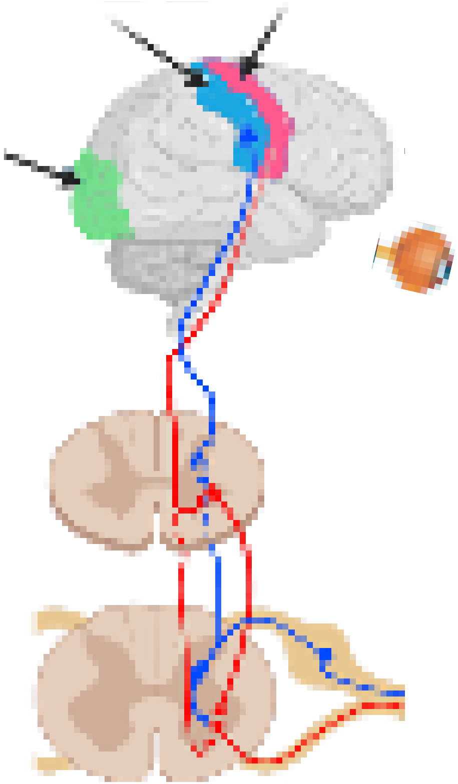
NX-414: Brain-like computation and intelligence

Martin Schrimpf

Lecture 4, March 12

Motivation for task-driven models

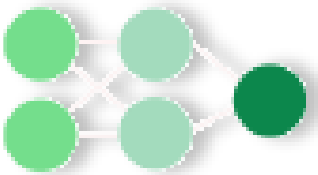
- We have seen that external information can be efficiently represented in the brain
- We have also considered the first “*representation learning model*” of this class: sparse coding
- Sparse coding is a powerful model and e.g. predicts simple cells for vision
- However, it’s based on a reconstruction loss, not a computational task …
- These kinds of approaches typically work best for well-parametrized stimulus regimes, but do not work well for many ecological behaviors



Normative frameworks

Information theoretic

e.g. sparse coding,
redundancy reduction,
mutual information …



Utilitarian

e.g. recognize objects,
chase prey, navigate …

Recurrent neural networks and path integration

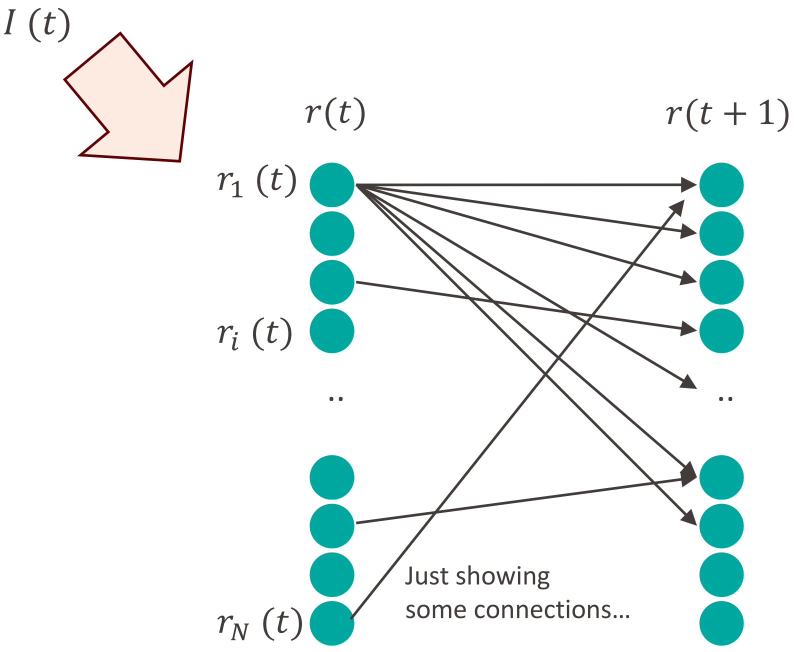
Today's questions:

1. How to *engineer* neural systems for path integration? -> ring attractors
2. How to *learn* path integration from scratch?
3. Are the solutions related?

Recap: path integration is a fundamental ability that depends on accumulating *velocity* signals (from the vestibular, proprioceptive …senses) to form a representation where one is in space.

In mammals, head direction, grid and place cells have been implicated.

Recurrent neural network (RNN)

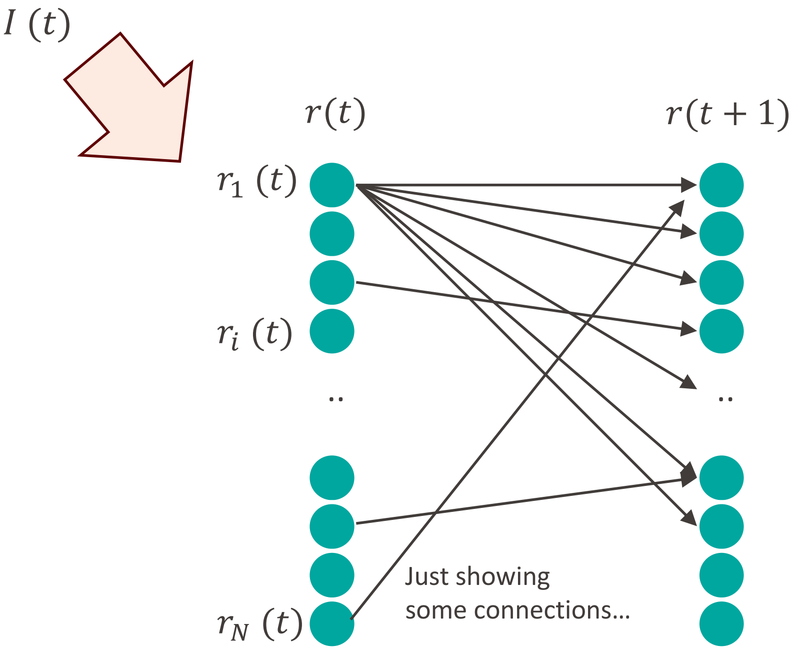


rate of neuron *i* at time *t*

Rate update equation

$$r(t+1) = W\ r(t) + I\ (t)$$

Recurrent neural network (RNN)



Just showing some connections…

Rate + membrane equation

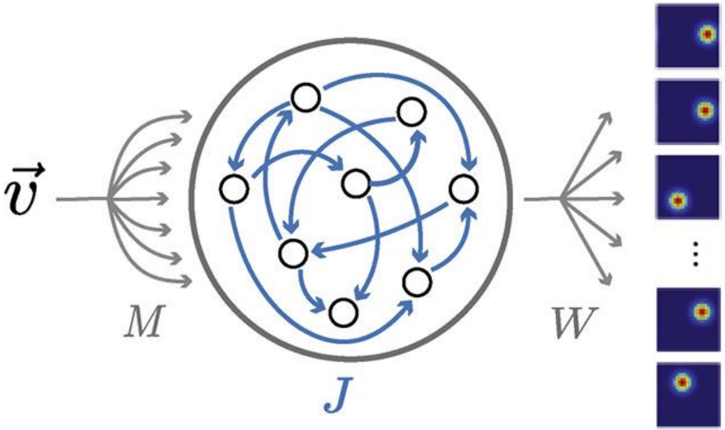
$$u(t+1) = Wr(t) + I(t)$$

$$\mathrm{r}(\mathrm{t}+1) = \sigma(u(t+1))$$

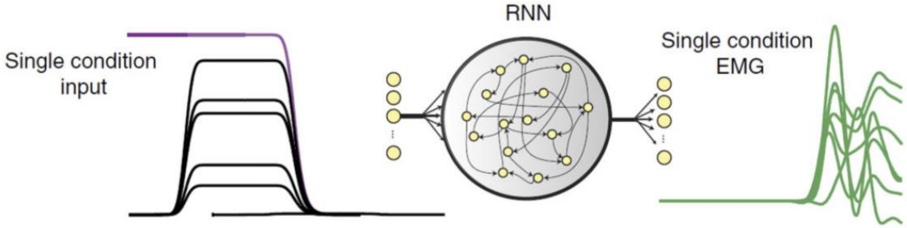
element-wise nonlinearity

Alternative depictions in the literature

Often (in the computational neuroscience literature) recurrent neural network models are depicted as in the two figure below.



Sorscher et al. Neuron 2022



Susillo et al. Nature Neuro 2015

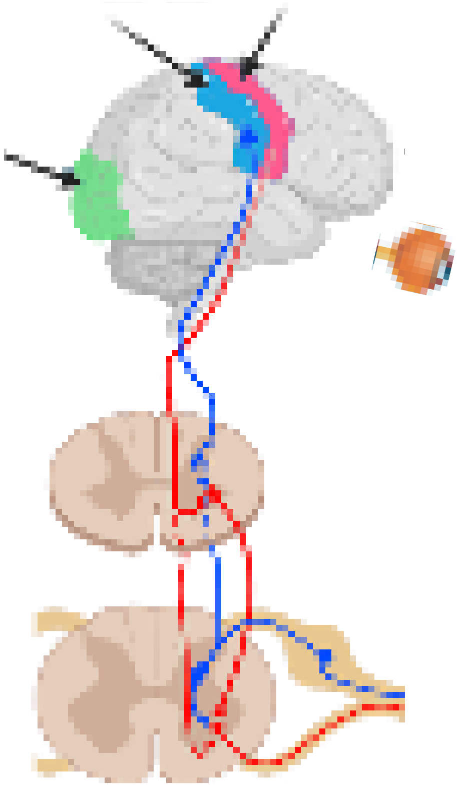
Here with input connectivity M, recurrent connectivity J and output connectivity W.

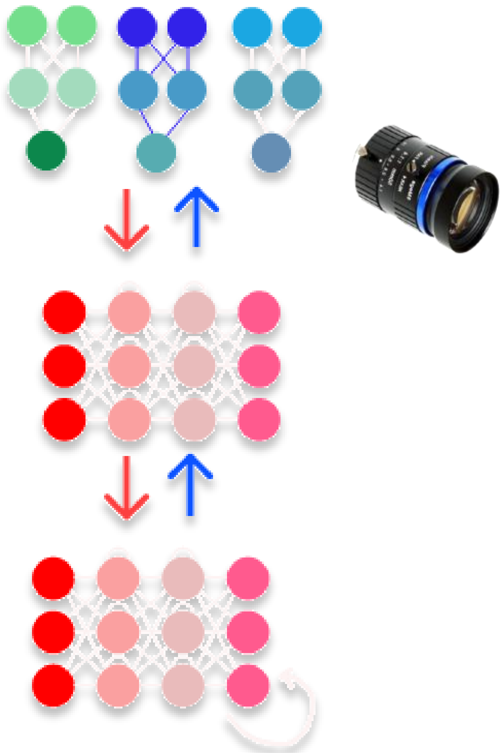
Intermediate summary

- Path integration is a key component of navigation
- Attractor models can perform path integration to explain (in a model) the head-direction, and grid cell system
- They make non-trivial predictions (see later)
- This is one of the first examples for a “*brain-like circuit for intelligence*” in this class.

Can such attractor models also be learned from the goal to navigate?

Biological Intelligence ↔ Artificial Intelligence



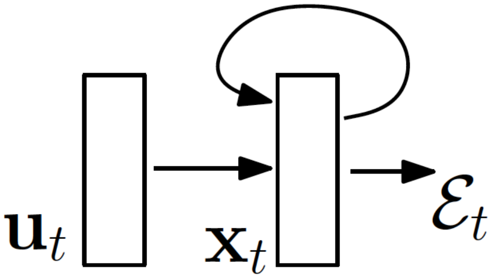


Hausmann & Marin-Vargas et al. 2021

Recurrent neural network (RNN)

Recurrent dynamics: $x_t = F(x_{t-1}, u_t, \theta)$

Generic RNN: $x_t = W_{rec}\sigma(x_{t-1}) + W_{in}u_t + b$ $\theta = W_{rec}, W_{in}, b$



For some task, where we want to predict: $\varepsilon_t = \mathcal{L}(x_t)$

The cost weights the individual costs per step: $\varepsilon = \sum_{t=1}^{T} \varepsilon_t$

How can we find parameters θ to minimize ε ?

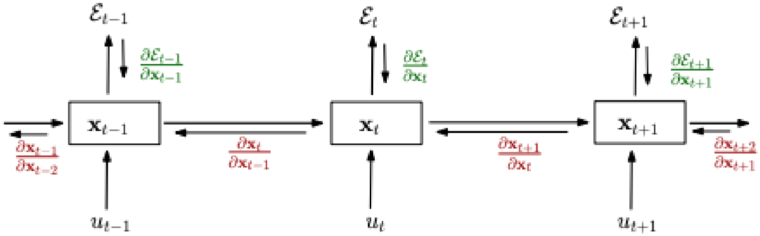
Pascanu, Mikolov & Bengio 2012

EPFL

Training recurrent neural networks

Backpropagation Through Time (BPTT)

$$x_t = W_{rec}\sigma(x_{t-1}) + W_{in}u_t + b$$



$\mathcal{E}_{t-1}$
$\mathcal{E}_t$
$\mathcal{E}_{t+1}$
$\frac{\partial \mathcal{E}_{t-1}}{\partial \mathbf{x}_{t-1}}$
$\frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t}$
$\frac{\partial \mathcal{E}_{t+1}}{\partial \mathbf{x}_{t+1}}$
$\mathbf{x}_{t-1}$
$\mathbf{x}_t$
$\mathbf{x}_{t+1}$
$\frac{\partial \mathbf{x}_{t-1}}{\partial \mathbf{x}_{t-2}}$
$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_{t-1}}$
$\frac{\partial \mathbf{x}_{t+1}}{\partial \mathbf{x}_t}$
$\frac{\partial \mathbf{x}_{t+2}}{\partial \mathbf{x}_{t+1}}$
u_{t-1}
u_t
u_{t+1}

$$\frac{\partial \varepsilon}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \varepsilon_t}{\partial \theta}$$

Total loss gradient is the sum of gradients at each step.

Using the chain rule: $$\frac{\partial \varepsilon_t}{\partial \theta} = \sum_{k=1}^{t} \left(\frac{\partial \varepsilon_t}{\partial x_t} \frac{\partial x_t}{\partial x_k} \frac{\partial^+ x_k}{\partial \theta} \right)$$

How is loss ε impacted by hidden state x

How does θ at step k impact later steps $t > k$

Treat x_{k-1} as constant with respect to differentiating θ

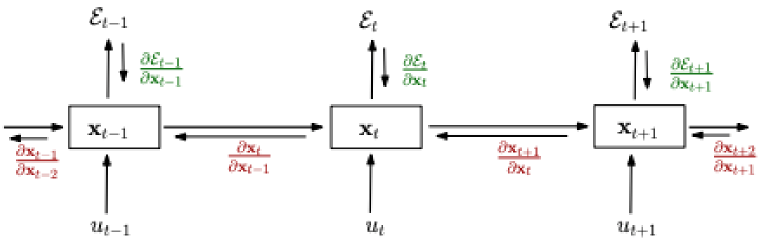
Do not propagate gradients beyond this depth

Pascanu, Mikolov & Bengio 2012

Training recurrent neural networks

Backpropagation Through Time (BPTT)

$$x_t = W_{rec}\sigma(x_{t-1}) + W_{in}u_t + b$$



$\mathcal{E}_{t-1}$
$\frac{\partial \mathcal{E}_{t-1}}{\partial \mathbf{x}_{t-1}}$
$\mathcal{E}_t$
$\frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t}$
$\mathcal{E}_{t+1}$
$\frac{\partial \mathcal{E}_{t+1}}{\partial \mathbf{x}_{t+1}}$
$\mathbf{x}_{t-1}$
$\mathbf{x}_t$
$\mathbf{x}_{t+1}$
$\frac{\partial \mathbf{x}_{t-1}}{\partial \mathbf{x}_{t-2}}$
$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_{t-1}}$
$\frac{\partial \mathbf{x}_{t+1}}{\partial \mathbf{x}_t}$
$\frac{\partial \mathbf{x}_{t+2}}{\partial \mathbf{x}_{t+1}}$
u_{t-1}
u_t
u_{t+1}

Problems:

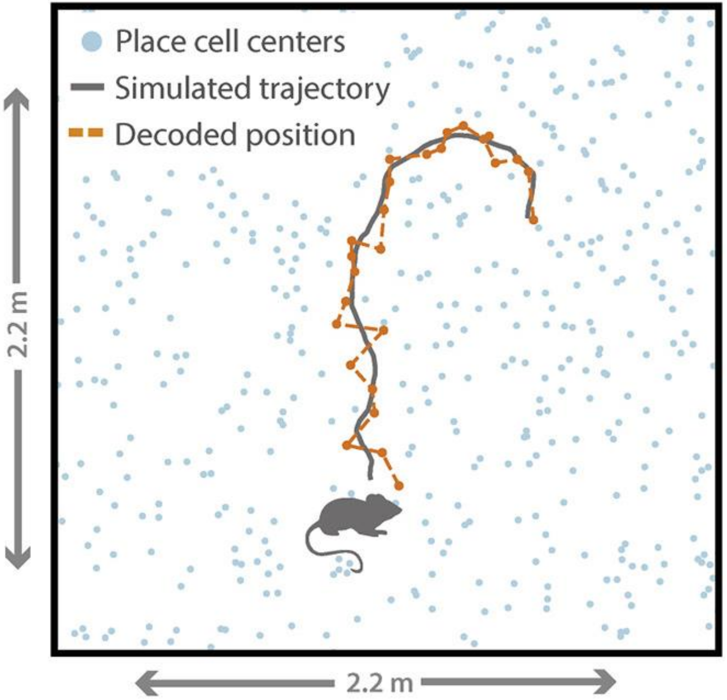
- Vanishing or exploding gradients
- Difficult to track long-range dependencies

Solutions:

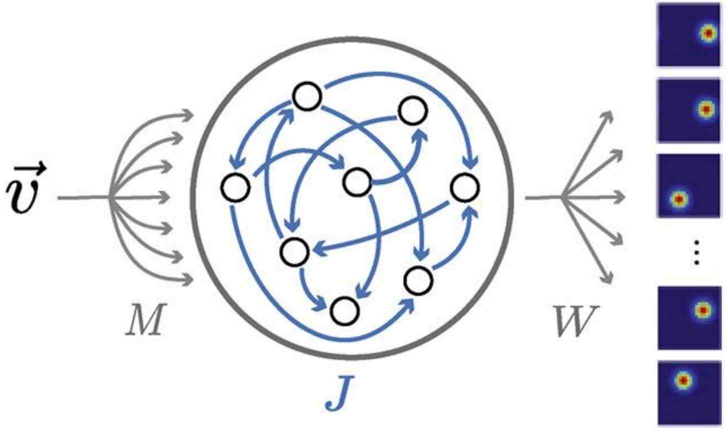
- Gating (LSTM, GRU)
- Feed-forward context integration (Transformers)

Pascanu, Mikolov & Bengio 2012

A model for path integration in mammals



place cell target $\hat{p}_i\ (t)$



Note that the model is directly trained to report place cells. Not really a normative task.

recurrency

$$r_i\ (t+1) = \sigma\left(\sum_{j=1}^{n} J_{ij} r_j\ (t) + M_{ix} v_x\ (t) + M_{iy} v_y\ (t)\right)$$

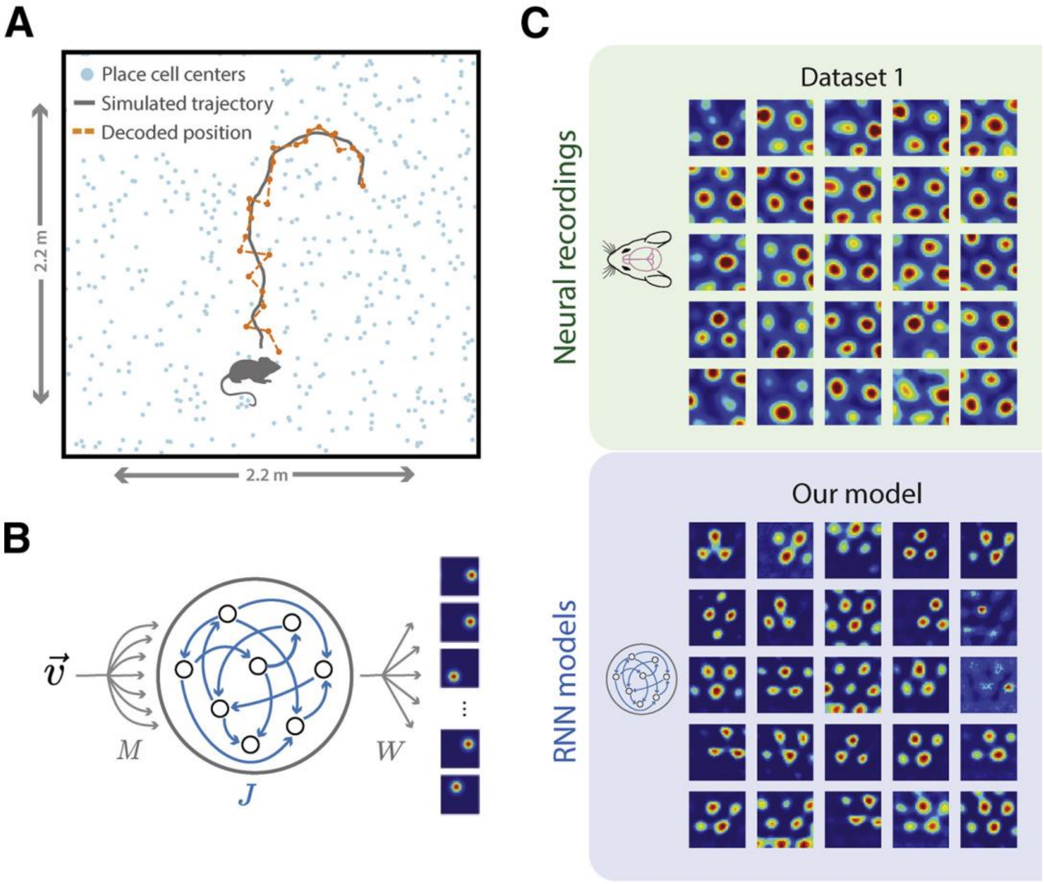
velocity inputs & weight matrices

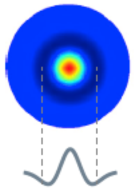
$$\hat{p}_i\ (t) = \sum_{j=1}^{n} W_{ij} r_j\ (t)$$

linear projection to place cells

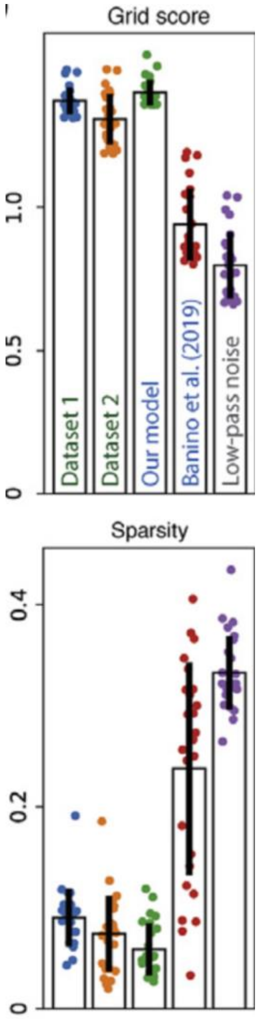
Sorscher & Mel et al. Neuron 2023

A model for path integration in mammals



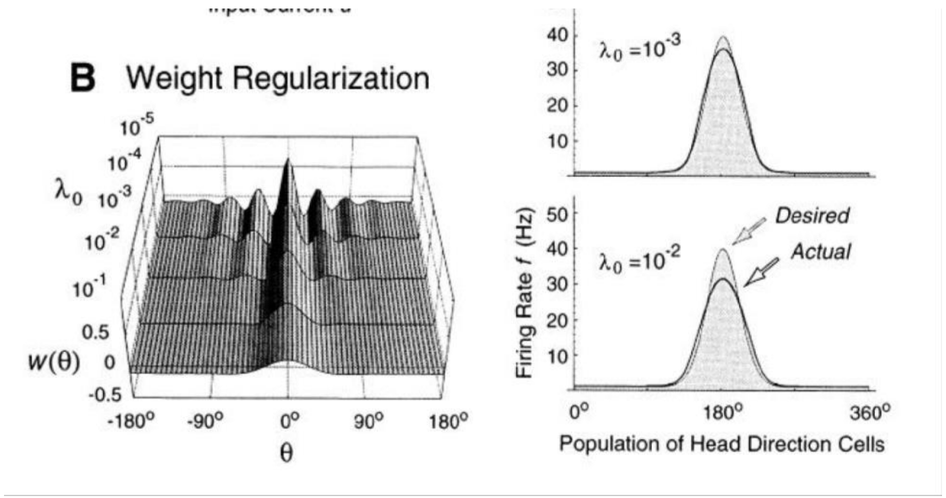


+ nonnegative



Sorscher & Mel et al. Neuron 2023

Reminder: “Mexican-hat” connectivity

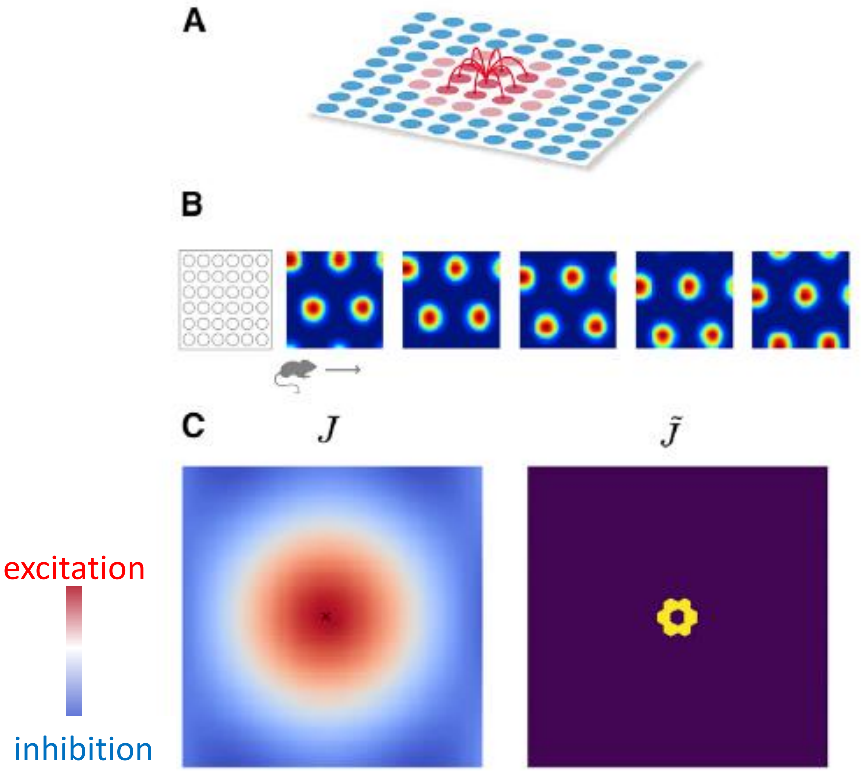


https://en.wikipedia.org/wiki/Sombrero

Local excitation, midrange inhibition

Mexican-hat connectivity in a hand-designed model

Idealized hand-designed model



Stable activity patterns on the neural sheet when the animal is at 5 successive positions in physical space.

Average outgoing connectivity profile:

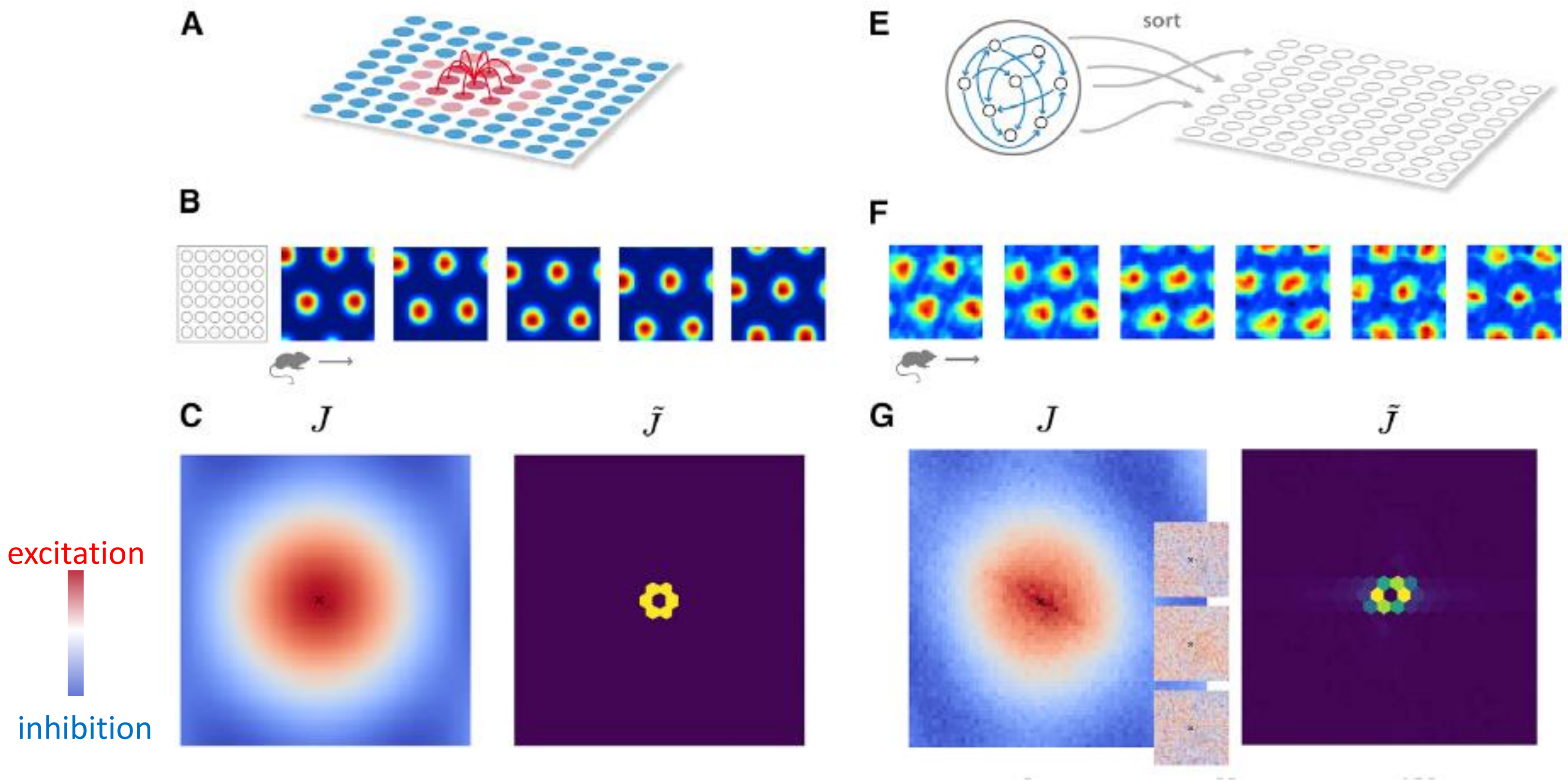
- Local excitatory connections (red)
- Long-range inhibitory connections (blue)
- Very local self-excitation (right, yellow)

Sorscher & Mel et al. Neuron 2023

Implicitly the model *learns* Mexican-hat connectivity

Idealized hand-designed model

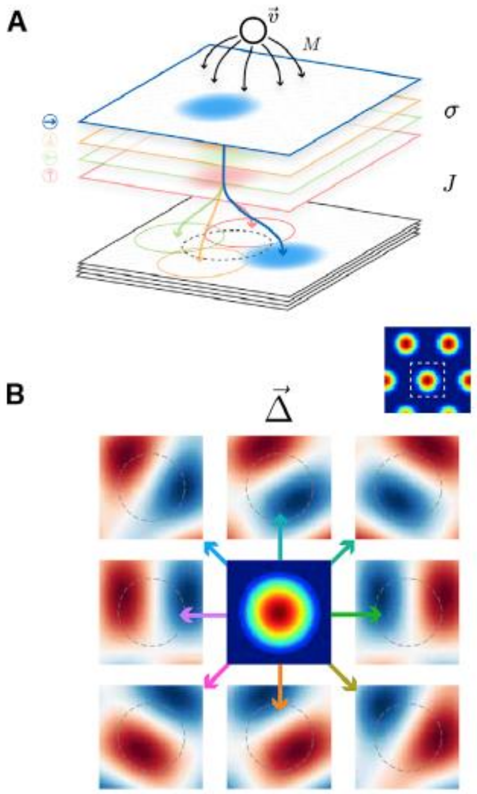
Learned model



Sorscher & Mel et al. Neuron 2023

Implicitly the model learns a shift-circuit

Idealized hand-designed model



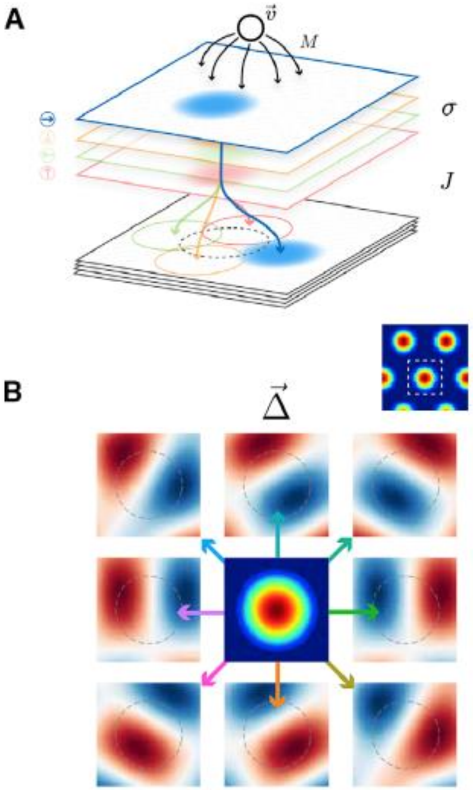
excitation

inhibition

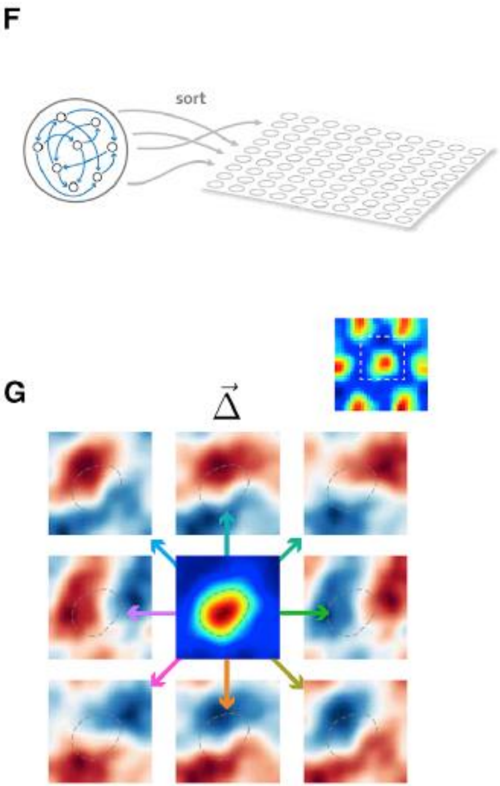
Sorscher & Mel et al. Neuron 2023

Implicitly the model learns a shift-circuit

Idealized hand-designed model



Learned model



excitation

inhibition

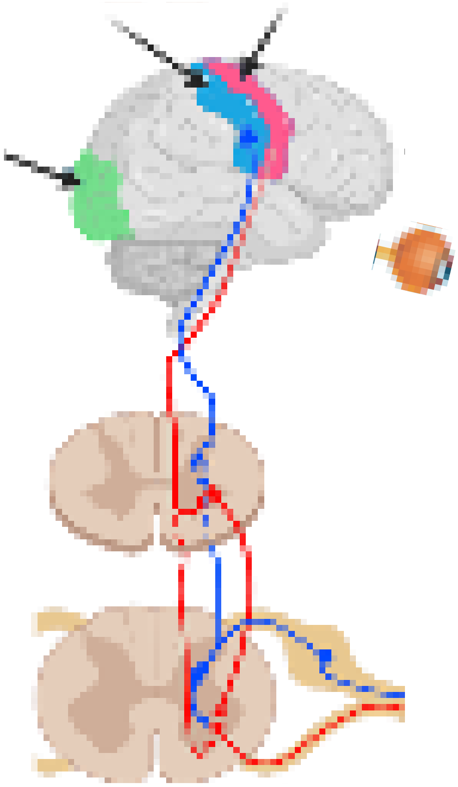
Sorscher & Mel et al. Neuron 2023

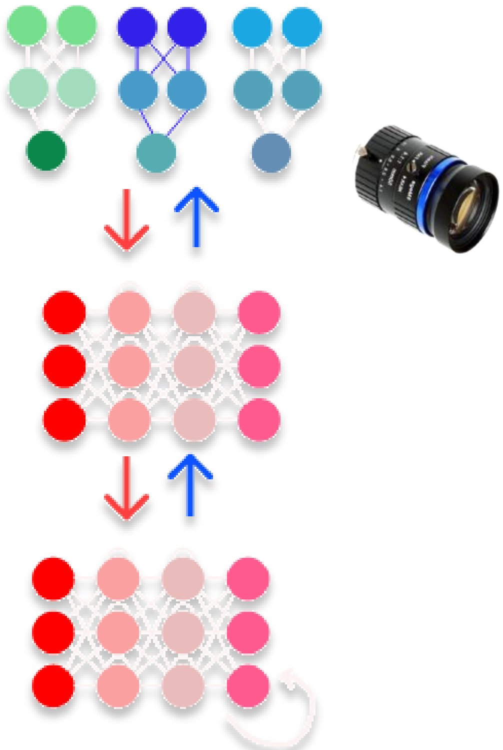
Take-home messages part 1

- First glimpse at *task-driven modeling*, we’ll see more in the next weeks
- Attractor models are powerful models of brain function (and make several non-trivial predictions that turn out to be true)
- Path integration is an important brain function and in mammals; the hippocampal formation supports this computation via specialized cell types
- We also highlighted recent circuits in Drosophila and zebrafish (last time)
- Attractor models can implement path integration, and learning to path integrate converges to similar solutions (with the right constraints)
- Attractor models are a first “brain-like circuit” in this class. Think about how this system computes vs., e.g., a CPU.

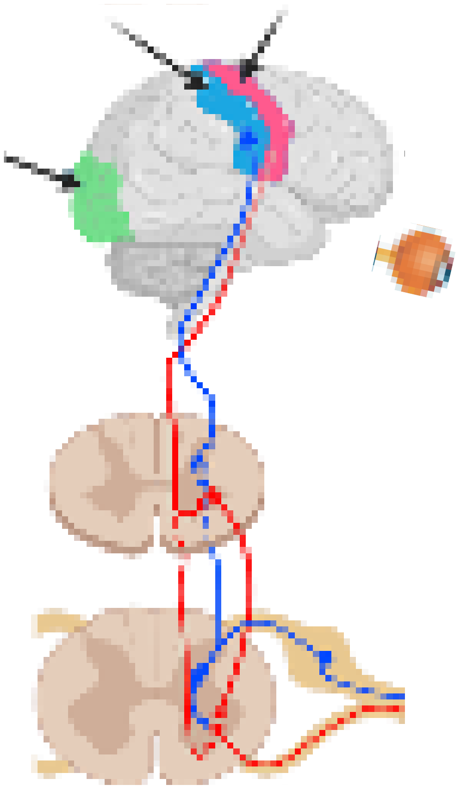
You will implement the Sorscher & Mel et al. model in the exercises!

Biological Intelligence ⟷ Artificial Intelligence





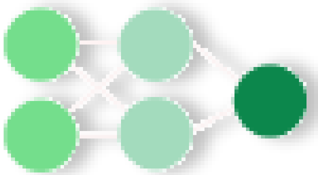
Hausmann & Marin-Vargas et al., 2021



Normative frameworks

Information theoretic

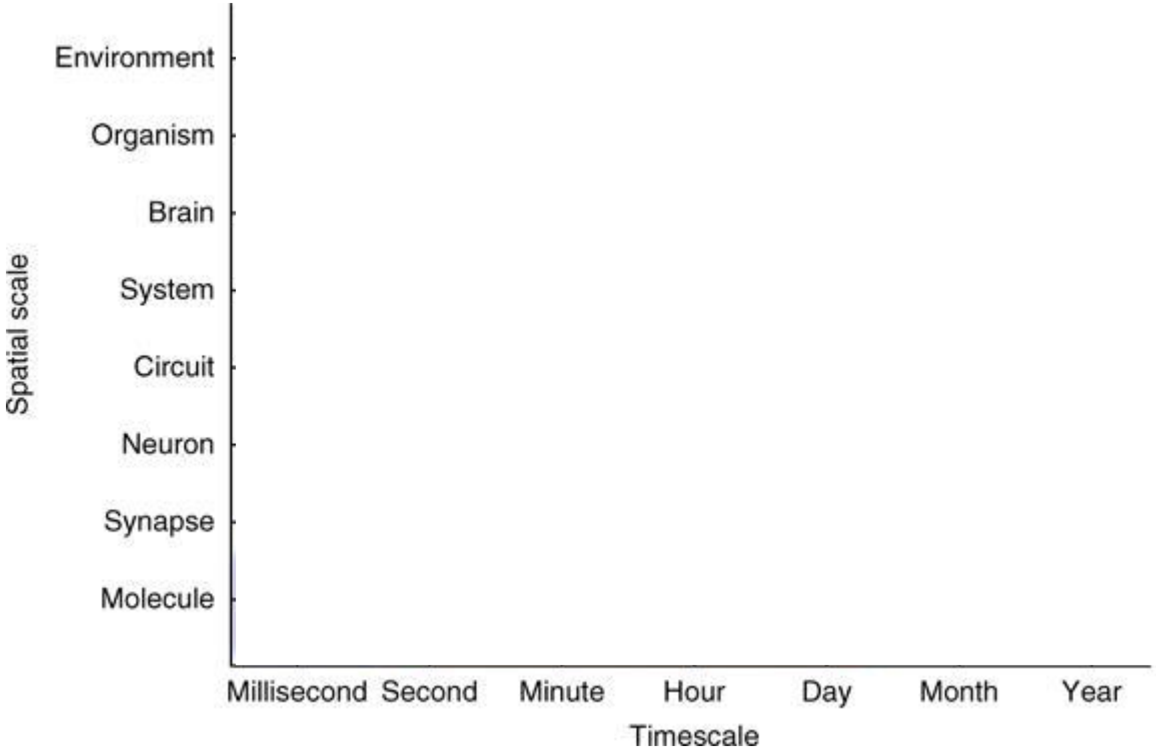
e.g. sparse coding, redundancy reduction, mutual information …



Utilitarian

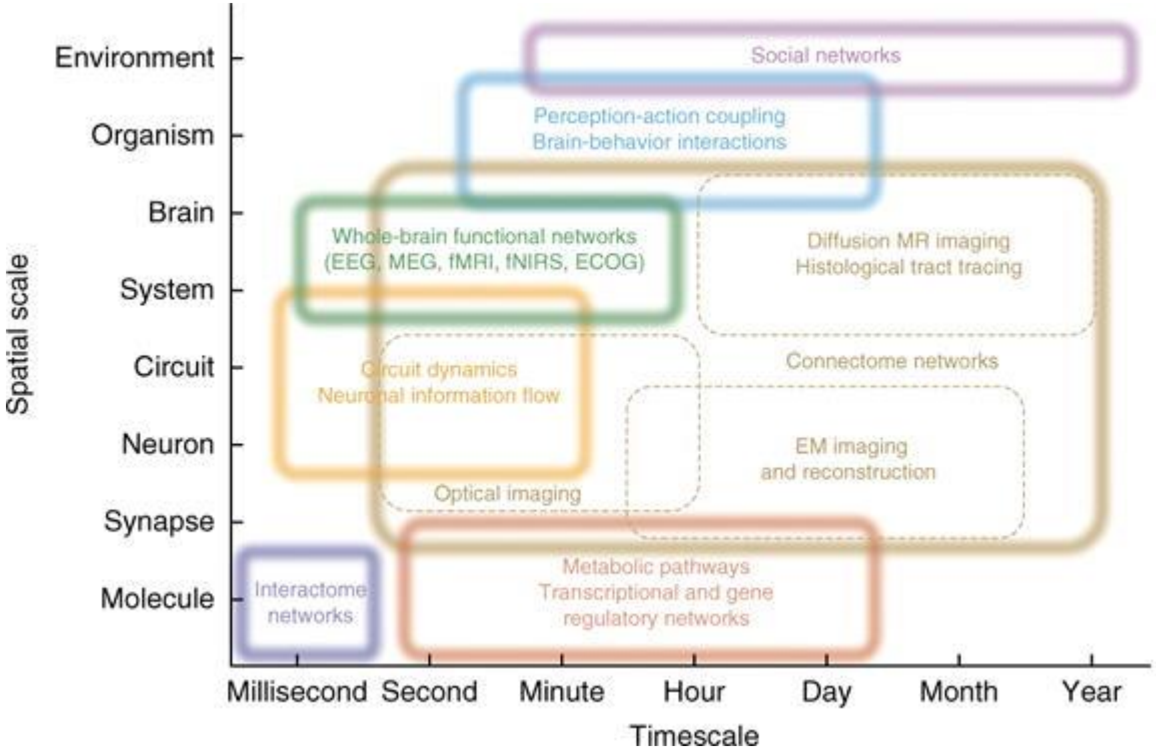
e.g. **recognize objects**, chase prey, navigate …

Temporal and spatial scales in neuroscience



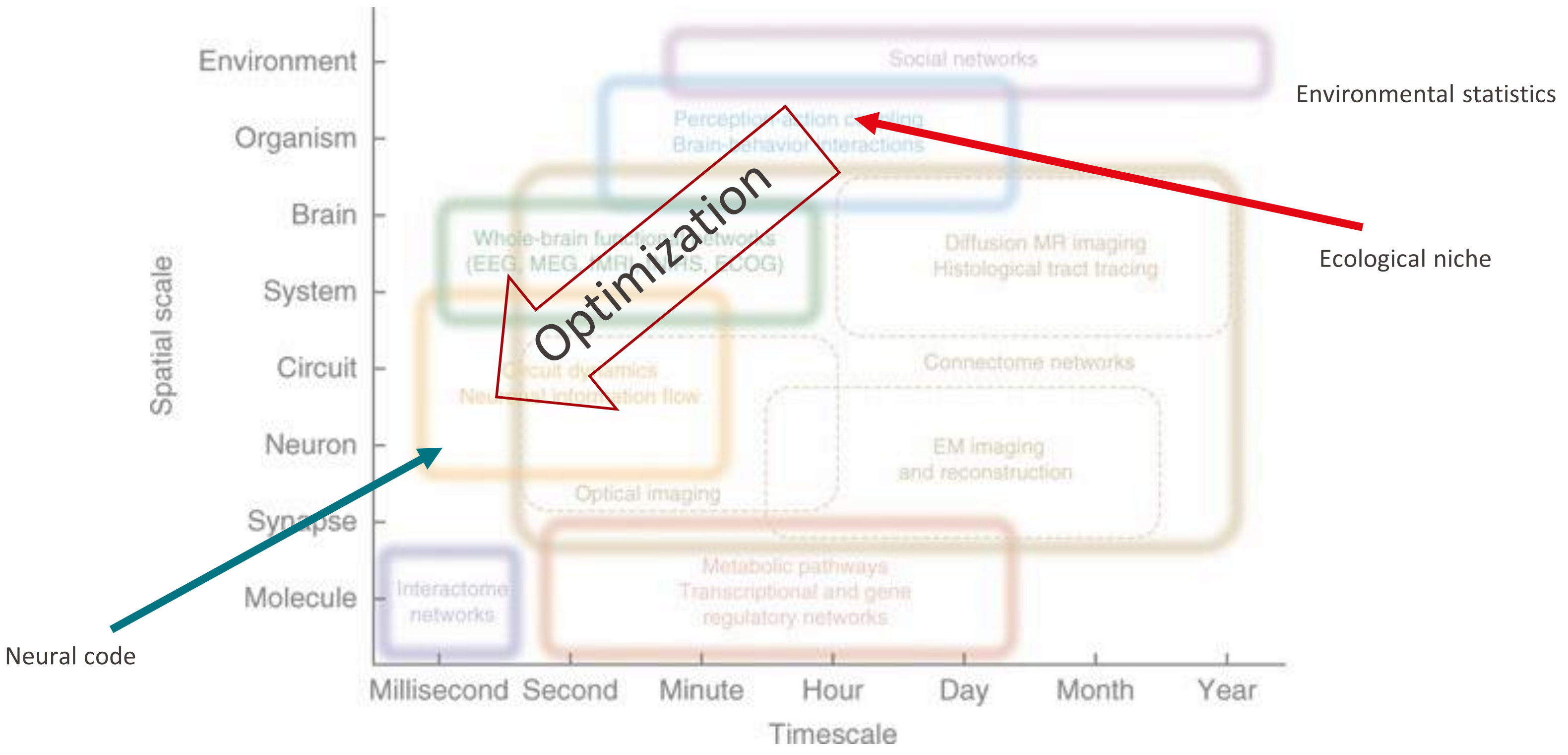
Basset, Nat Neuro 2017

Temporal and spatial scales in neuroscience



Basset, Nat Neuro 2017

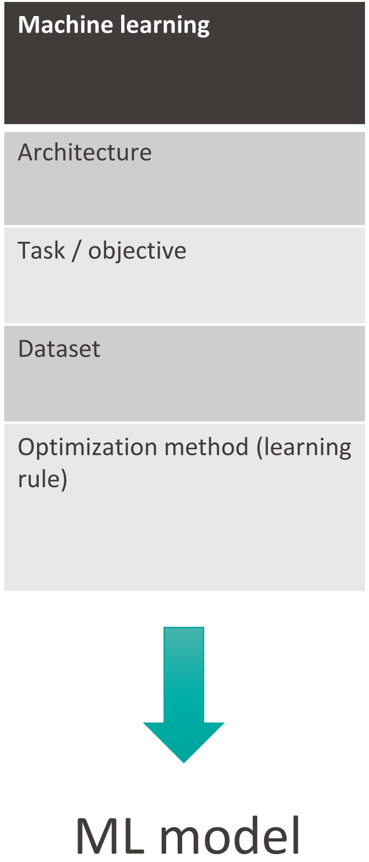
Task-driven modeling: linking behavior to circuits



Environmental statistics

Ecological niche

Neural code



Machine learning

Architecture

Task / objective

Dataset

Optimization method (learning rule)

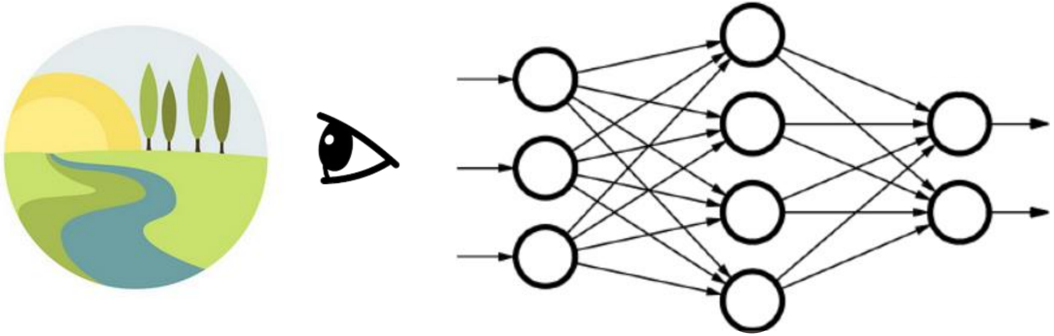
ML model

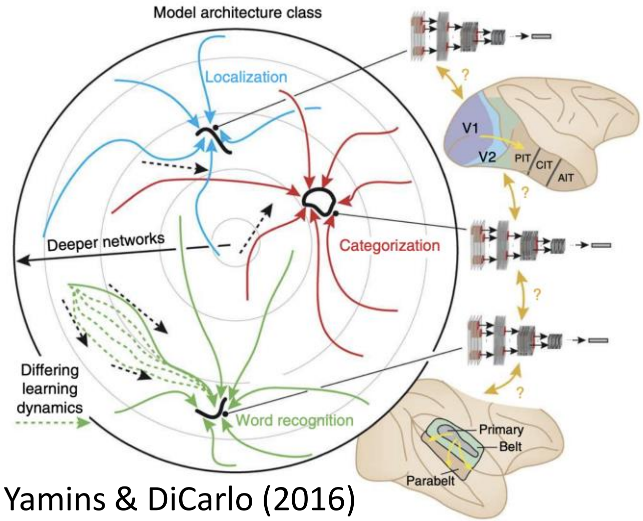
Machine learning	Neuroscience
Architecture	Circuits
Task / objective	Ecological niche
Dataset	Environment
Optimization method (learning rule)	Natural selection + synaptic plasticity

↓ ↓

ML model 🧠

Using deep neural networks as goal-driven models of a system





Yamins & DiCarlo (2016)

Vision: object recognition. Yamins & Hong et al. (2014), Schrimpf & Kubilius et al. (2018)

Audition: speech recognition, speaker & sound identification. Kell et al. (2018)

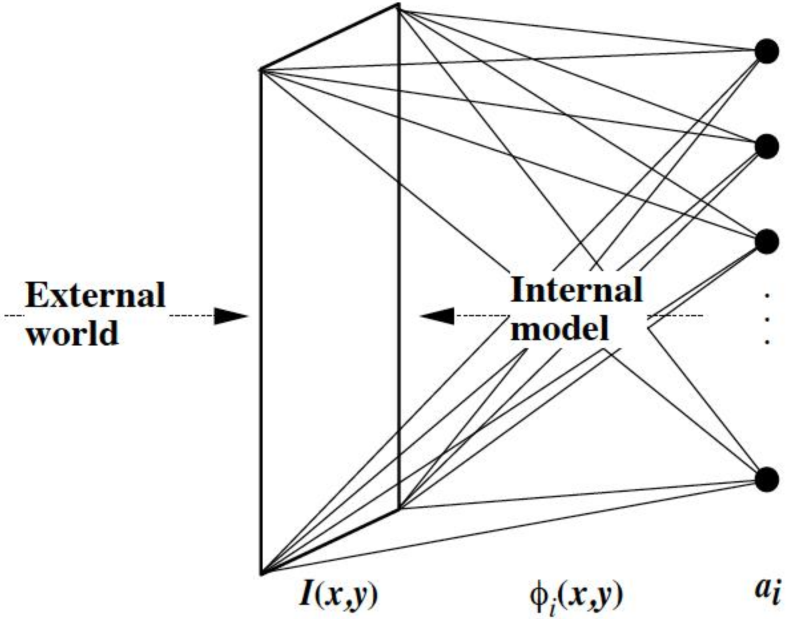
Somatosentation: shape recognition. Zhuang et al. (2017)

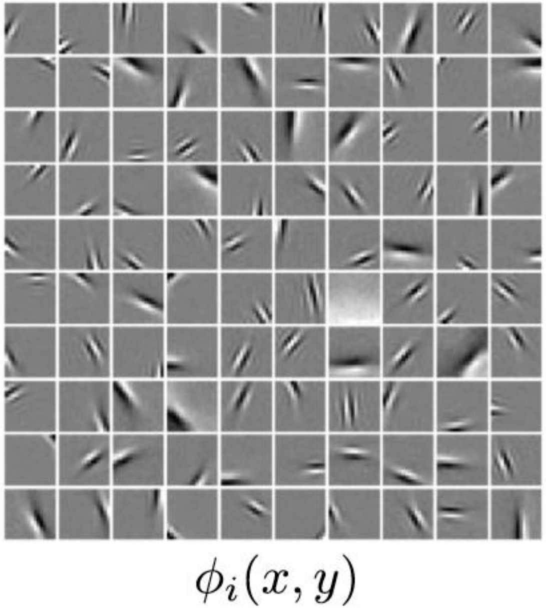
Language: next-word prediction. Schrimpf et al. (2021)

Decision making: context-dependent choice. Mante & Sussilo et al. (2013)

Proprioception: action recognition. Sandbrink et al. (2023)

Reminder: Explaining edge detectors with sparse autoencoding





$$I(x,y) = \sum_i a_i\,\phi_i(x,y) + \epsilon(x,y)$$

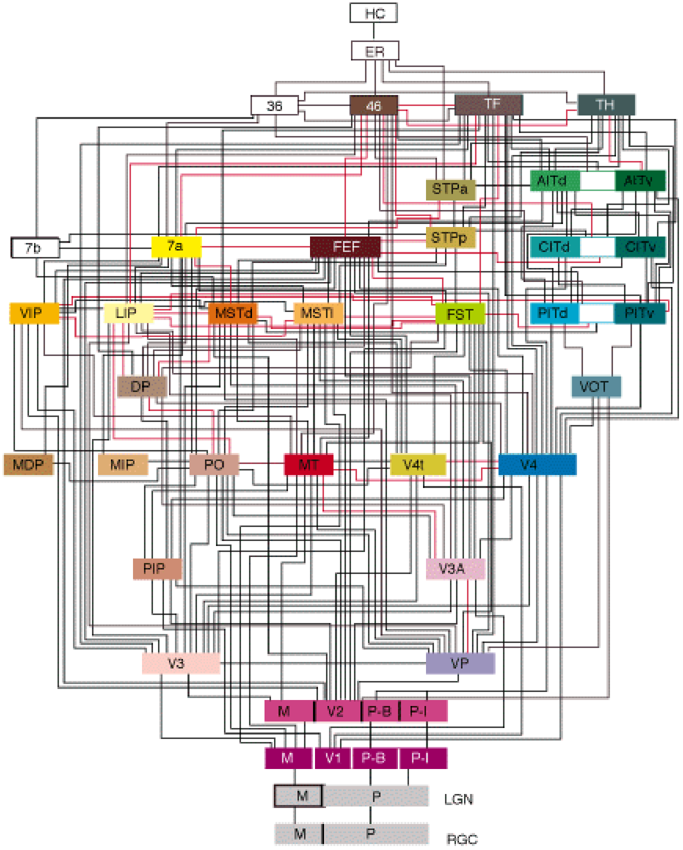
Olshausen & Field, 1996 Nature

Object recognition

- Recognizing objects seems easy, but
  - we can recognize objects among thousands of possibilities
  - we do so in the fraction of a second (Thorpe et al., 1996)
  - we do so despite tremendous variation (size, angle, …)

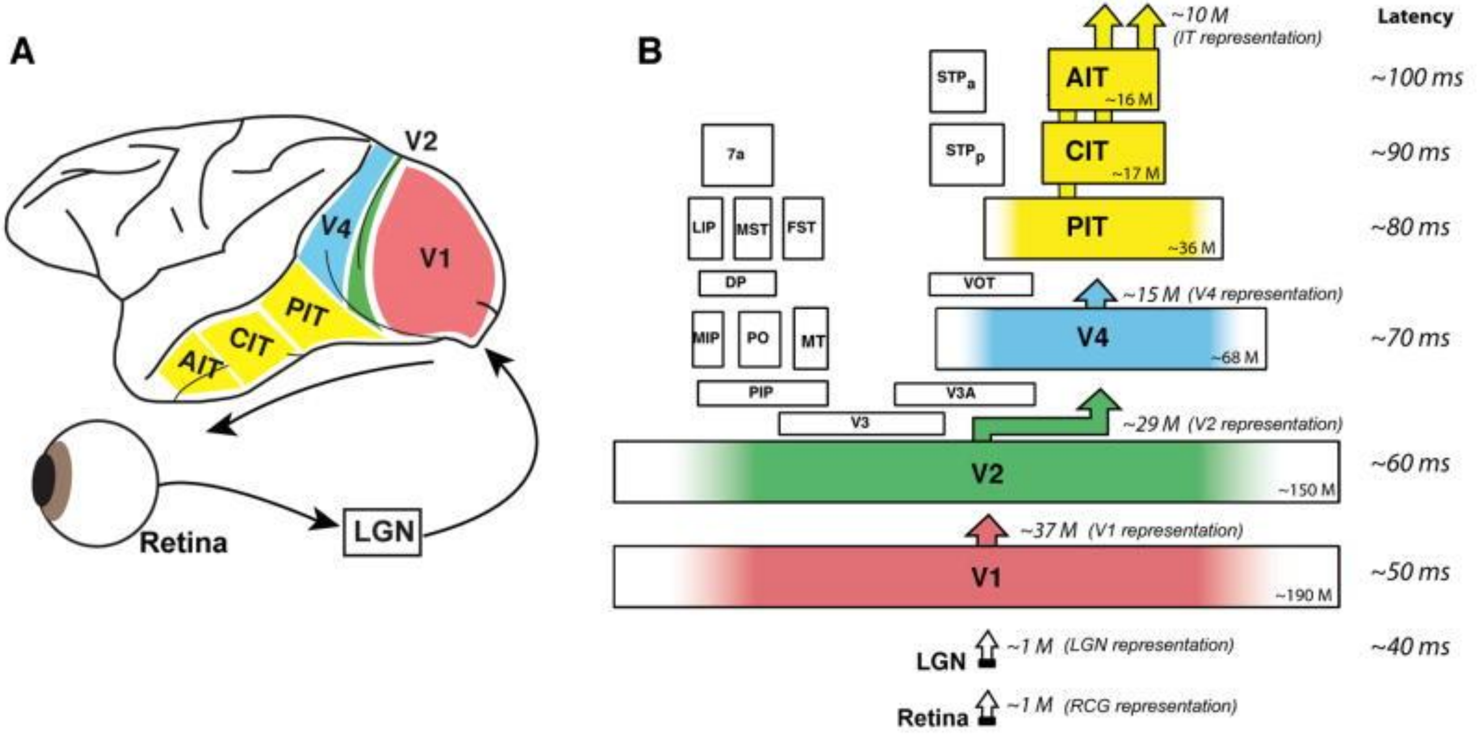
- Recognizing objects must be hard,
  - Half of the primate neocortex is devoted to vision (Felleman & Von Essen, 1991)
  - Despite all CV advances, machines still struggle with *robust* vision! I.e. on benchmarks like ImageNet, they are as good/better than humans but they are subject to adversarial robustness

Visual system



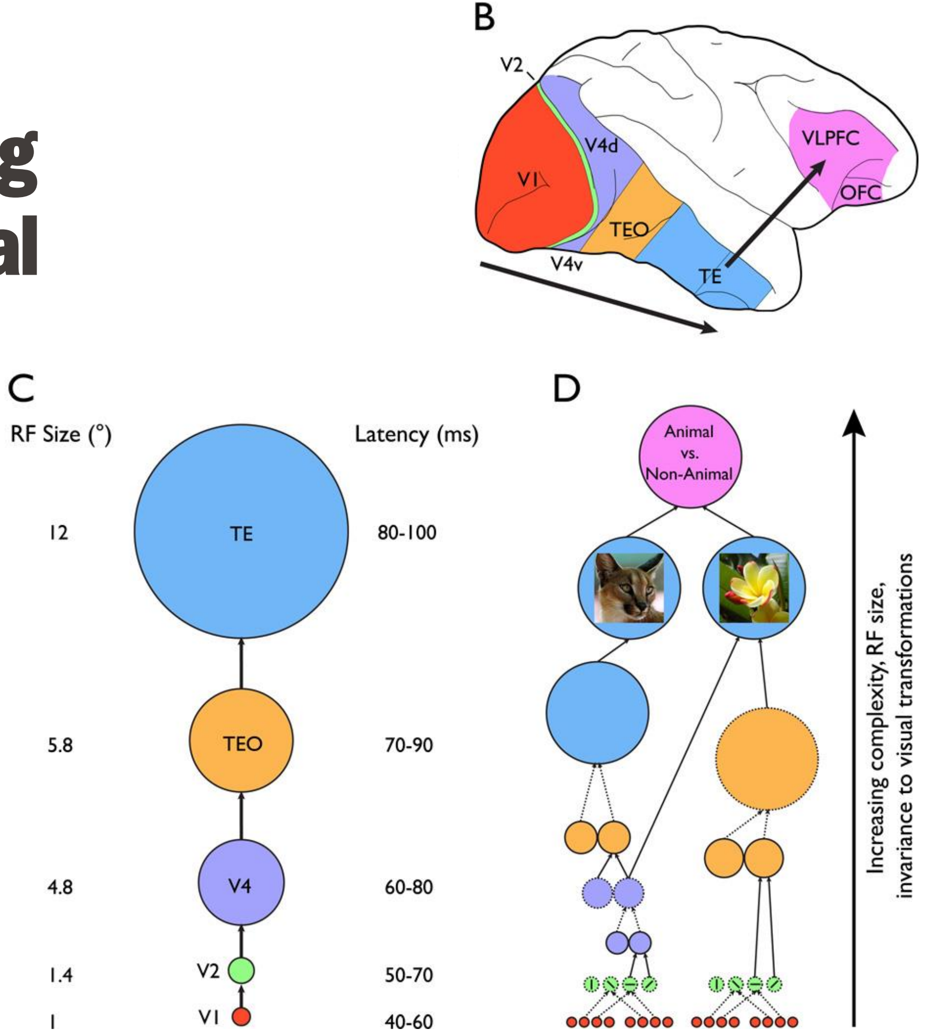
Felleman and Van Essen *Cerebral Cortex* 1991

Ventral visual pathway



DiCarlo, Neuron 2012

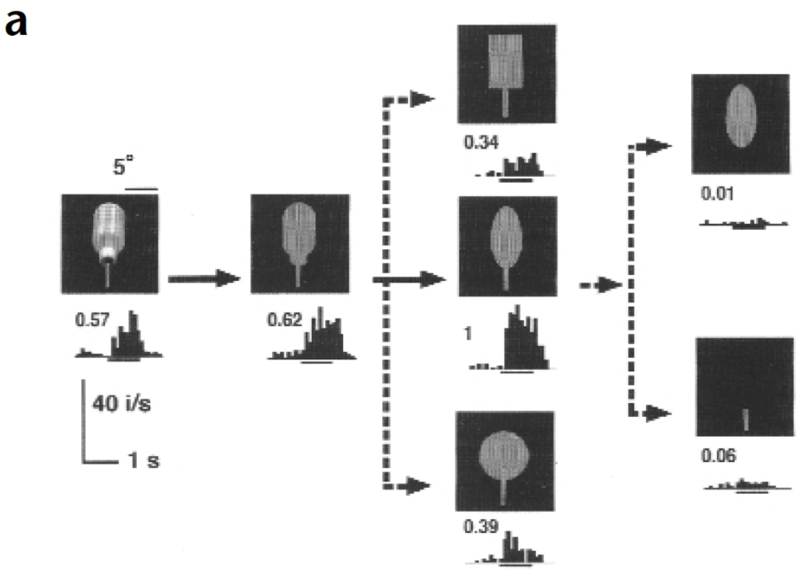
Increasing complexity along the visual ventral stream



Kravitz et al. 2012

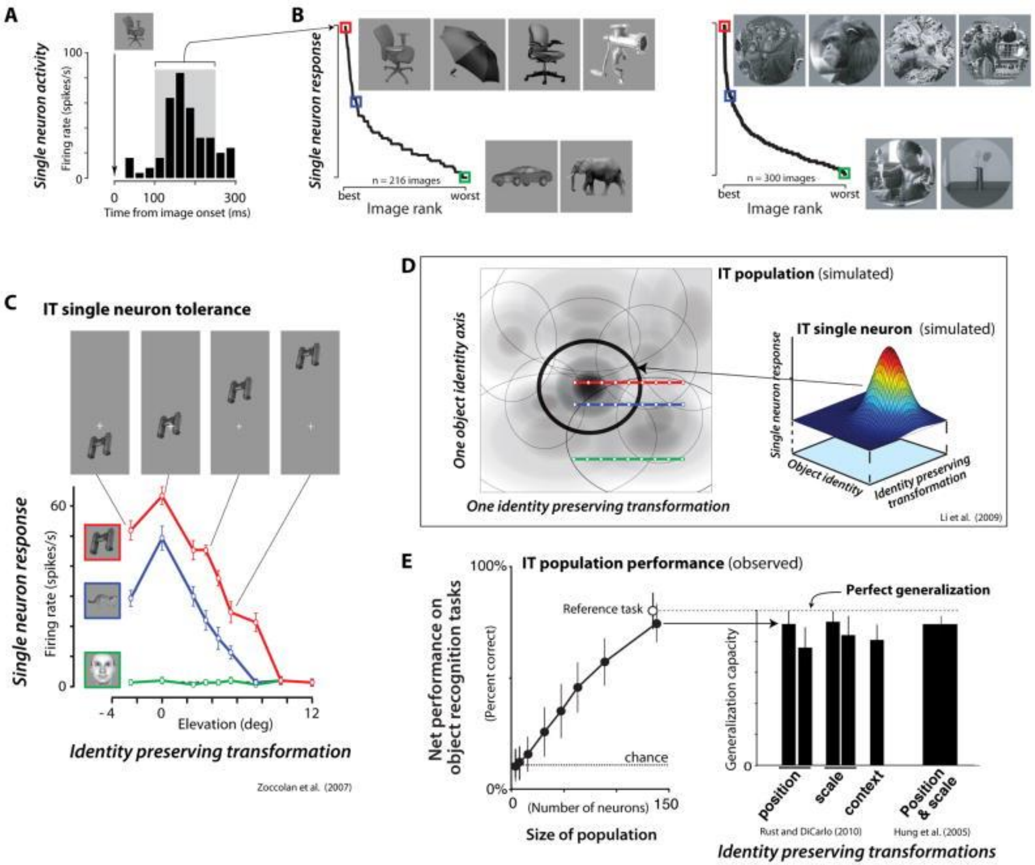
IT neurons are nonlinear

Example neuron



Wang, Tanifuji, Tanaka 1998

What does IT do?



Object-centric representations,

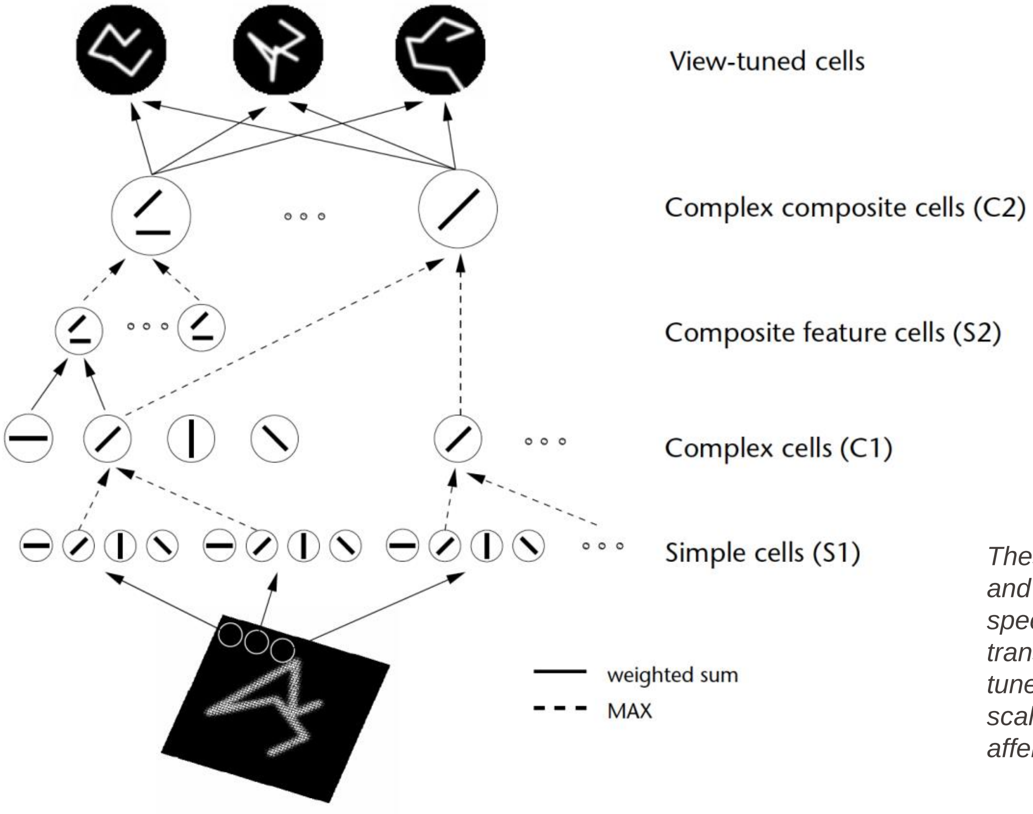
invariance to viewpoint variations

DiCarlo, Neuron 2012

How can we achieve invariance (to viewing parameters) & selectivity to identity?

Sketch of the HMAX model.

This model is an extension of Hubel &Wiesel's complex cell model and earlier work by Fukushima (Neocognitron).



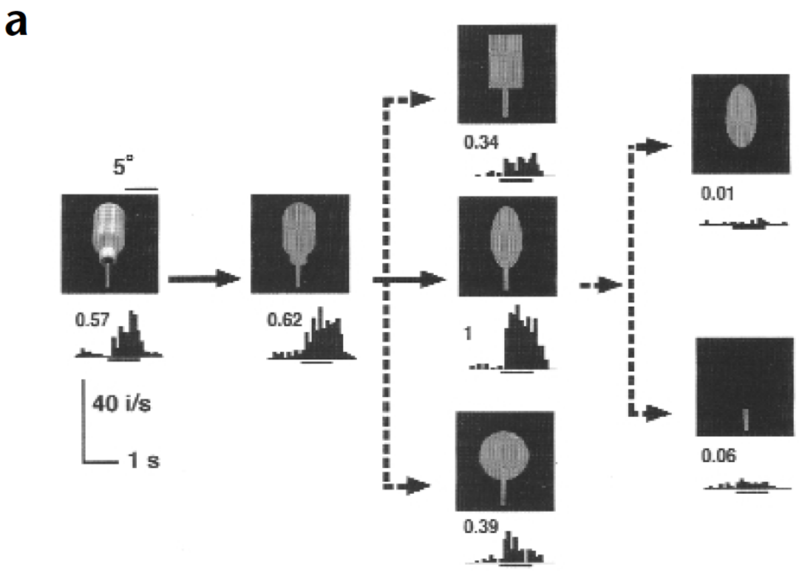
These two types of operations (max and linear sum) provided pattern specificity and invariance to translation, by pooling over afferents tuned to different positions, and to scale (not shown), by pooling over afferents tuned to different scales.

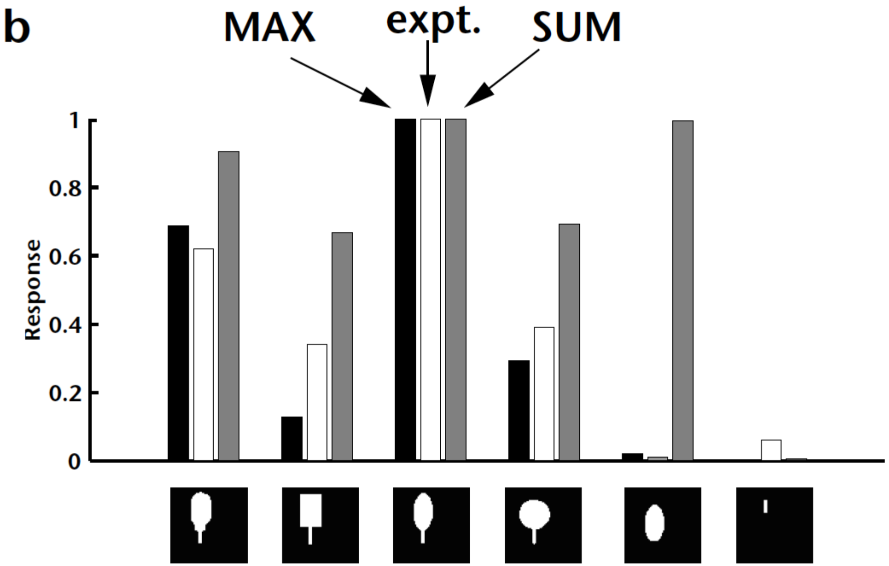
Riesenhuber & Poggio, Nat Neuro 1999

Highly nonlinear response properties

IT recordings

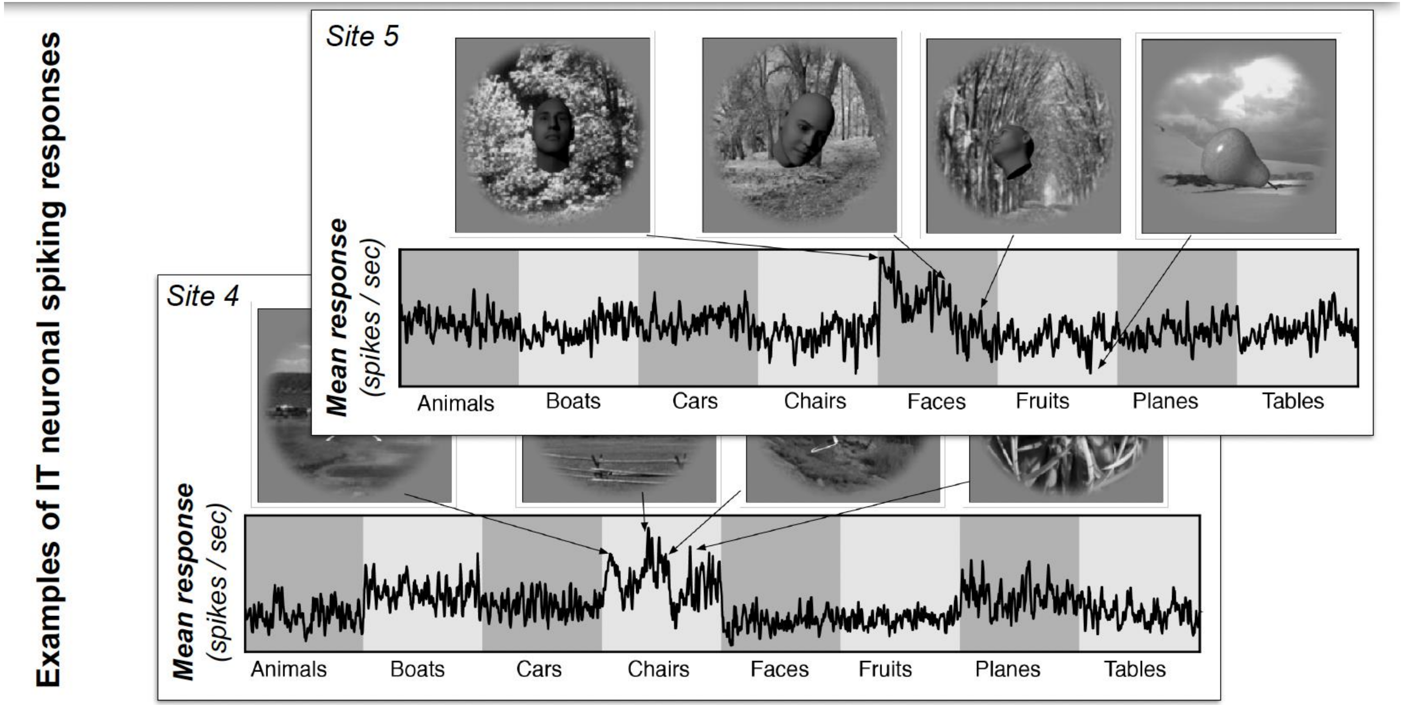
Model predictions (with max vs. sum)





Riesenhuber & Poggio, Nat Neuro 1999

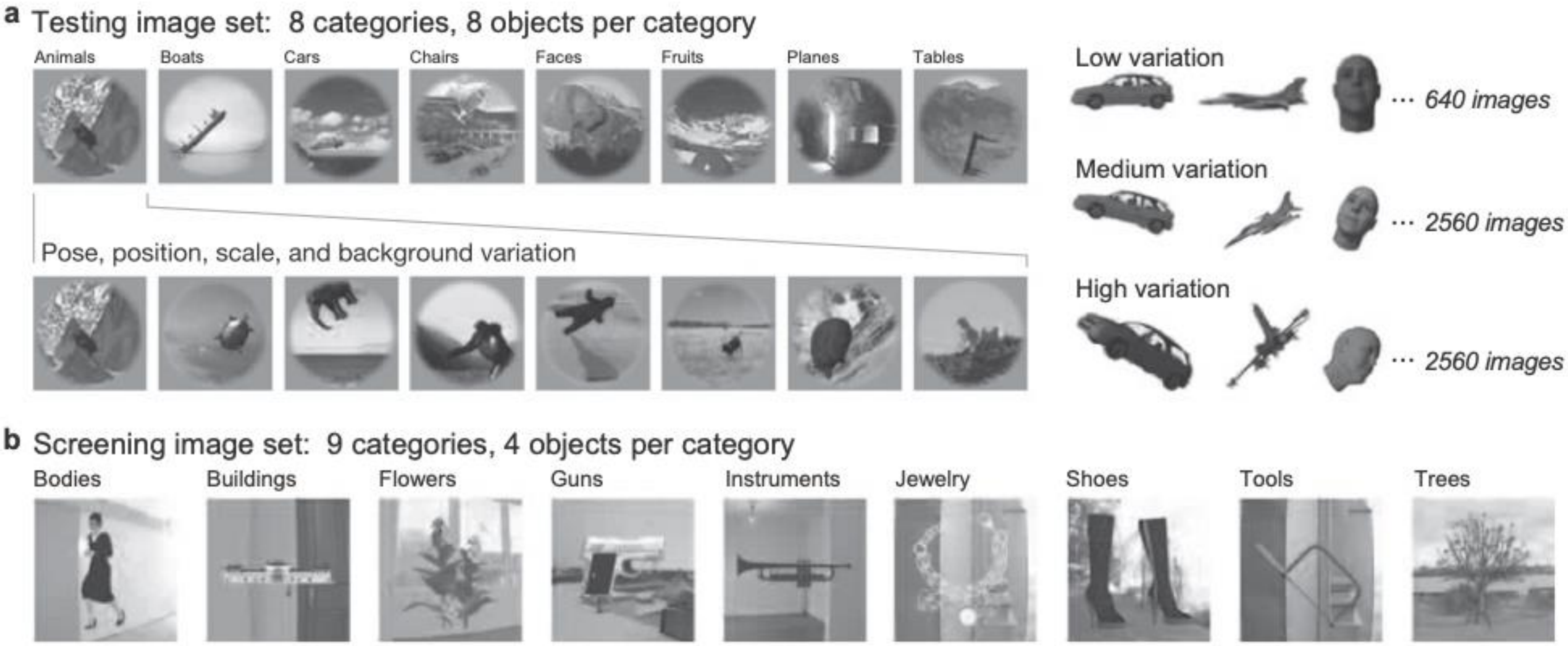
Example higher-order visual cortex responses



(1600 images tested here)

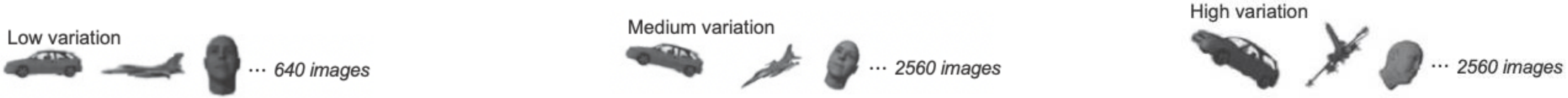
Slide from Jim DiCarlo, MIT

Core-object recognition paradigm



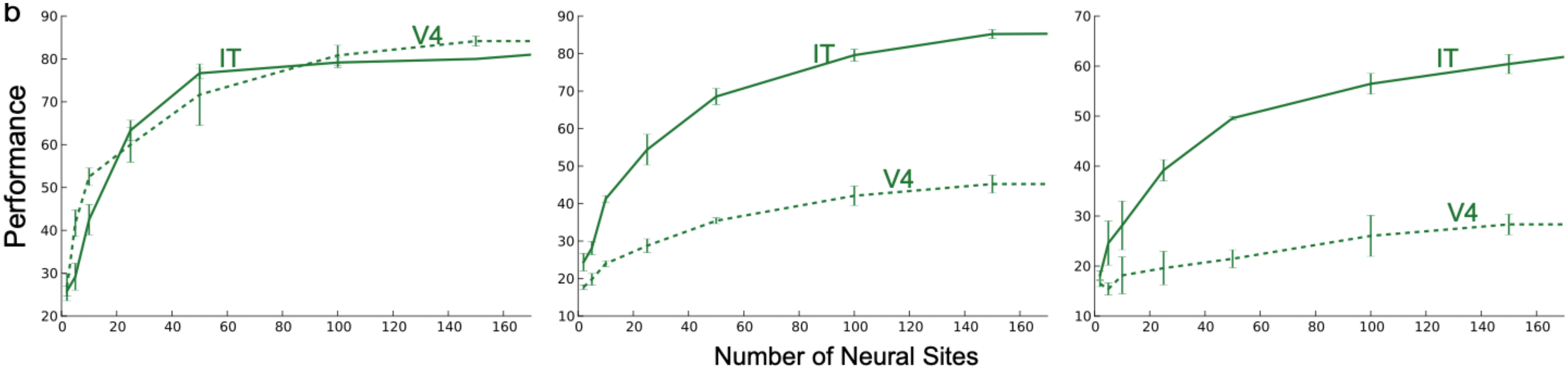
Hong & Yamins et al., NatNeuro 2016

Decoding object identity from neural data



Increasing difficulty

Increasing gap between V4 and IT



Hong & Yamins et al., NatNeuro 2016

Take-home messages part 2

- Visual pathway: increased invariance to variations in viewpoint, culminating in most complexity in inferotemporal cortex (IT)
- Object preferences in IT
- Increased performance in object decoding with more IT sites (but not V4)
- Normative models in vision: learn neural activity via behavior
- HMAX as an early model of hierarchical invariance via simple and complex cells