Predictive Error-Driven Learning in the Brain

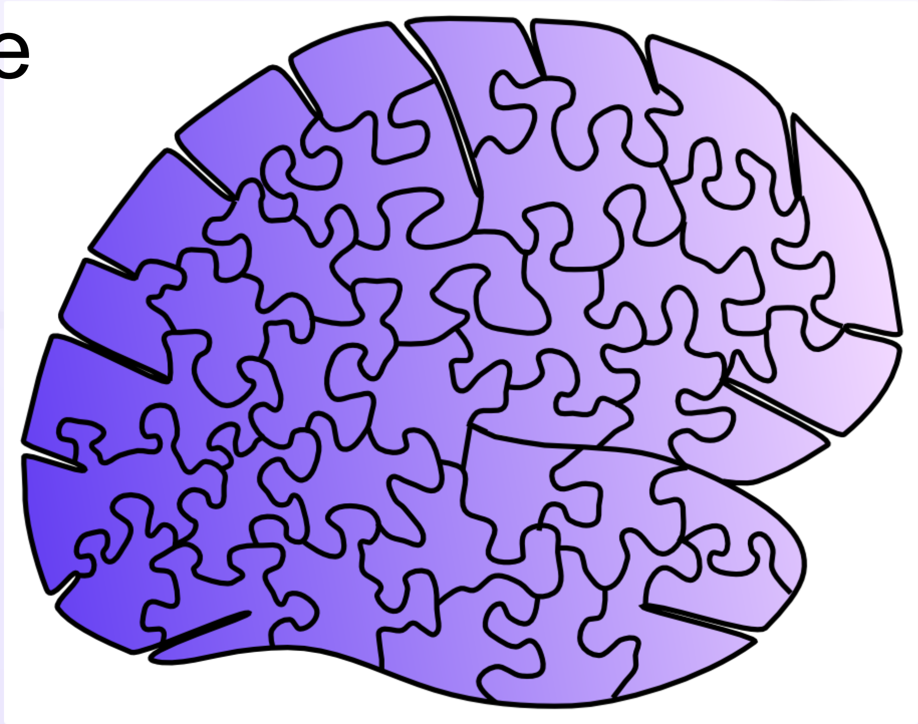
Randall C. O'Reilly

University of California Davis

eCortex, Inc

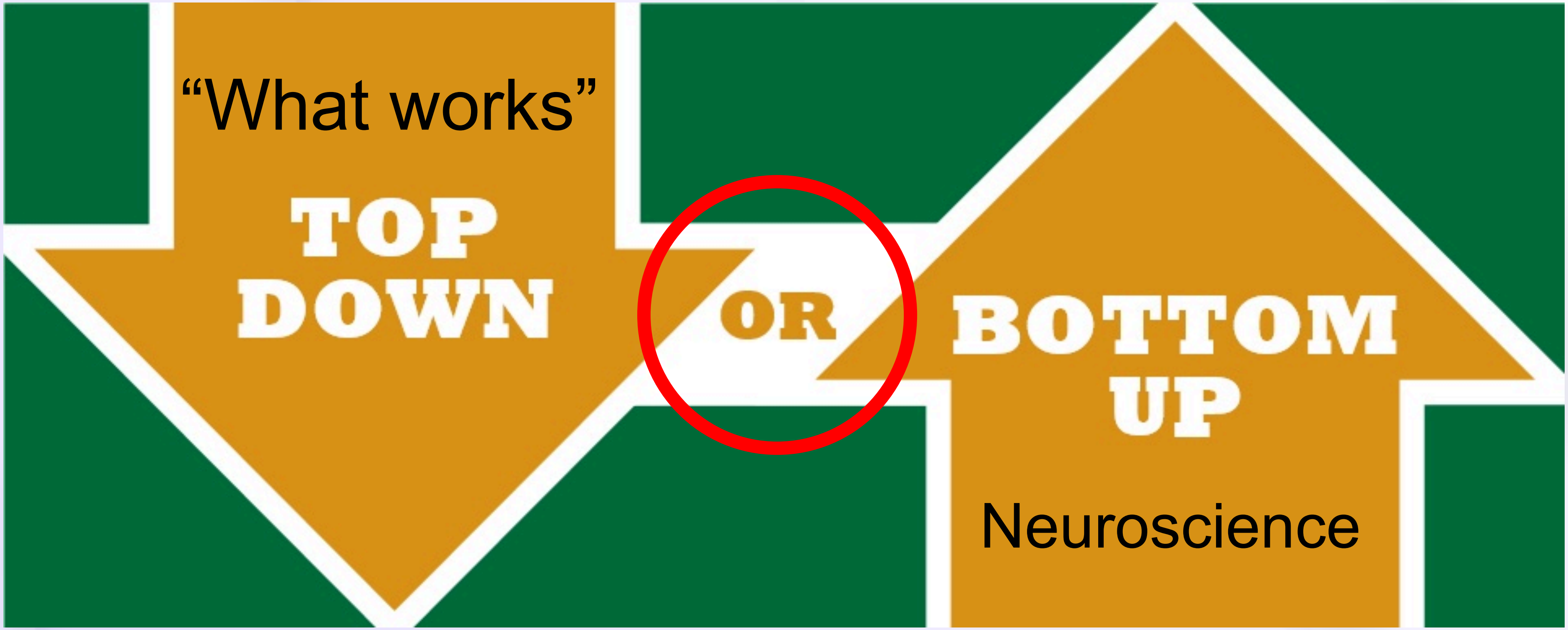
Learning is an embarrassment… (of potential riches!)

It is embarrassing how little we know about learning in the one place that really matters: the *neocortex*.



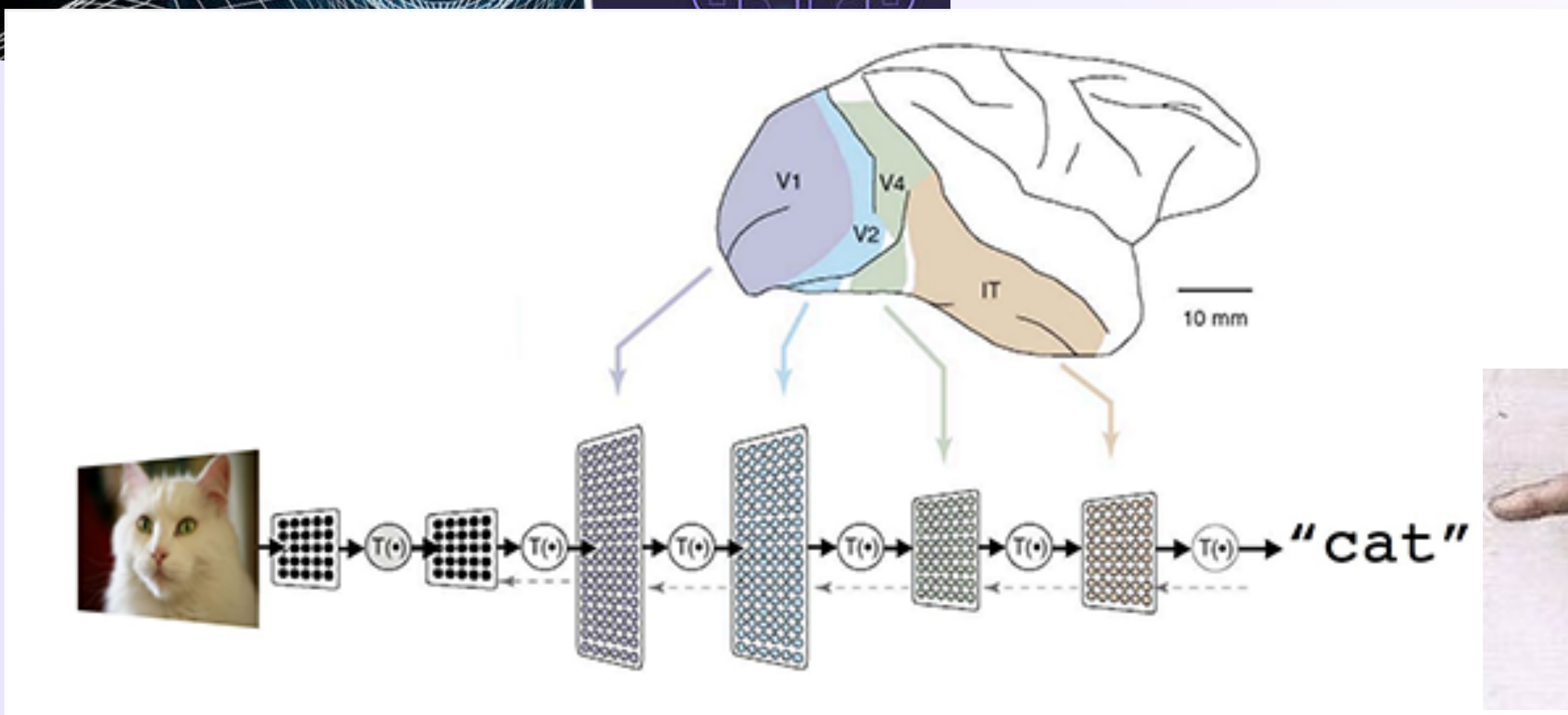
We know more about most other brain areas:

Basal Ganglia, Cerebellum, Hippocampus..



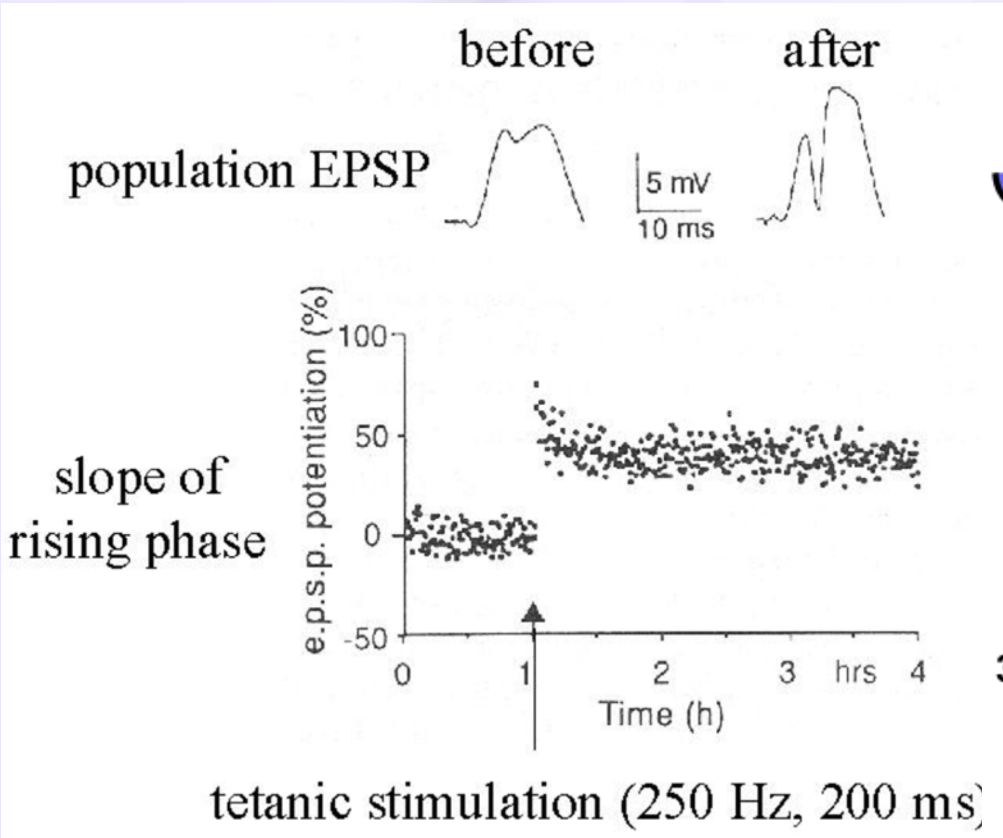
How do we get to “and”?

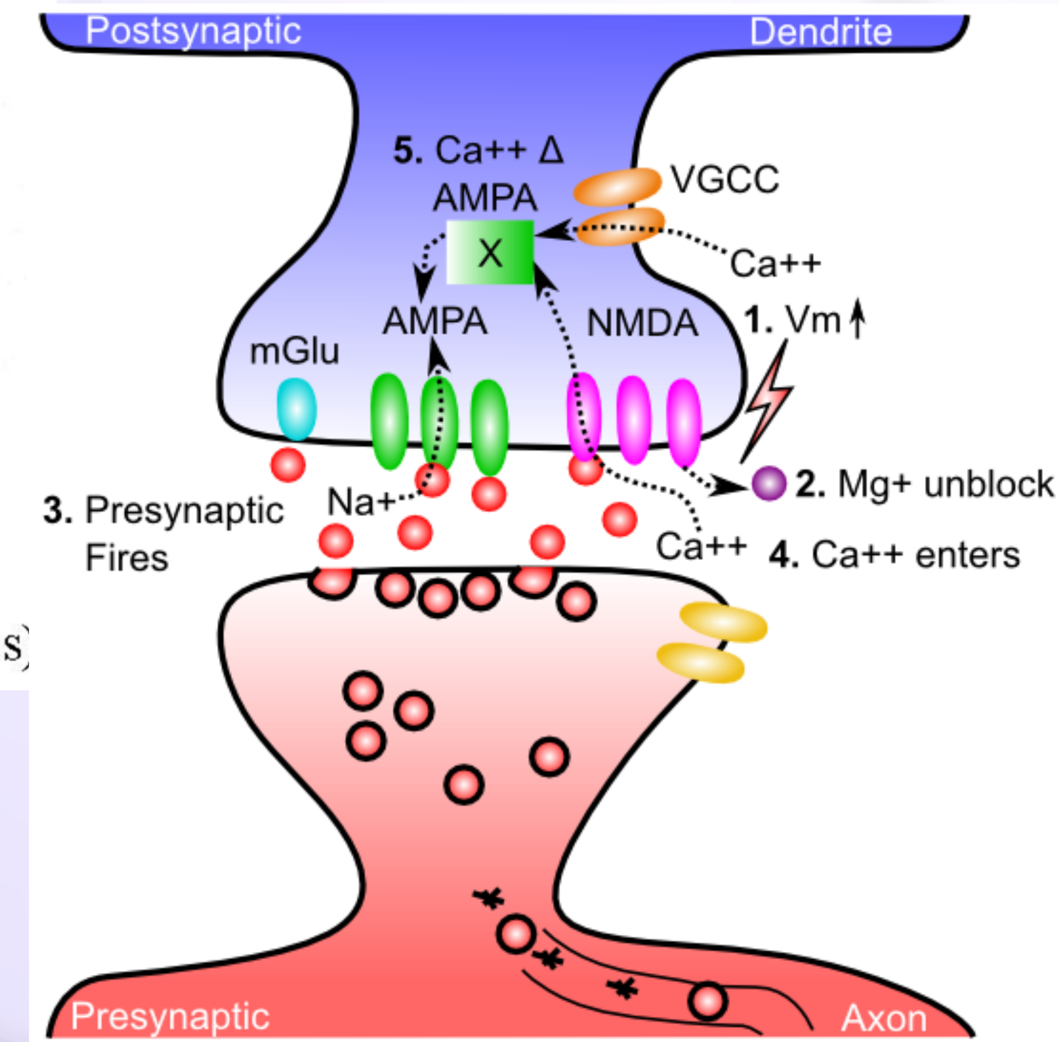
What Works:



Not biological: error backpropagation
Not psychological (where is that hand when you need it!?)

Neuroscience: Hebbian

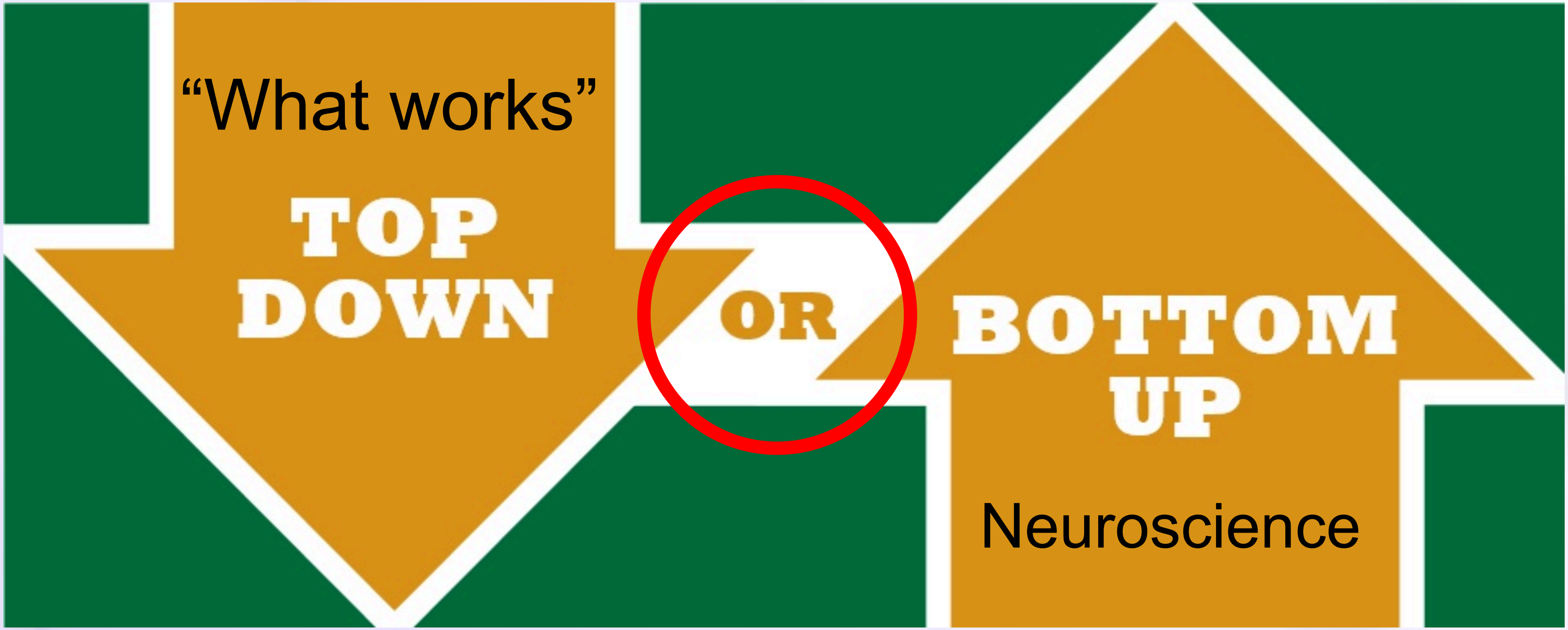




Hebbian is too Dumb!

Babies exhibit some serious active, theory-like learning abilities!

Not just passive soaking up of statistics..

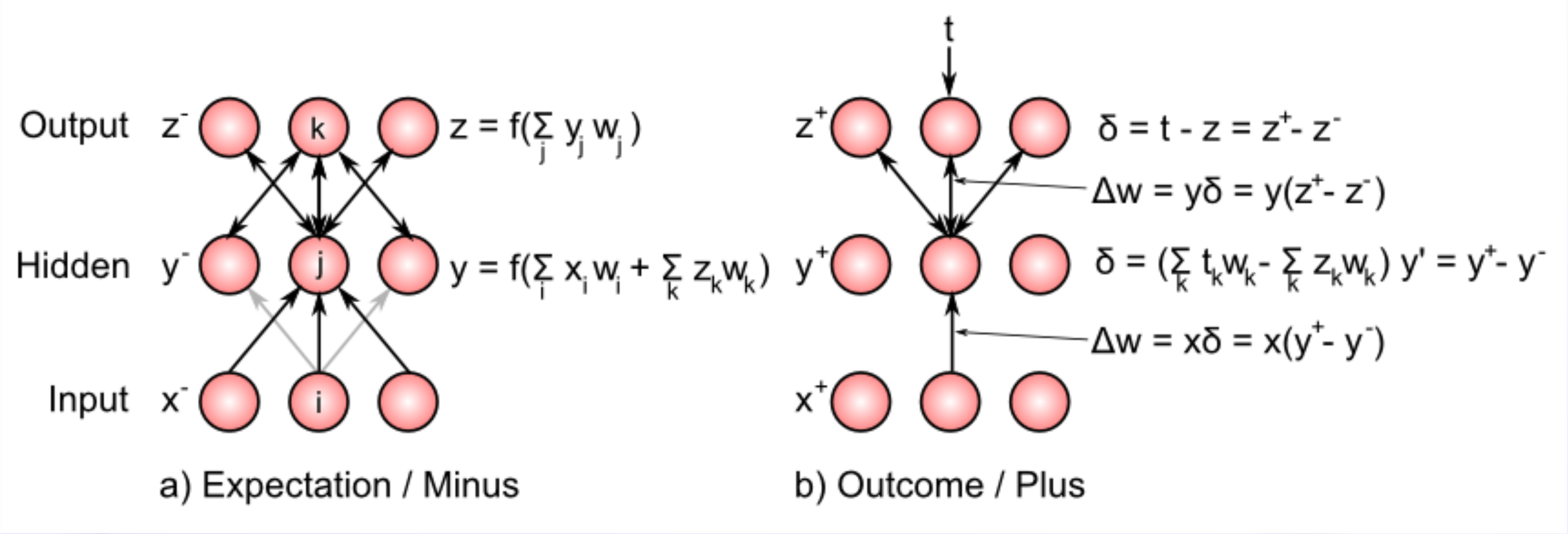


How do we get to "and"?

Three Levels

- Computational level: *error-driven & predictive learning*
- Implementational level: *thalamocortical loops*
- Functional level: *does it actually work?*

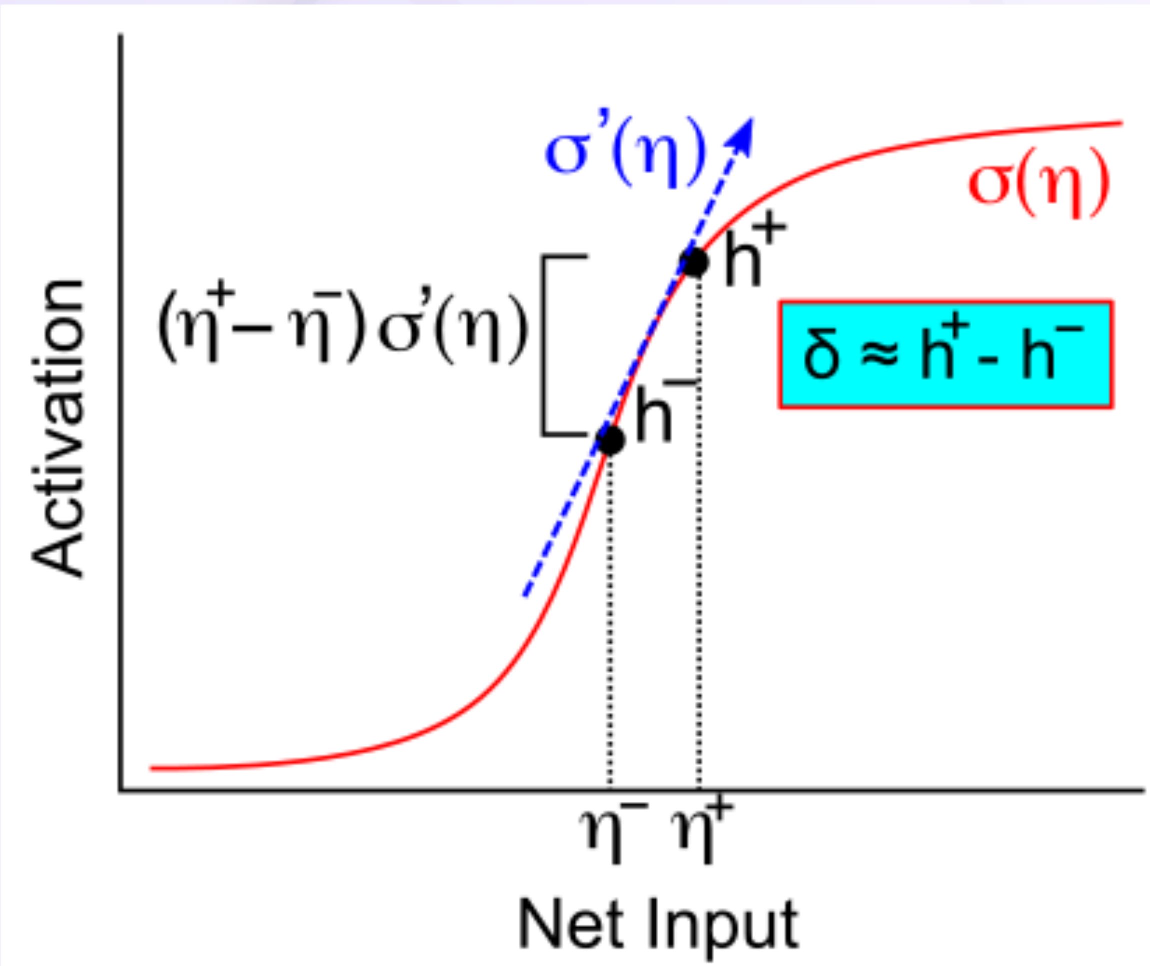
Bidirectional Connections Carry Error Gradients (GeneRec; O'Reilly, 1996)



$$dW = x^+y^+ - x^-y^-$$

(Midpoint integration + symmetry = Contrastive Hebbian = DBM)

Activation Diffs Implicitly Compute Derivatives (GeneRec; O'Reilly, 1996)



Free to use arbitrarily complex activation functions!

Many different approaches..

(Whittington & Bogacz Review, TICS, 2019)

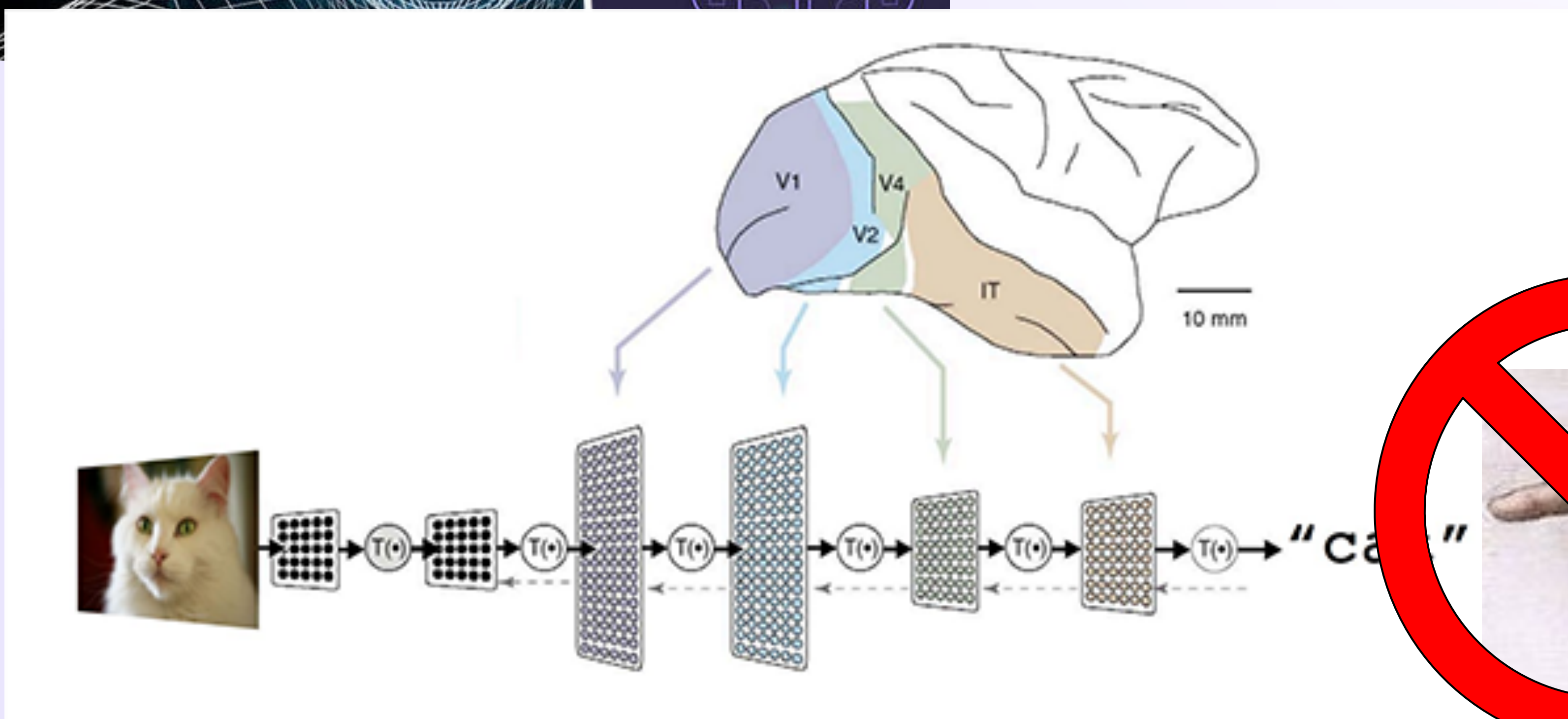
Table 1. Comparison of Models

		Temporal-error model		Explicit-error model	
		Contrastive learning	Continuous update	Predictive coding	Dendritic error
Properties[a]	Control signal	Required	Required	Not required	Not required
	Connectivity	Unconstrained	Unconstrained	Constrained	Constrained
	Propagation time	L-1	L-1	2L-1	L-1
	Pre-training	Not required	Not required	Not required	Required
Error encoded in		Difference in activity between separate phases	Rate of change of activity	Activity of specialised neurons	Apical dendrites of pyramidal neurons
Data accounted for		Neural responses and behaviour in a variety of tasks	Typical spike-time-dependent plasticity	Increased neural activity to unpredicted stimuli	Properties of pyramidal neurons
MNIST performance[b]		~2–3	–	~1.7	~1.96

Scellier, B. and Bengio, Y. (2017) Equivalence of equilibrium propagation and recurrent backpropagation. arXiv preprint arXiv:1711.08416

Whittington, J.C.R. and Bogacz, R. (2017) An approximation of the error backpropagation algorithm in a predictive coding net- work with local Hebbian synaptic plasticity. Neural Comput. 29, 1229–1262

What Works:

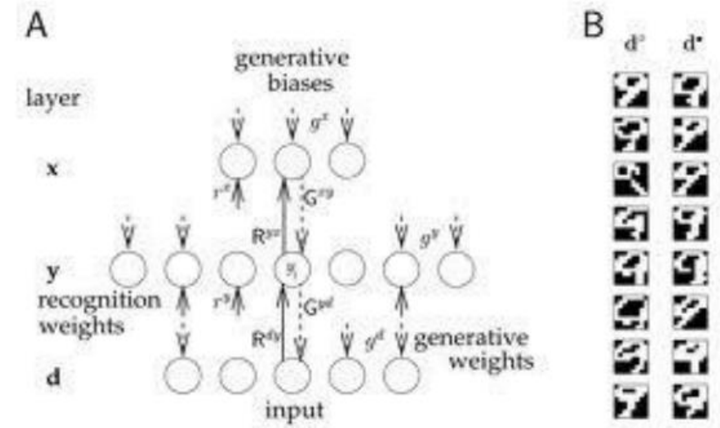


Not biological: error backpropagation
Not psychological (where is that hand when you need it!?)

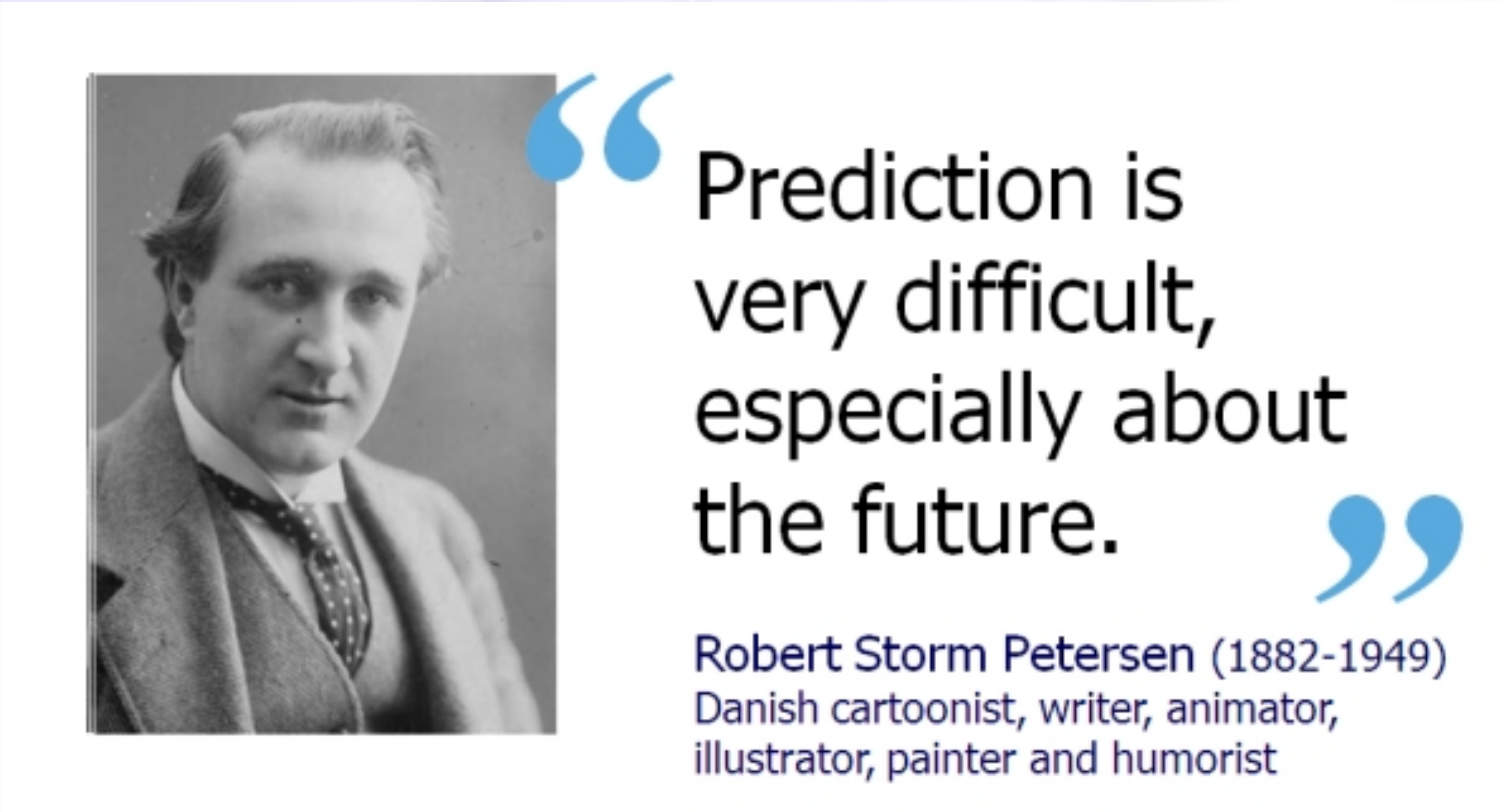
Helmholtz: Recognition by Synthesis

Helmholtz Machine

- Device or scheme that uses a generative model to furnish a recognition density. They learn hidden structure in data by optimising the parameters of generative models.

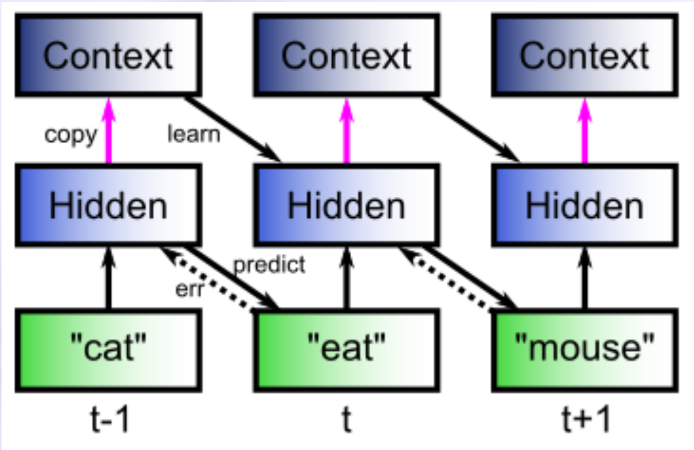


Auto-encoders, Bayesian models, Rao & Ballard, Friston et al..

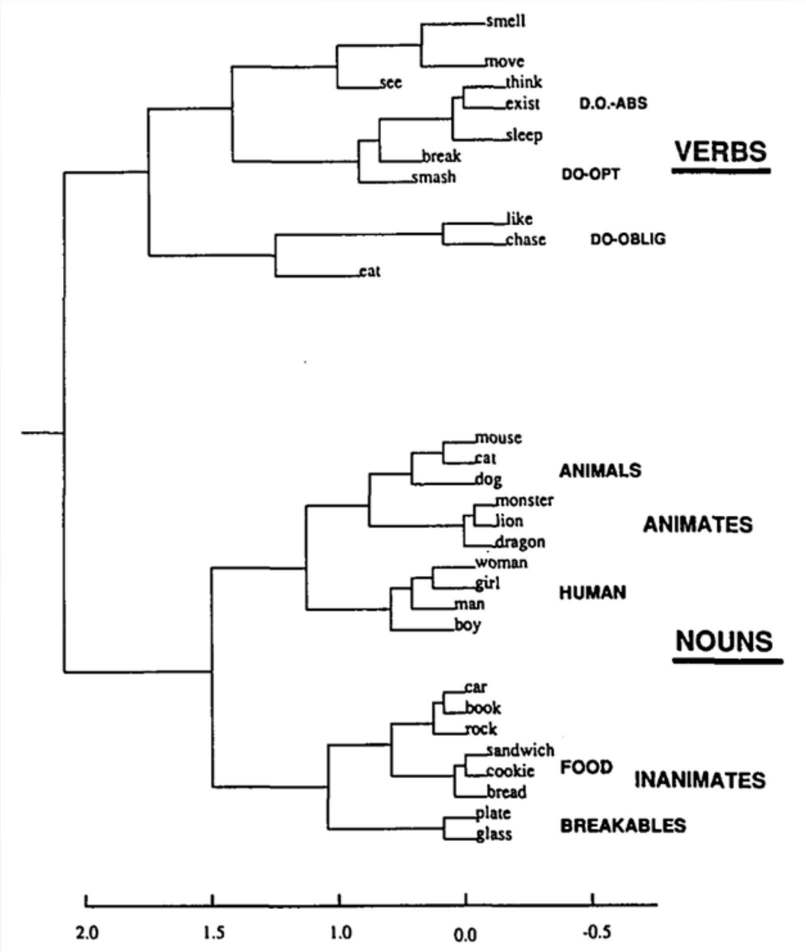


Key idea: We learn by constantly generating hypotheses or predictions about *what will happen next!*

The Predictive Bootstrap
(Elman 1990; Elman, Bates, *et al.* 1996)



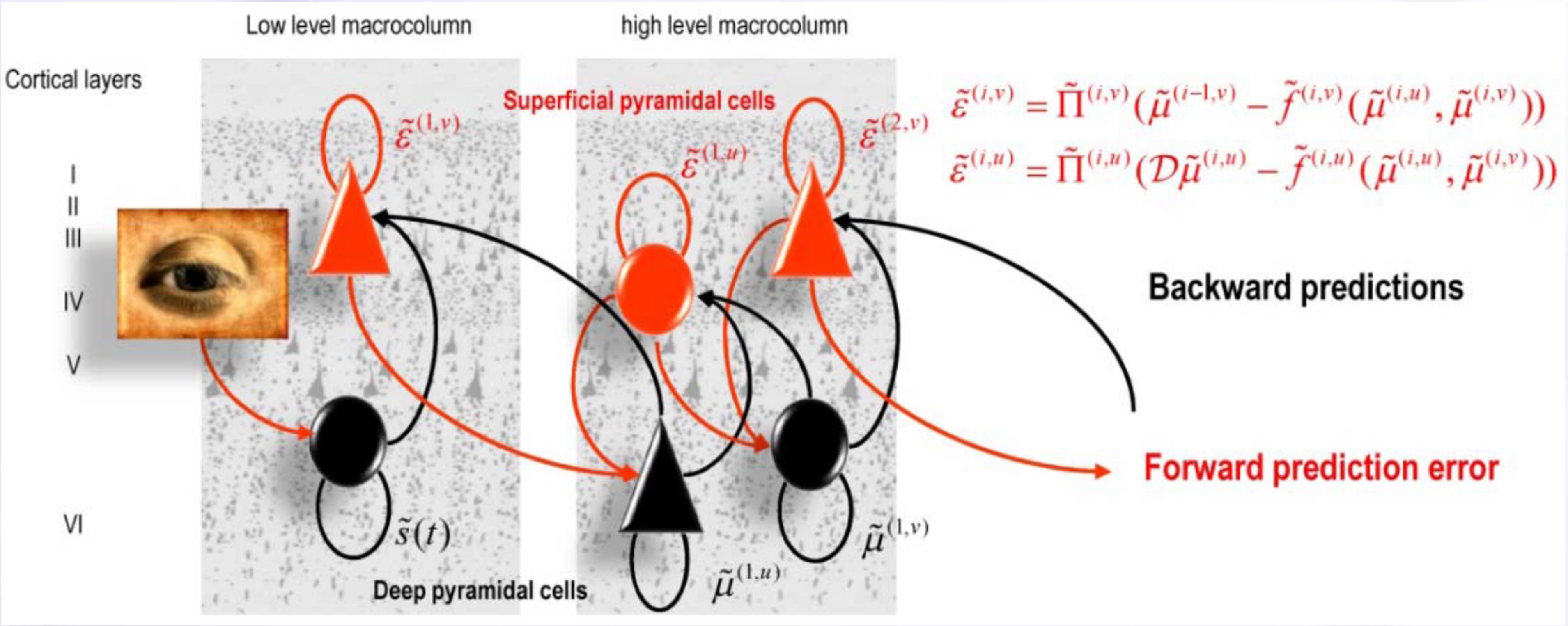
Learning to predict words induces verb vs noun category representations in hidden layer of SRN



The future is *free*! *If* you can predict it, you *know* it!

Biologically, how does it work?

“Standard” approach
(Friston et al)

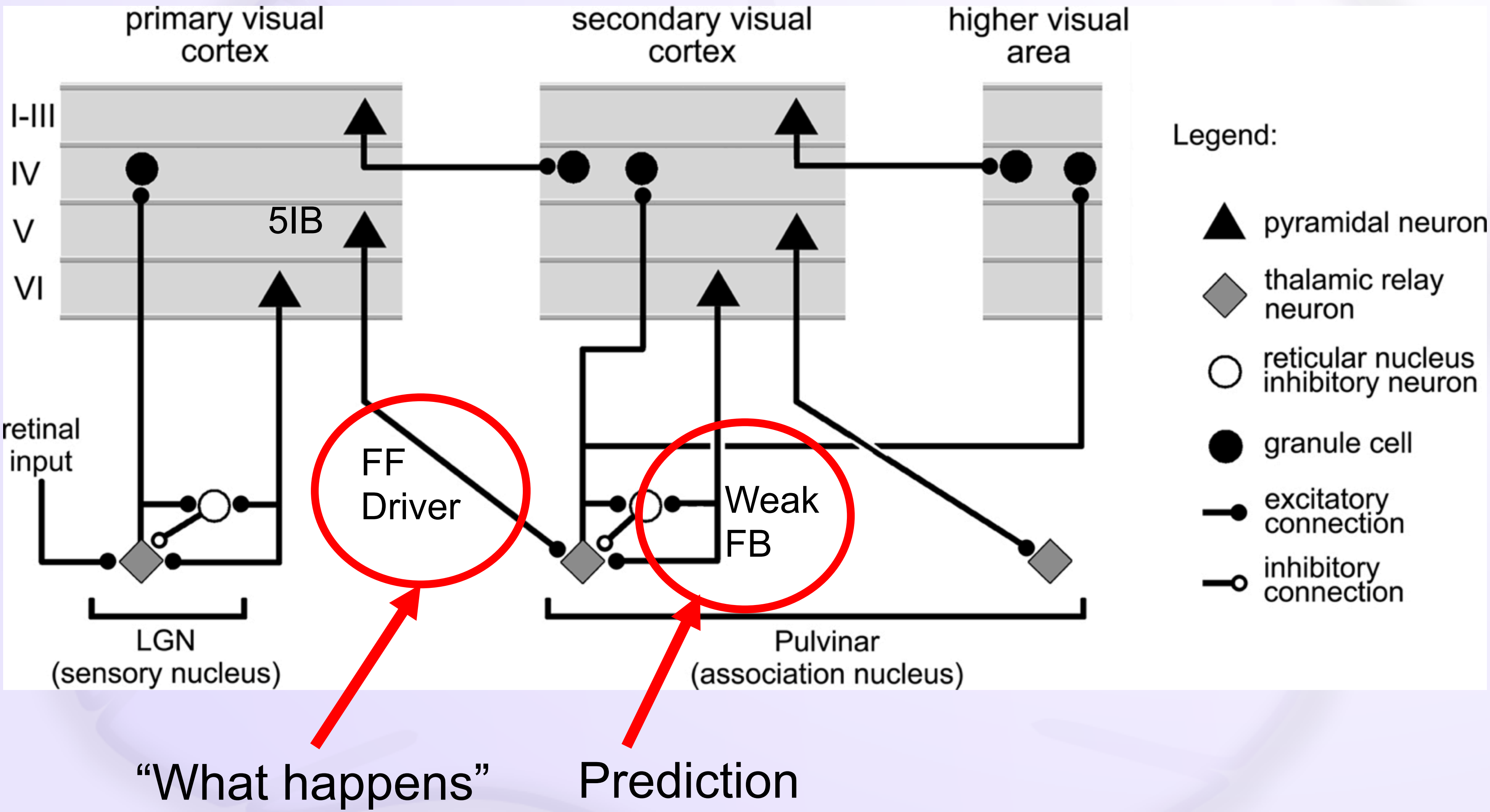


Explicit error-coding neurons subtracting top-down vs. bottom up

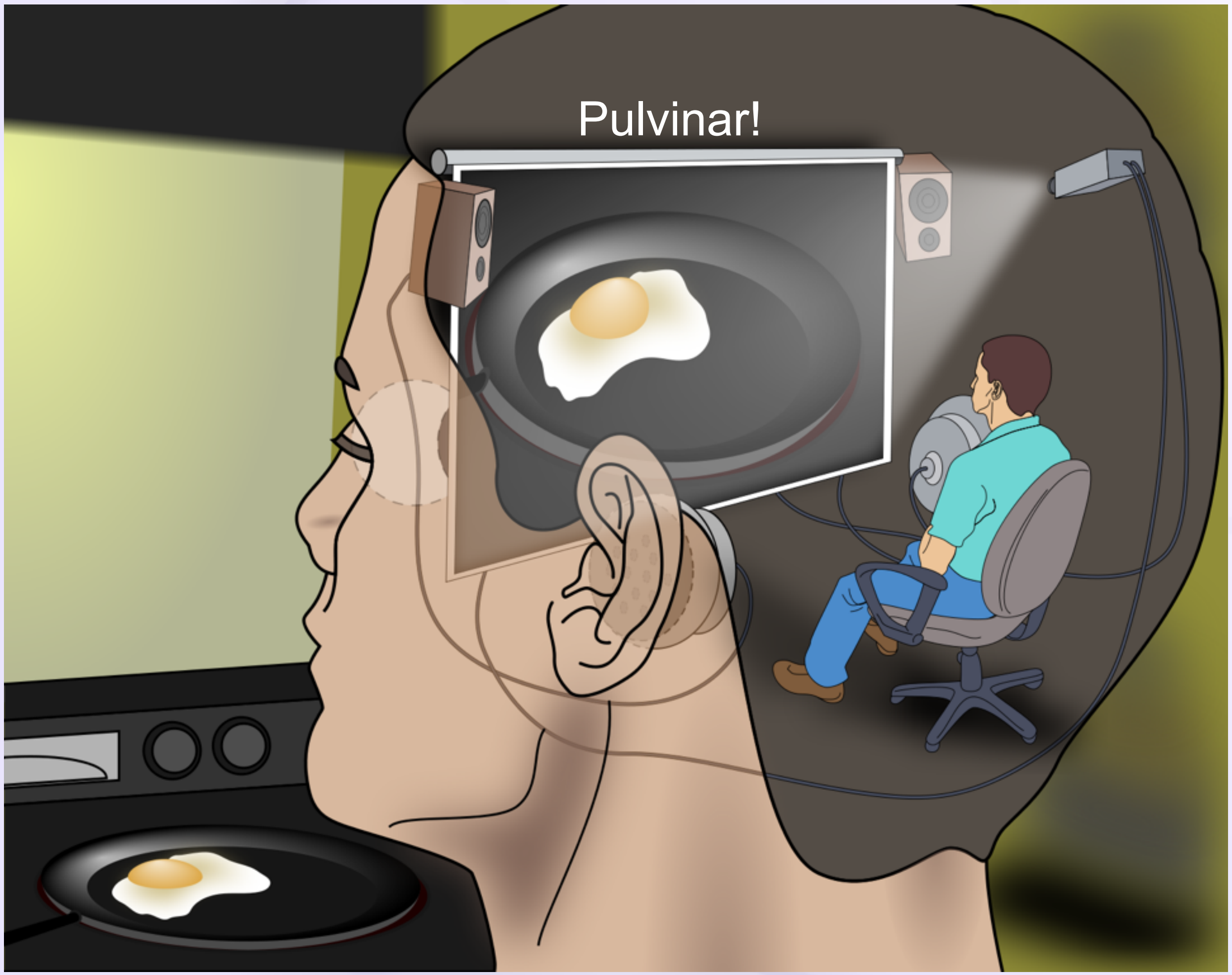
But: no evidence of such cells!

Thalamocortical Loop Biology

(Sherman & Guillery, 2006)



The Pulvinar = Projection Screen
(c.f. Mumford, 1991 “blackboard”)



Pulvinar receives connections from all over visual cortex

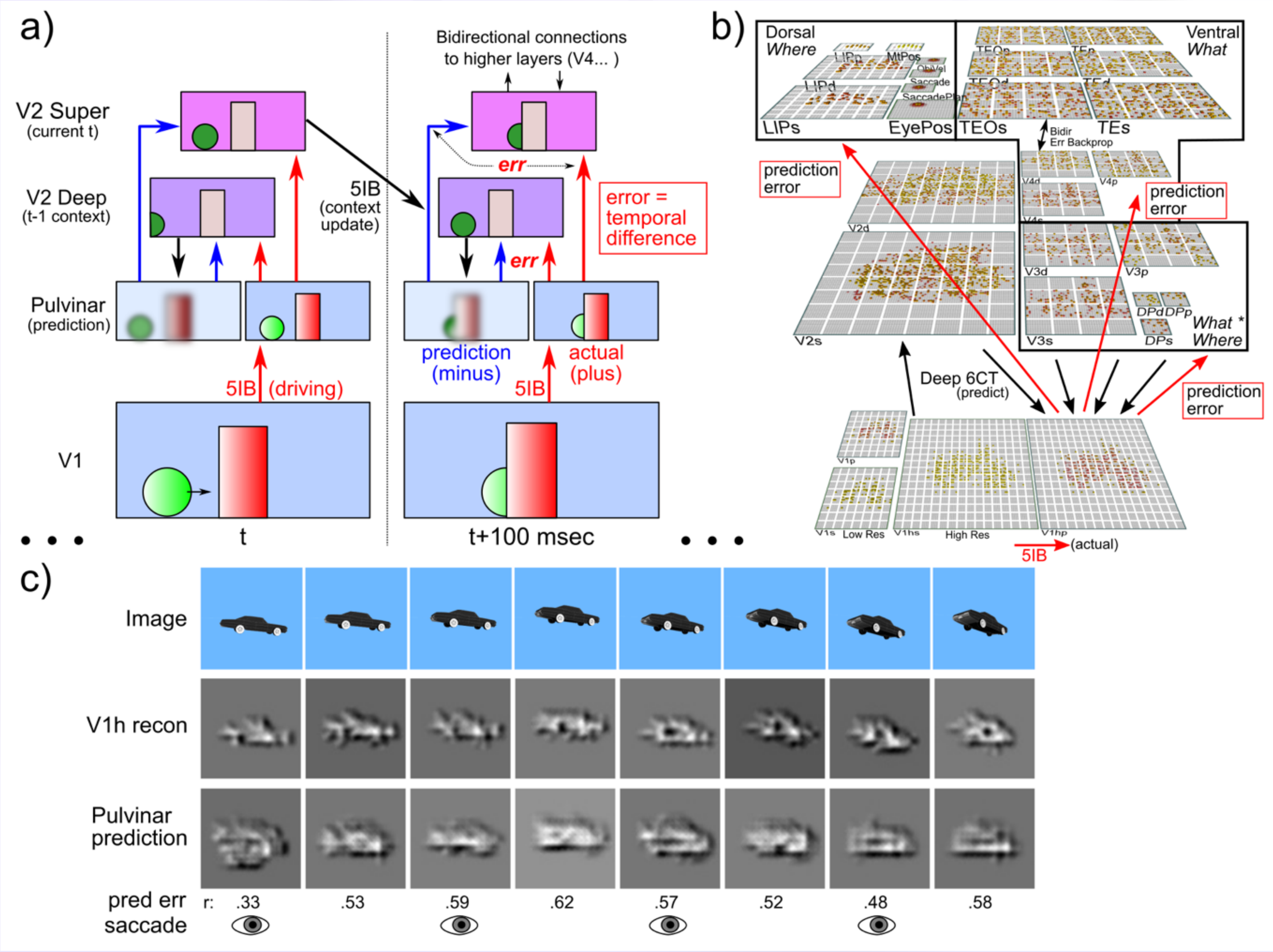
and projects back out to these same areas

Two inputs:

1. Few strong feedforward: *“what happens”*

2. Many weaker feedback: *prediction*

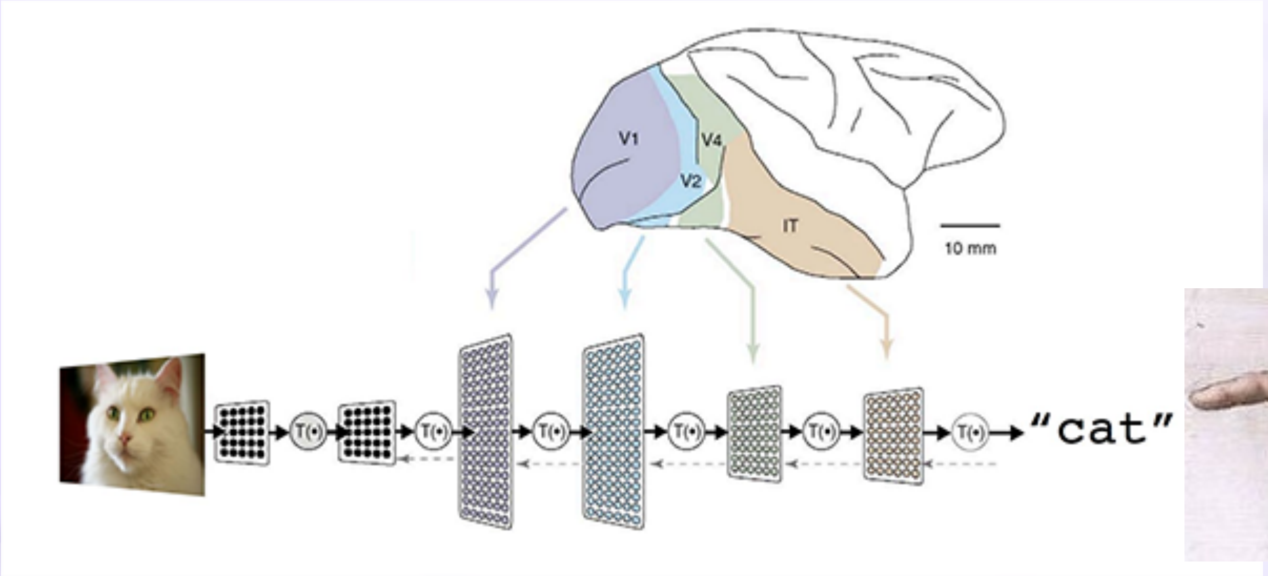
Deep Predictive Learning

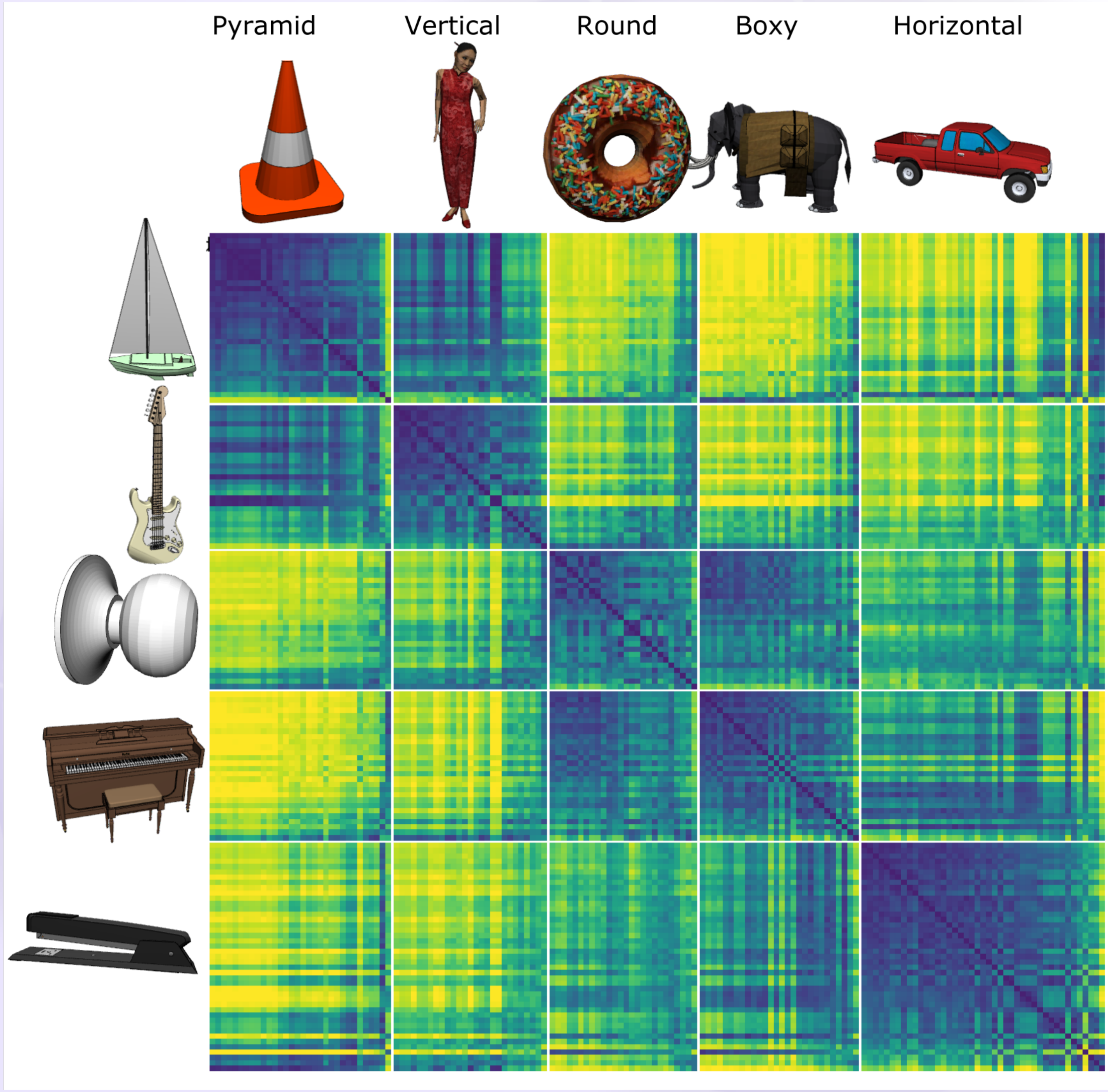


Functionally: does it work?

Can merely predicting low-level sensory inputs produce higher-level abstract representations?

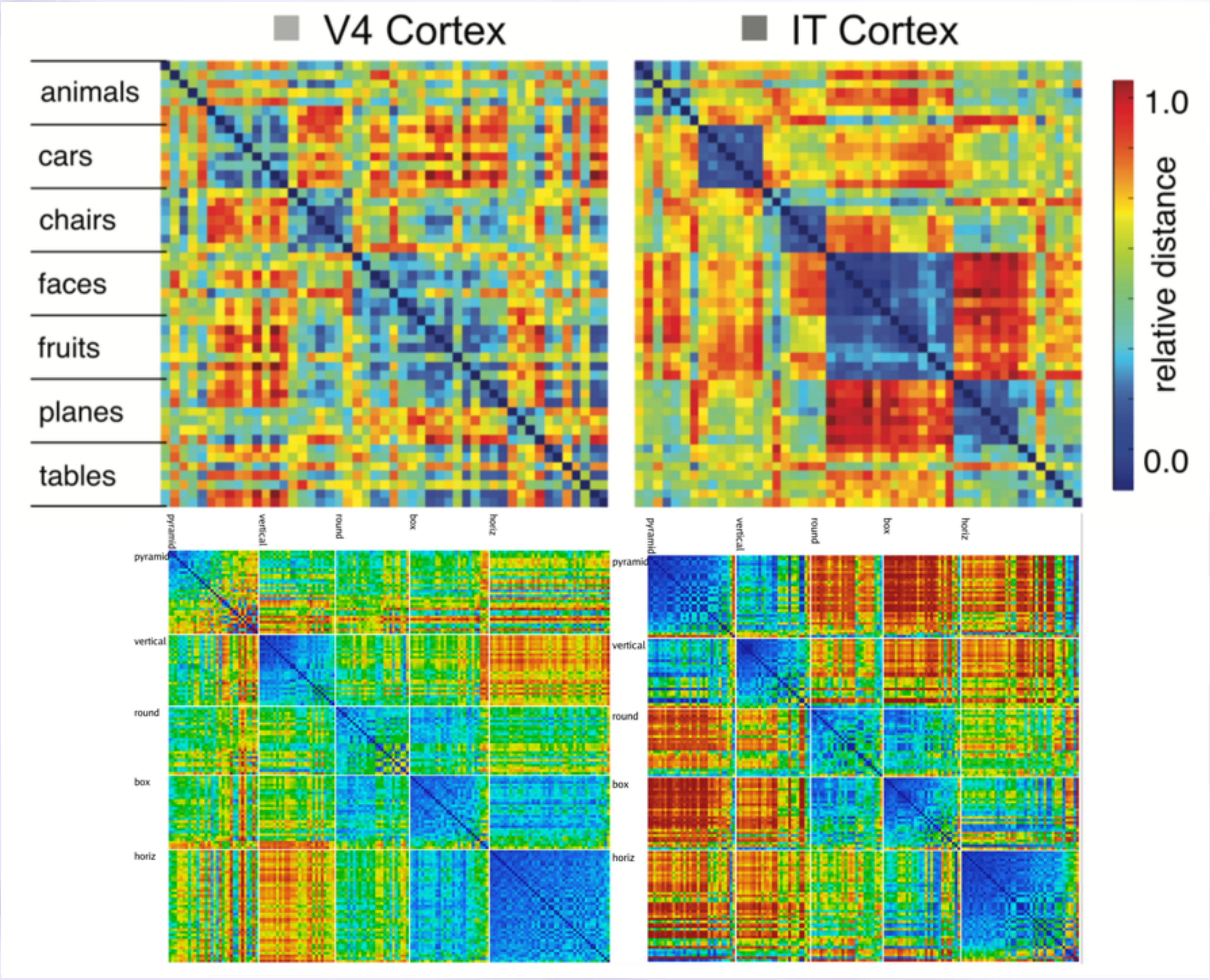
If not, maybe we still need that hand??



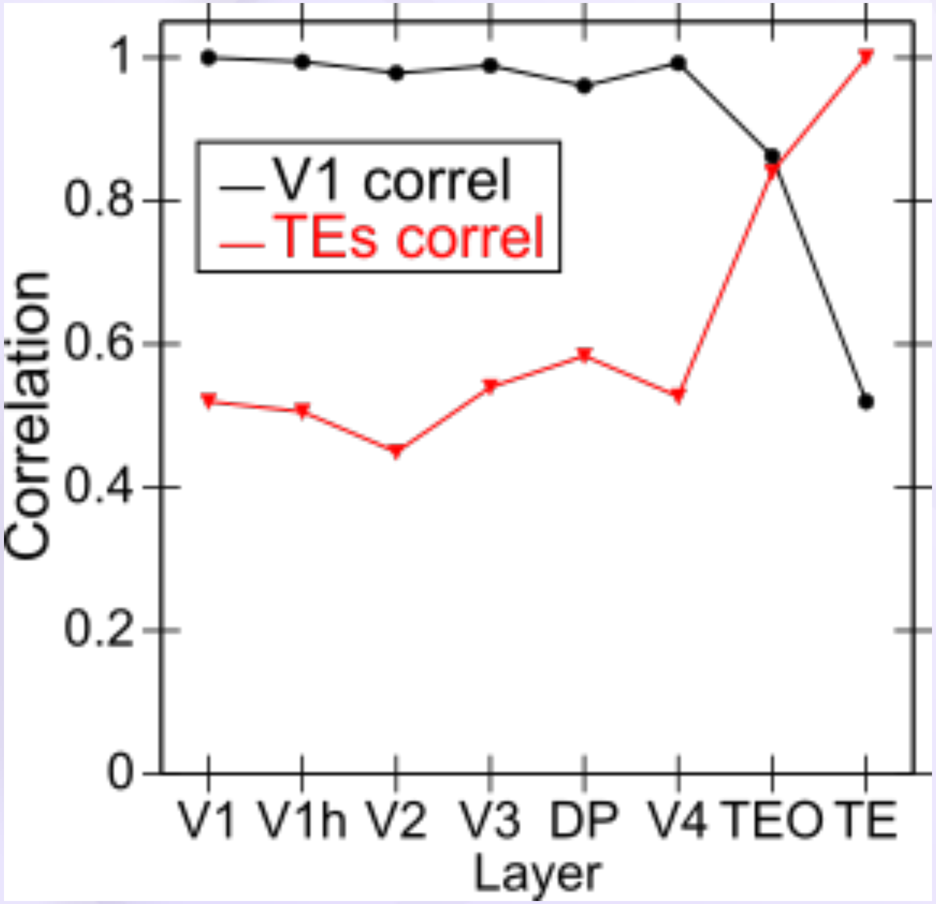


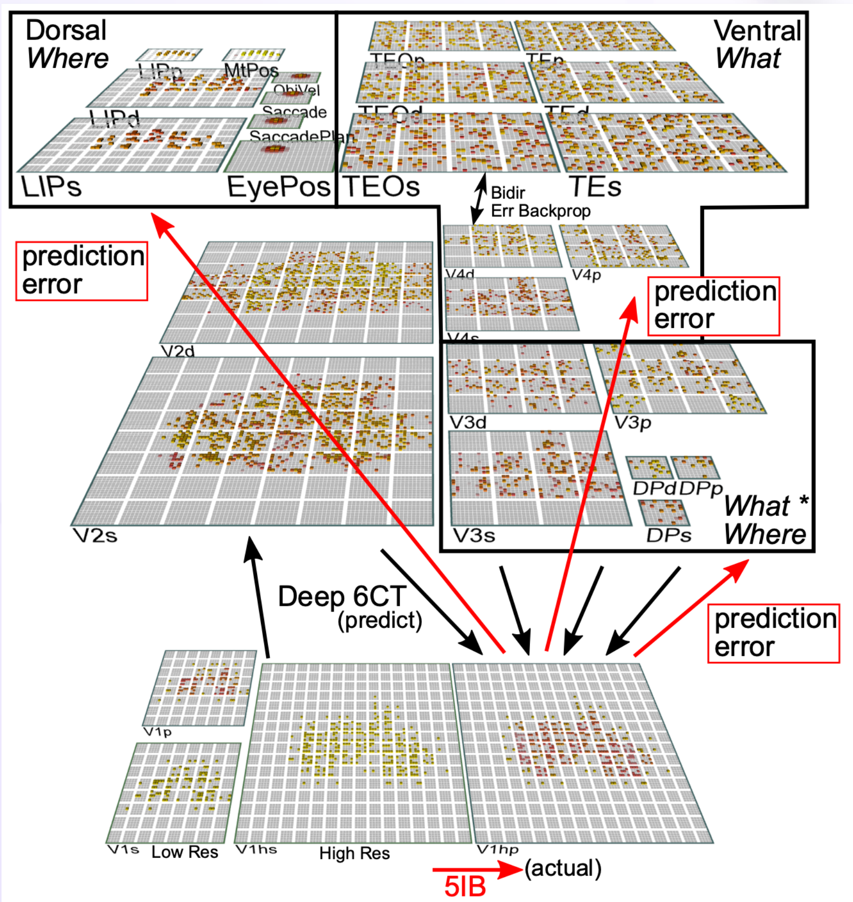
Model discovered
Shape categories!

Model vs. Monkeys: Categories Emerge in Higher Layers



Categories not in V1; Emerge in IT Obj Rec Pathway



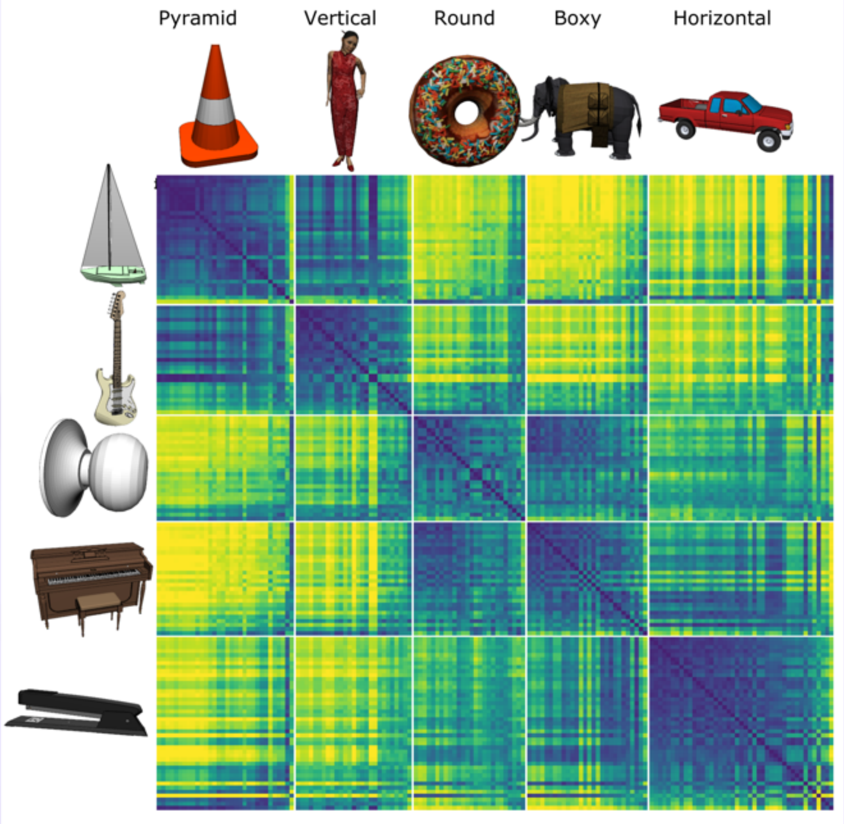


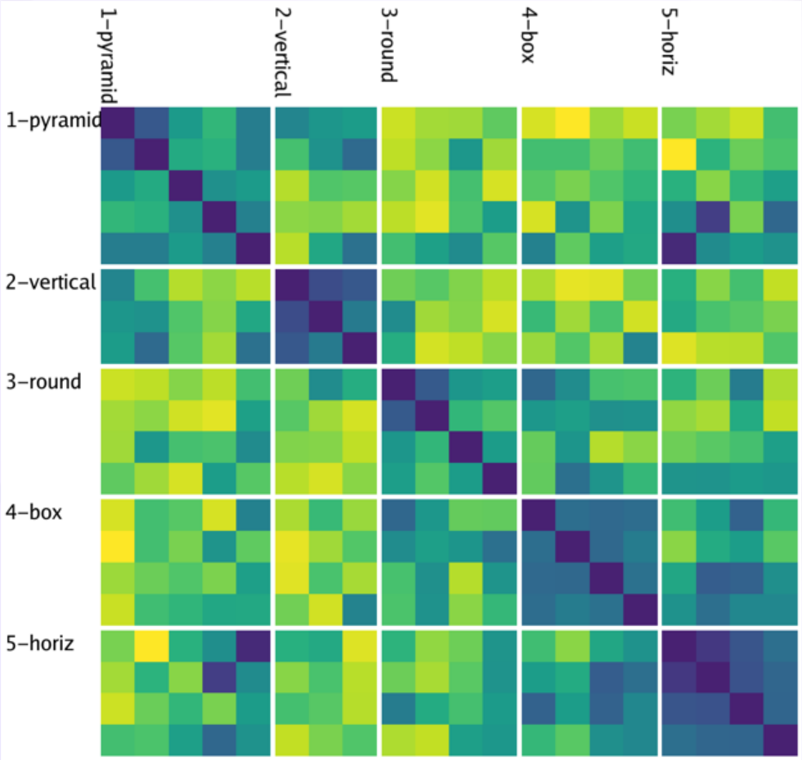
Model vs. Humans

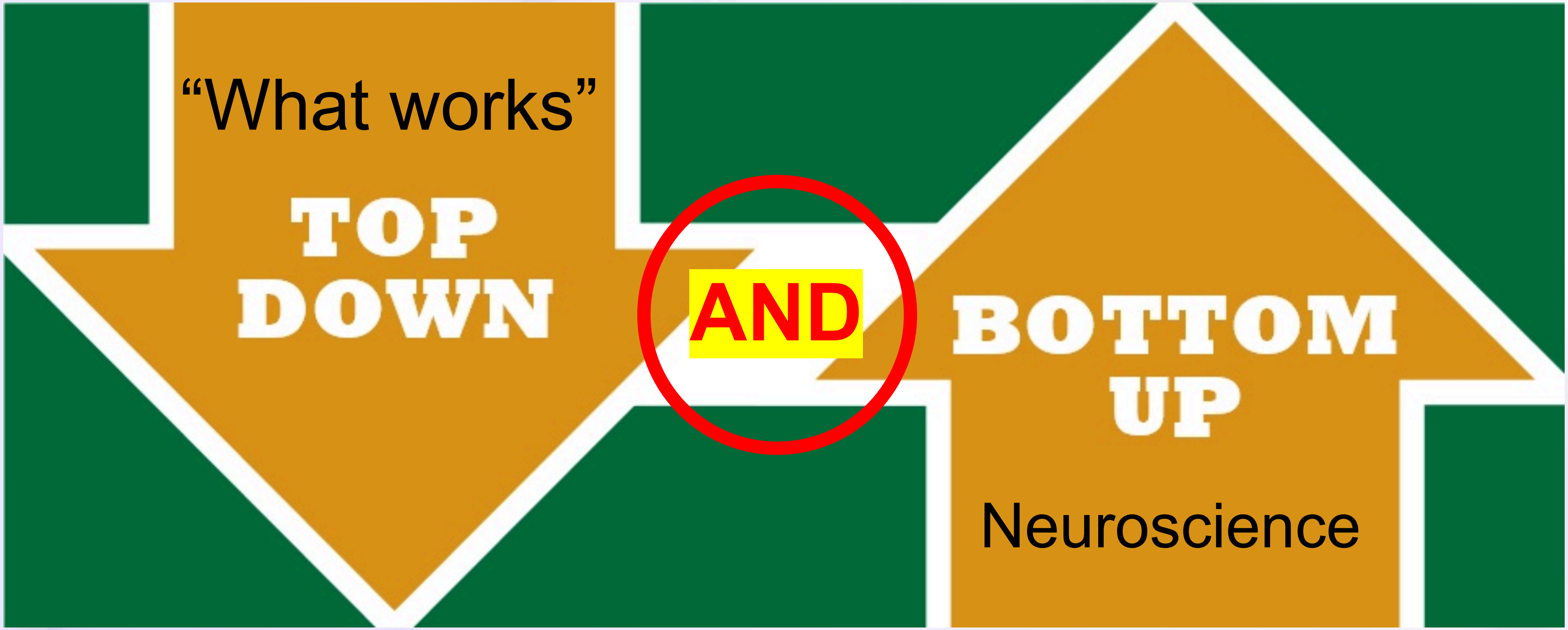
Which pair is more similar in terms of *overall shape?*

Model

People

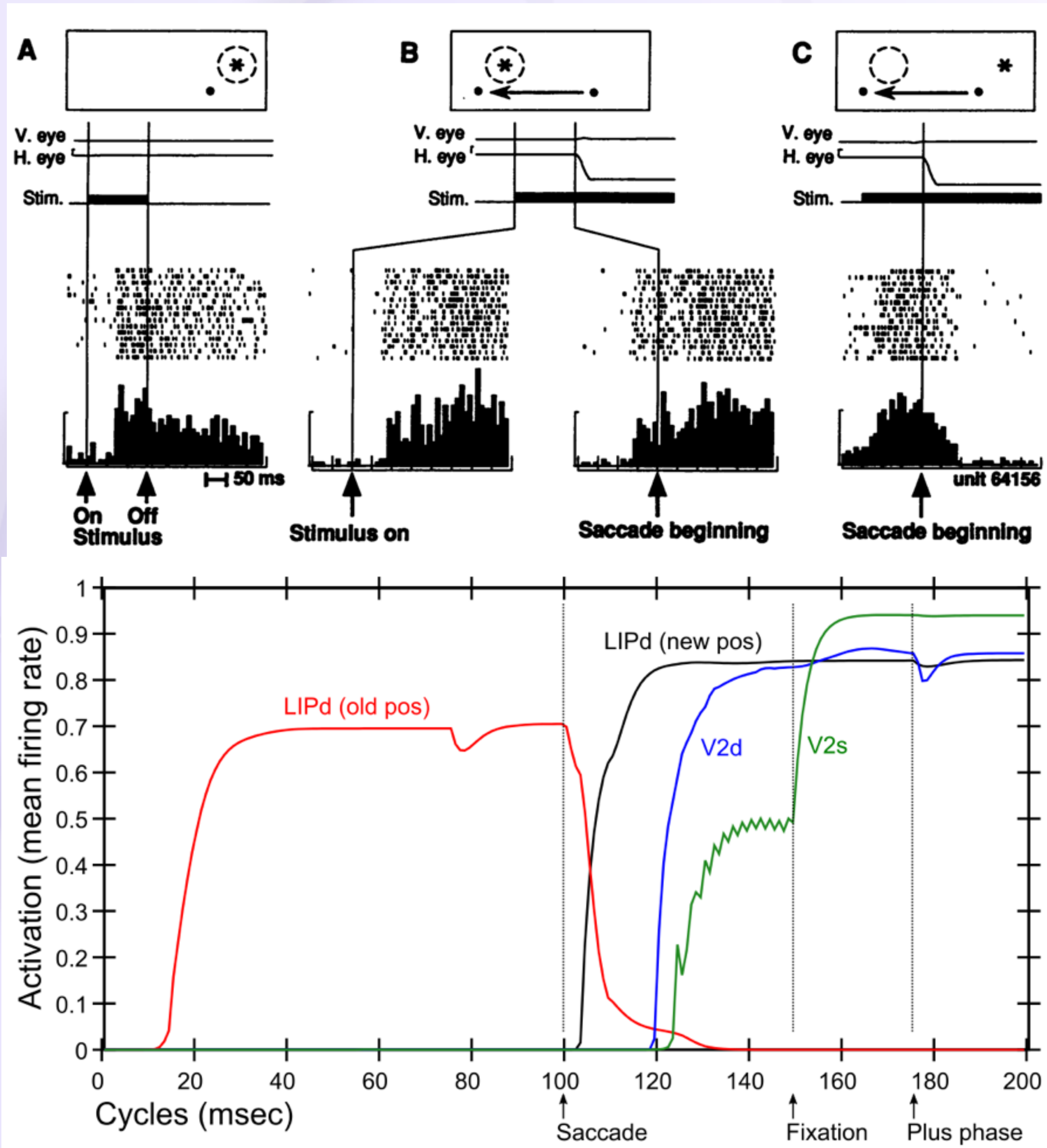






Deep predictive learning:
- Works
- Fits with lots of biology
- Extends to motor, cross-modal predictions

Predictive Remapping



Duhamel et al. (1992):

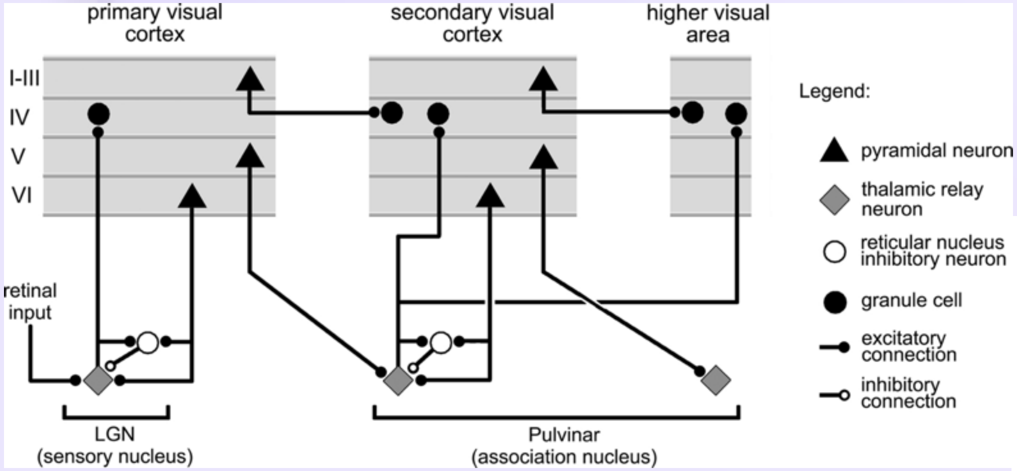
LIP neurons anticipate effect of saccade, start firing for new location before fixation lands (even before saccade)

Model:

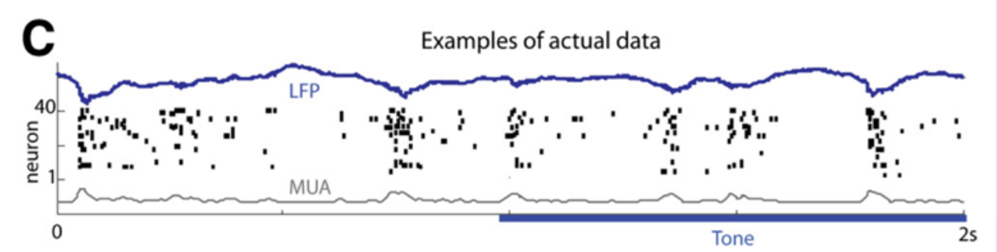
LIPd remaps first at high, abstract level, drives top-down remapping in lower areas – consistent with Cavanagh et al. (2010)

Key Biological Data

- Strong, synchronized, low-frequency modulation of cortex (at the alpha frequency).
- Specificity of alpha modulation to deep layers & thalamus, not superficial layers.
- Nature of deep-layer connectivity to pulvinar: numerous, weaker, plastic pathway (for generating a prediction) *and* sparse, strong, fixed pathway (for ground truth target).
- Synchronization of this strong pathway input with the alpha cycle.
- Broad connectivity of pulvinar with different visual pathways (afferent and efferent).
- Lack of direct bottom-up superficial projections into the deep layers (would short-circuit prediction), but presence of these projections top-down (beneficial).
- Bidirectional (top-down and bottom-up) connectivity between superficial layers.
- Early development of the *Where* (MT, LIP) pathway.
- Organization into three separable (yet highly interconnected) visual pathways, particularly a third putative *What*Where* integration pathway.



(Buffalo et al., 2011; van Kerkoerle et al., 2014; Shipp, 2003; VanRullen & Koch, 2003; Luczak et al., 2013..)



Conclusions

- Peculiar features of connectivity between cortex and thalamus support form of predictive learning (many diffs from Friston etc).
- Computational model shows that predictive learning from raw visual “movies” self-organizes abstract categorical object representations (based on *shape*, not texture!)

Key Diffs From Friston / Bayes

Friston et al: errors go up, predictions come down

Us: *full* activation goes up & down, predictions go to pulvinar, *errors are temporal differences.*

Both models account for increased activity for unexpected outcomes.

Thanks To

CCN Lab

- Will Chapman
- Tom Hazy
- Seth Herd
- Kai Krueger
- Jessica Mollick
- Ananta Nair
- John Rohrlich
- Jake Russin
- Dean Wyatte
- Maryam Zolfaghar

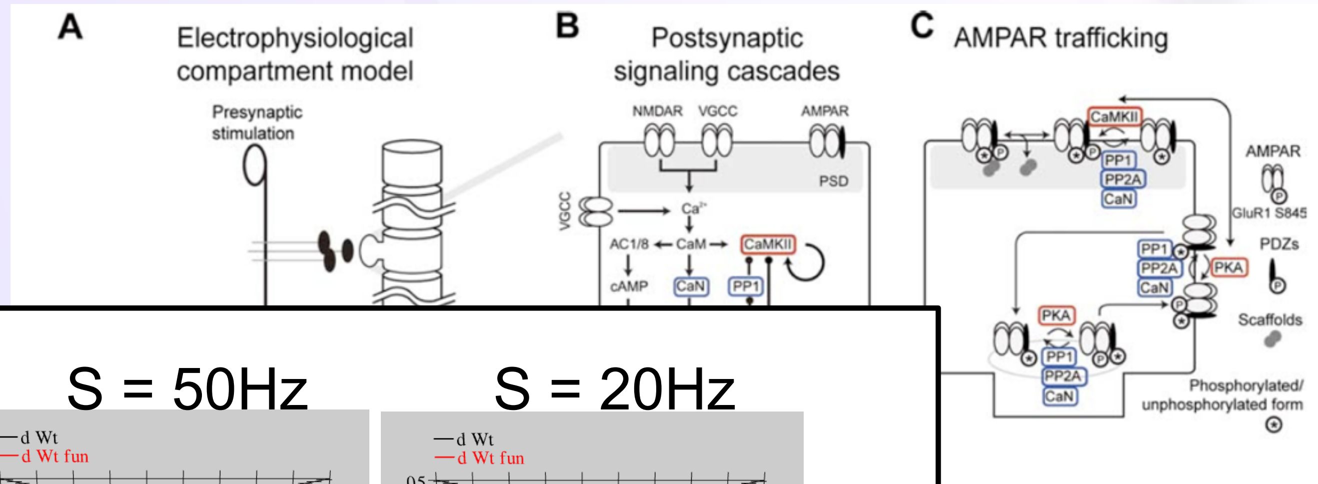
Collaborators

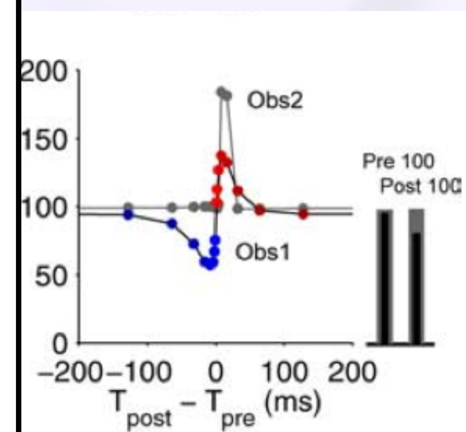
- Jonathan Cohen
- Alex Petrov
- Christian Lebiere
- Tim Curran
- David Sheinberg

Funding

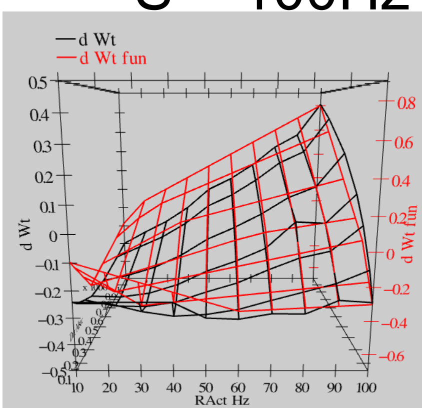
- **ONR** – Hawkins & McKenna

Synaptic Plasticity: XCAL Model
(reduction of Urakubo et al, 2008 STDP model)

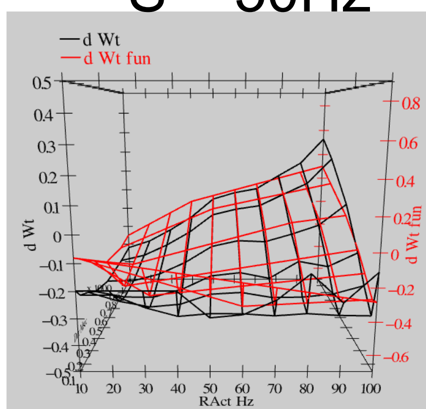




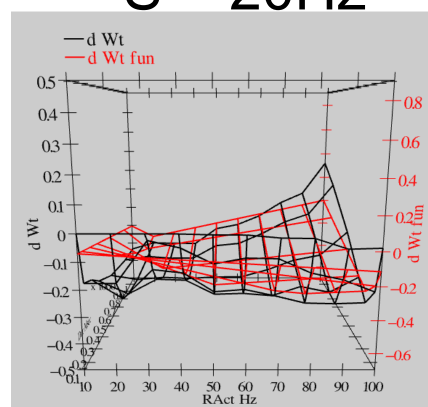
S = 100Hz



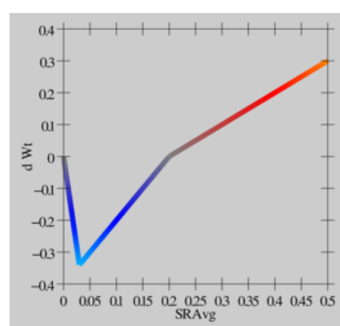
S = 50Hz



S = 20Hz

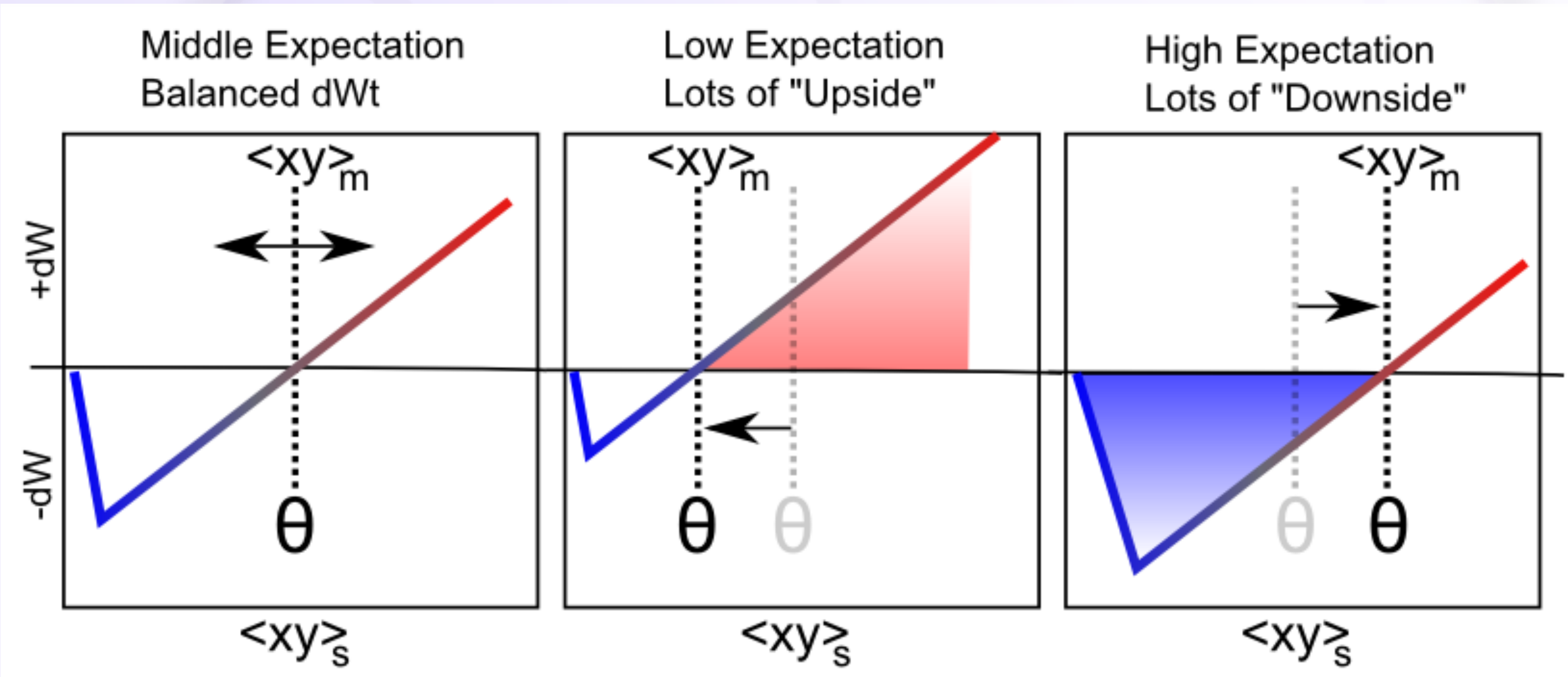


dW = f(send * recv) =
(spike rate * duration)



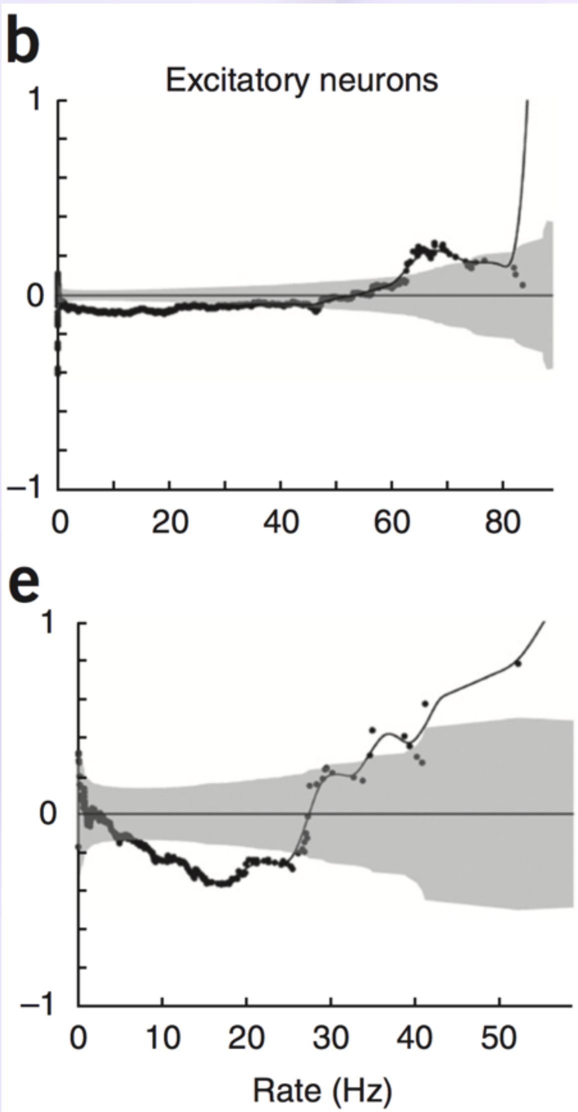
r=.894

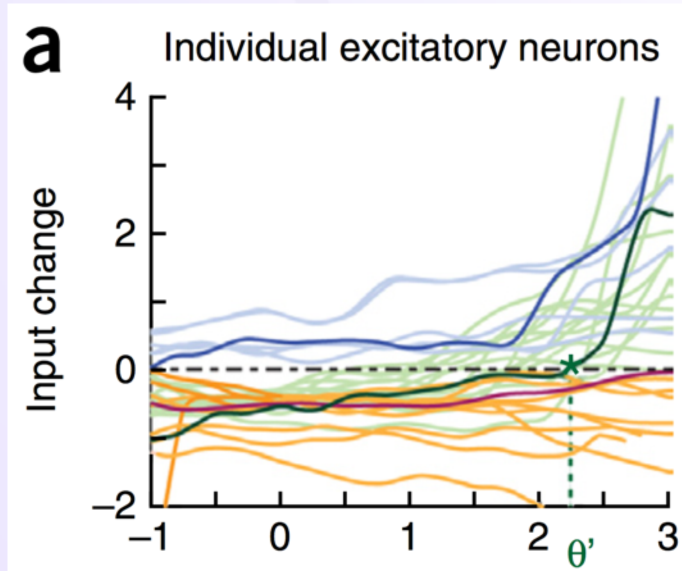
Floating Threshold = Medium Term Synaptic Activity (Error-Driven)

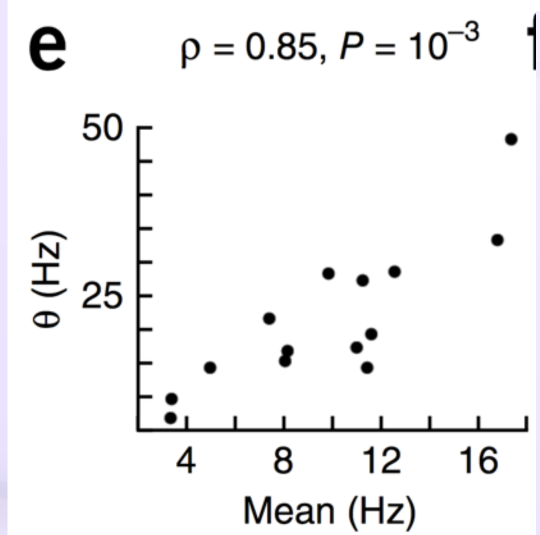


dW = Outcome – Expectation = $\langle xy \rangle_s$ - $\langle xy \rangle_m$

Evidence of Dynamic Thresholds
(Lim, McKee, Woloszyn et al., 2015)



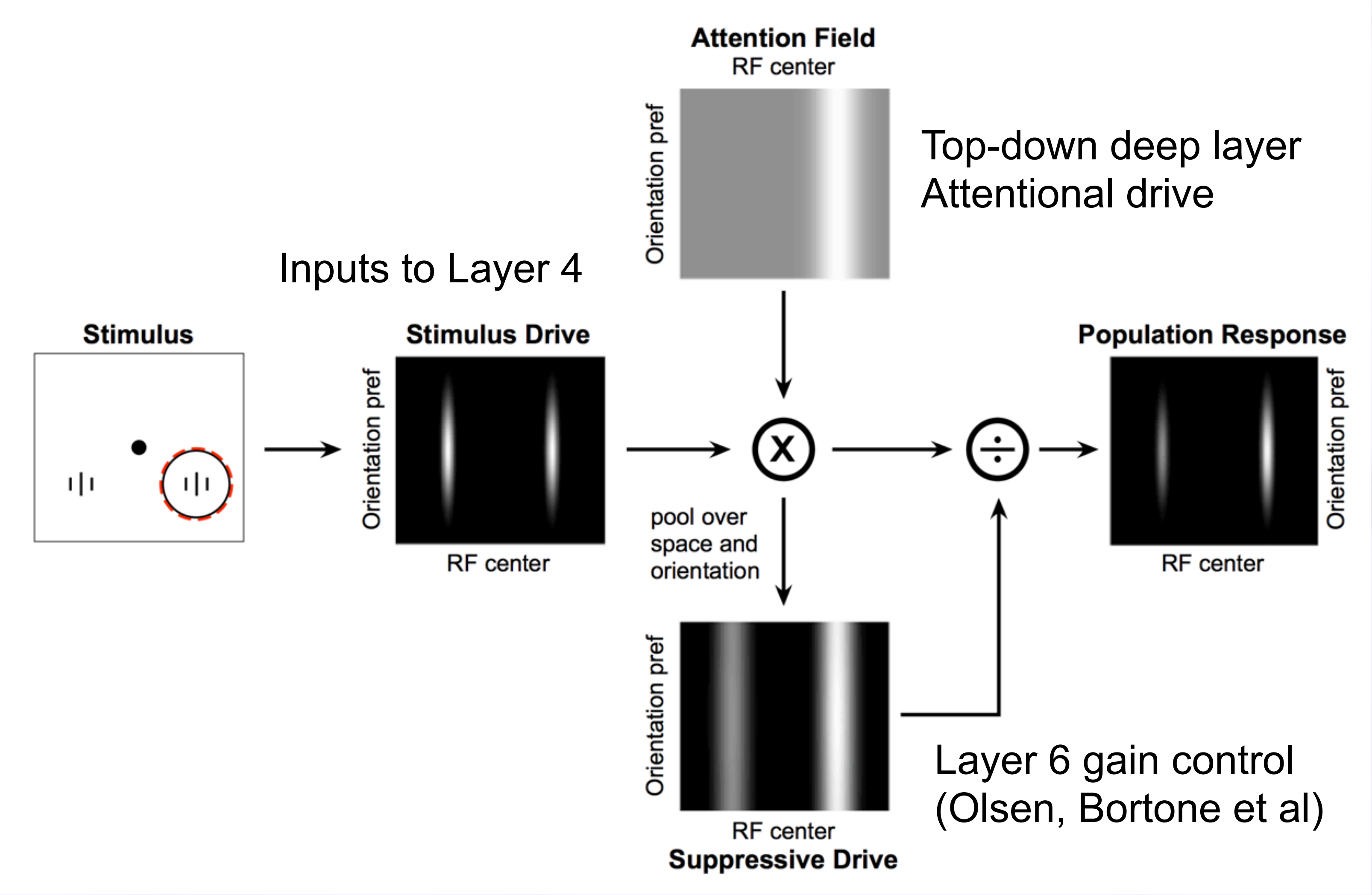




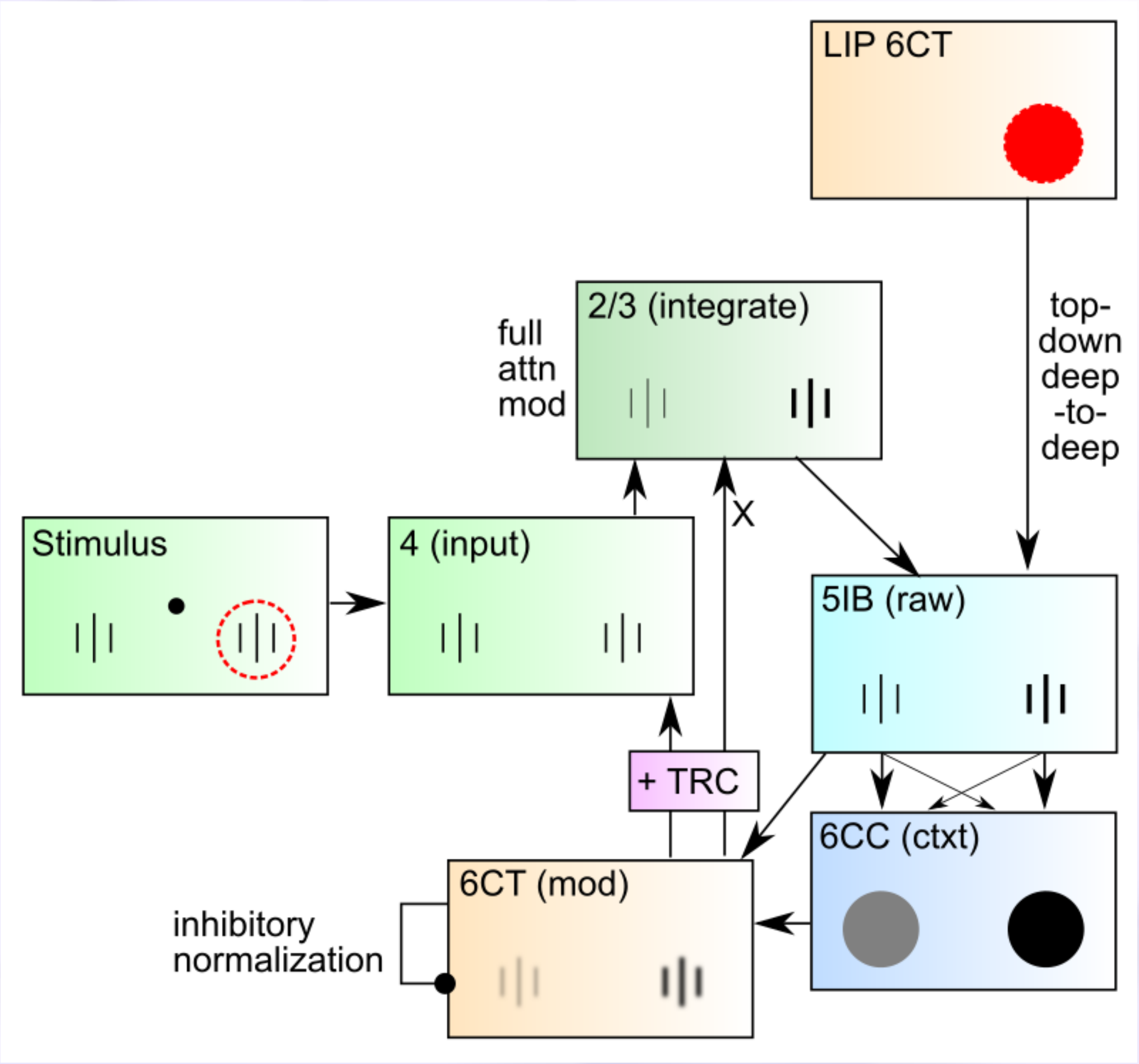
b = passive viewing
e = active task

Threshold changes dynamically on a rapid time scale, as a function of short-term activity level!

Deep Attentional Dynamics
(Reynolds & Heeger, 2009; Grossberg, 1999)



DeepLeabra Attentional Dynamics



Attentional Dynamics Results