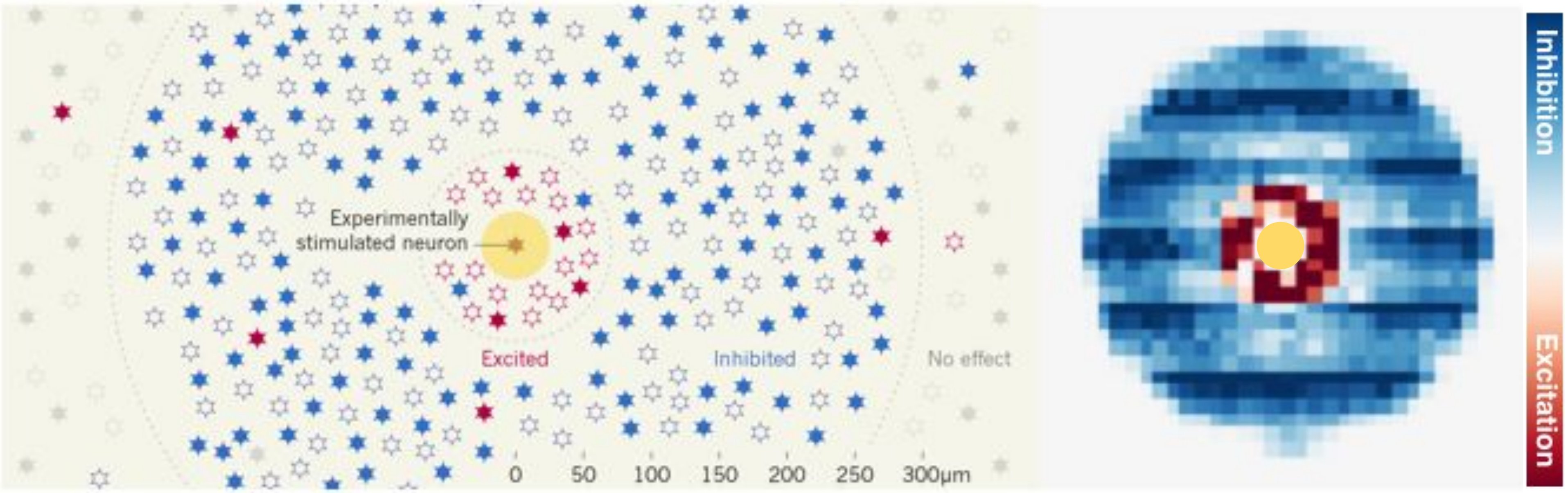
Optogenics

on the circuit model



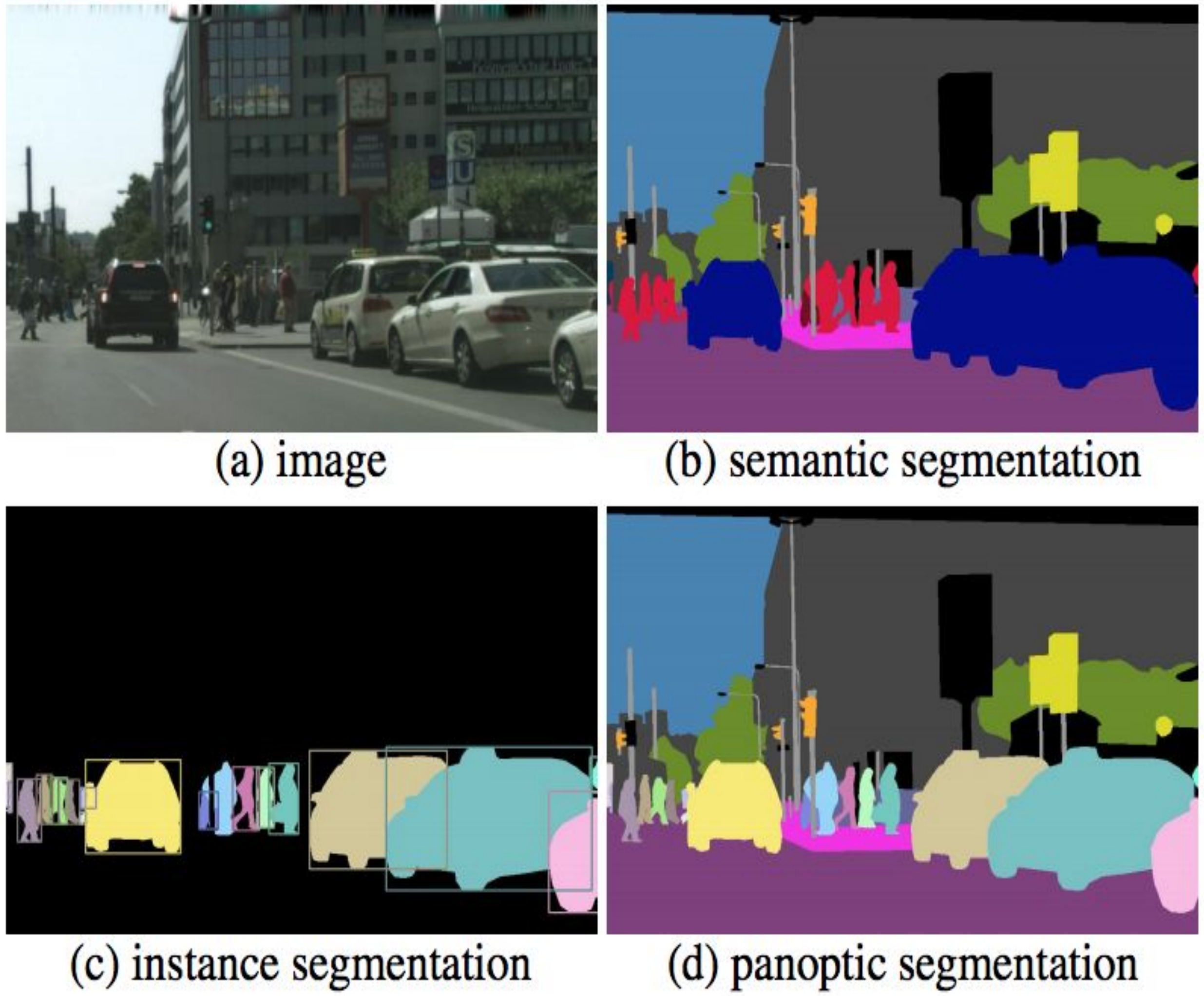
CRF Top-down only Horizontal only

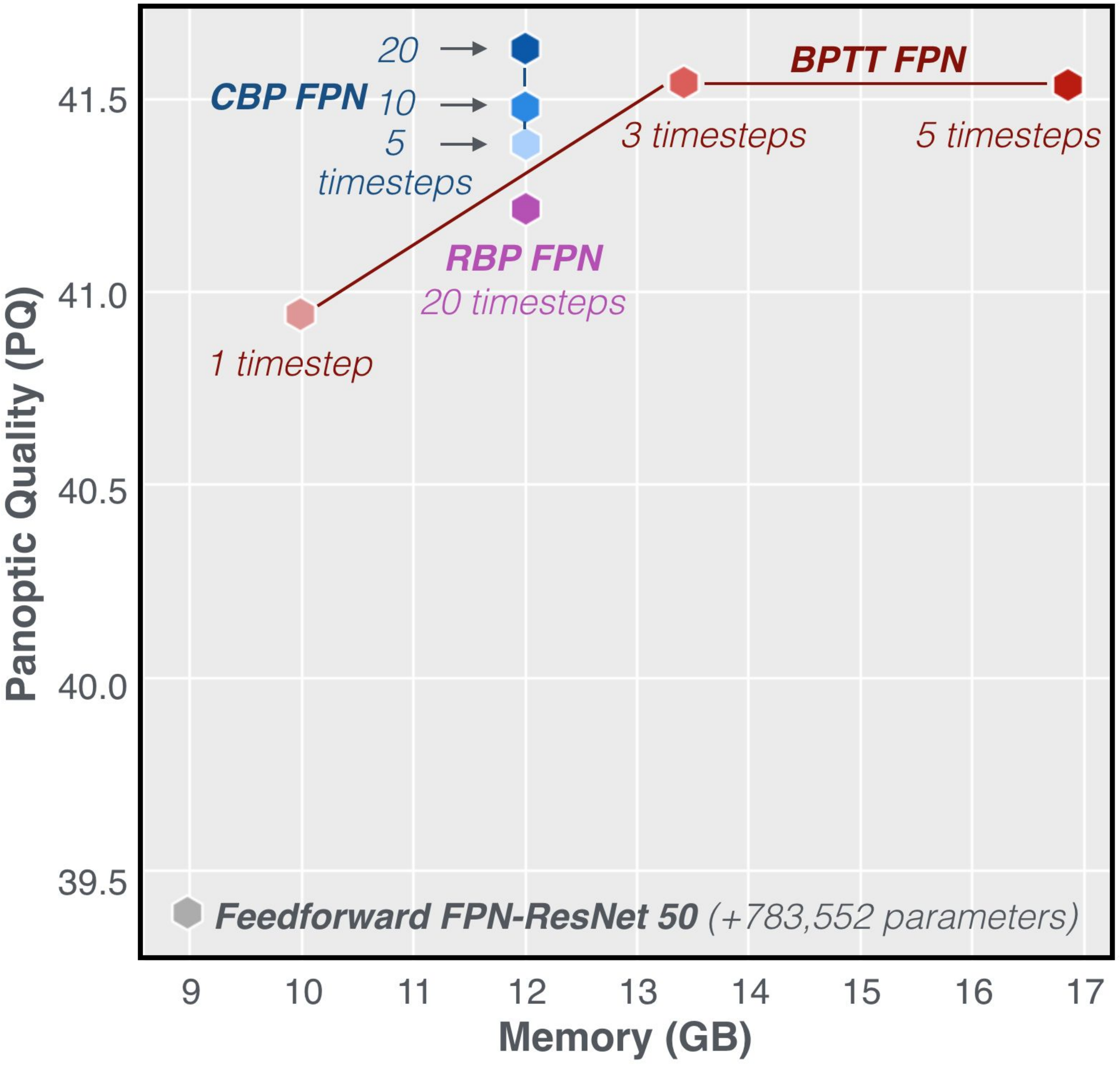
Chettih & Harvey 2019



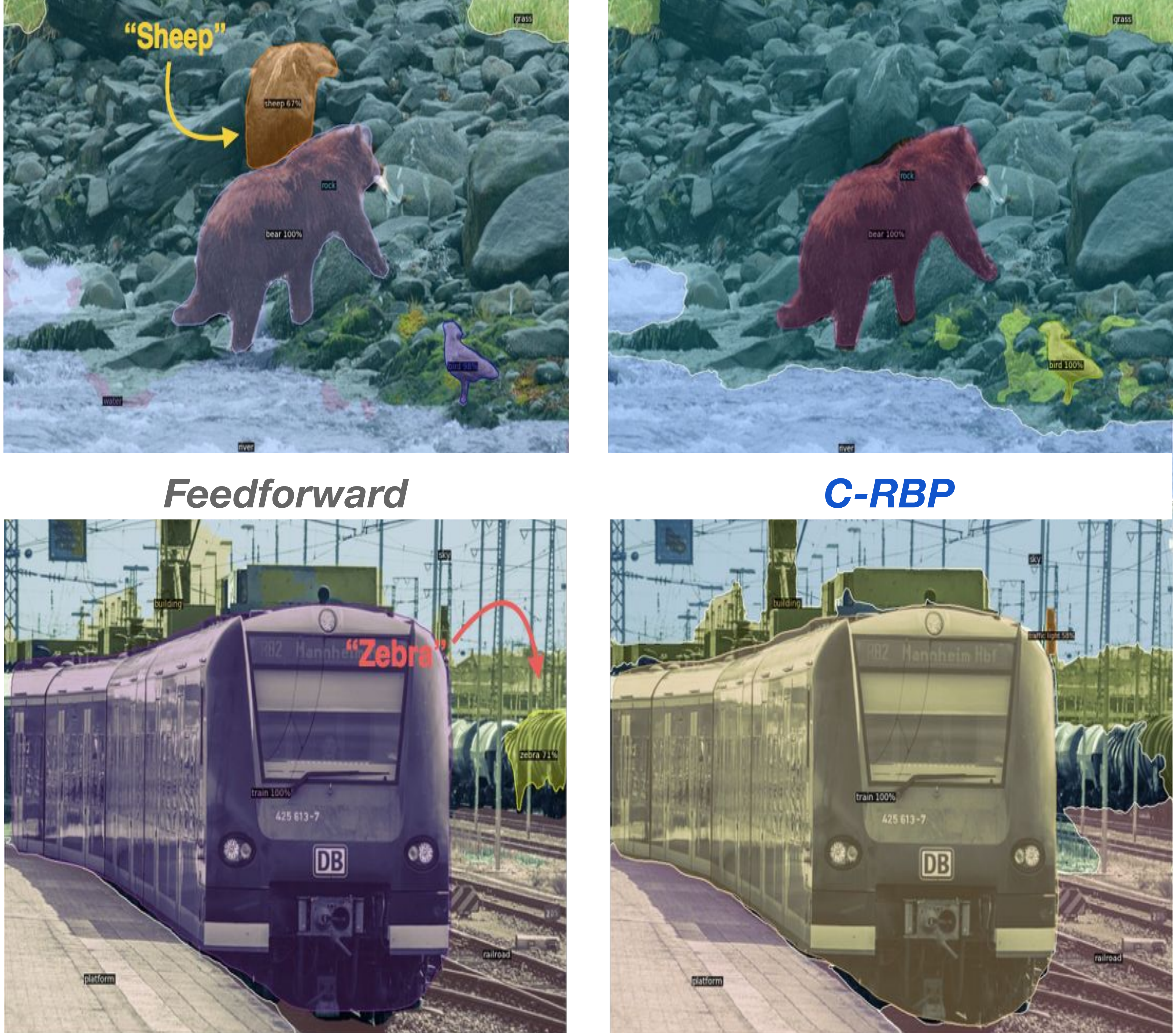
Linsley Govindarajan Serre in prep

Results on large-scale
computer vision challenges



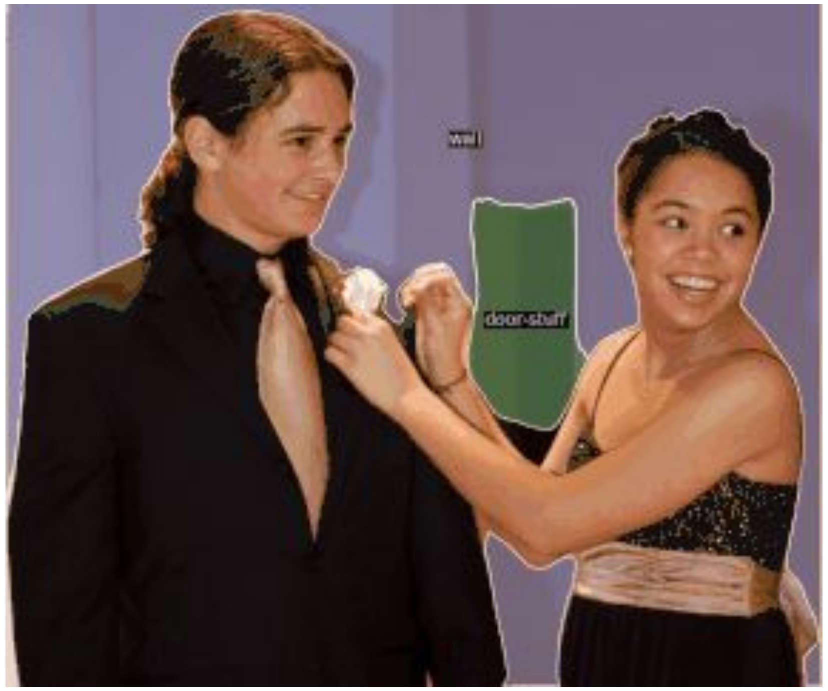


Linsley* Ashok* Govindarajan* Liu & Serre NeurIPS 2020



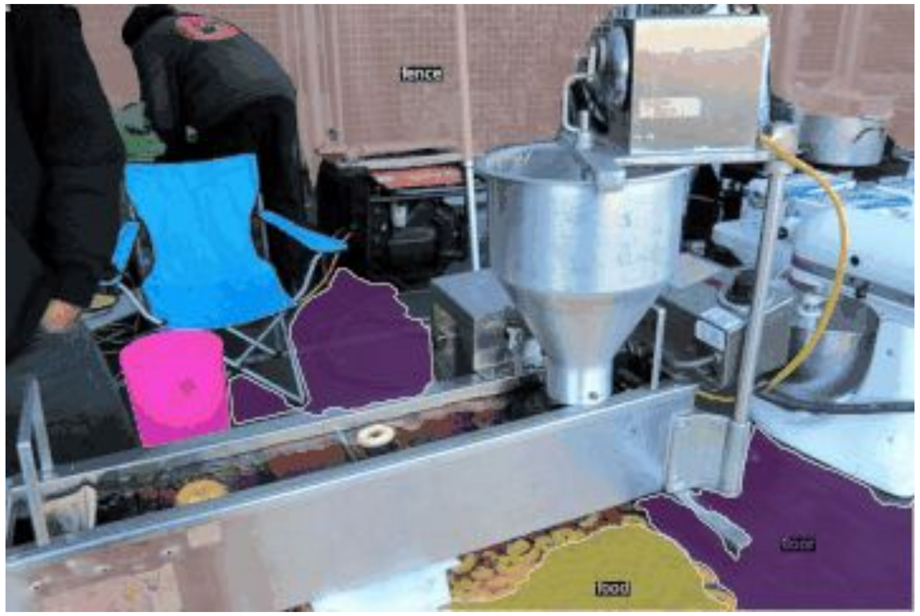
Linsley* Ashok* Govindarajan* Liu & Serre NeurIPS 2020

“Coloring” emerges
from stable recurrent circuits



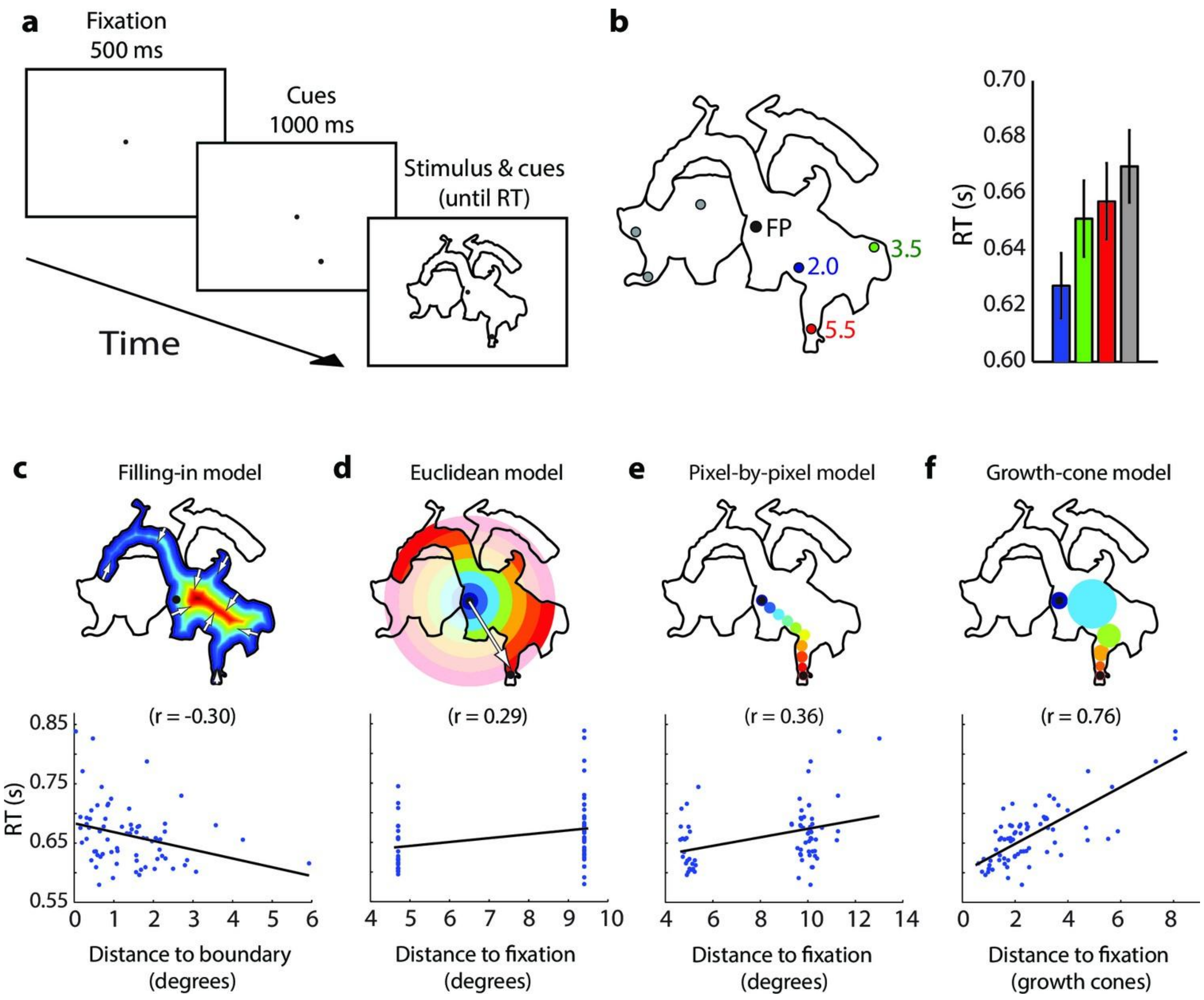
Linsley* Ashok* Govindarajan* Liu & Serre NeurIPS 2020

"Coloring" emerges
from stable recurrent circuits



Linsley* Ashok* Govindarajan* Liu & Serre NeurIPS 2020

"Coloring" emerges
from stable recurrent circuits



Jeurissen et al eLife 2016

Summary

- New breed of highly-recurrent neural nets inspired by cortical circuits found in visual cortex
- Tilt illusion (and others!) as byproduct of neural circuits to help biological visual systems achieve robust and efficient perception
  - 2 opponent surround regions (exc. vs. inh)
  - “Tuned” (like-to-like) connections from eCRF
  - Asymmetry of exc. and inh.
- SOTA accuracy for contours detection, neural tissue segmentation and panoptic segmentation

Acknowledgments

Drew Linsley

Matt Ricci

Lakshmi Govindarajan

Alekh Ashok

Rex Liu

Junkyung Kim (now DeepMind)

David Mely (now Google Brain)