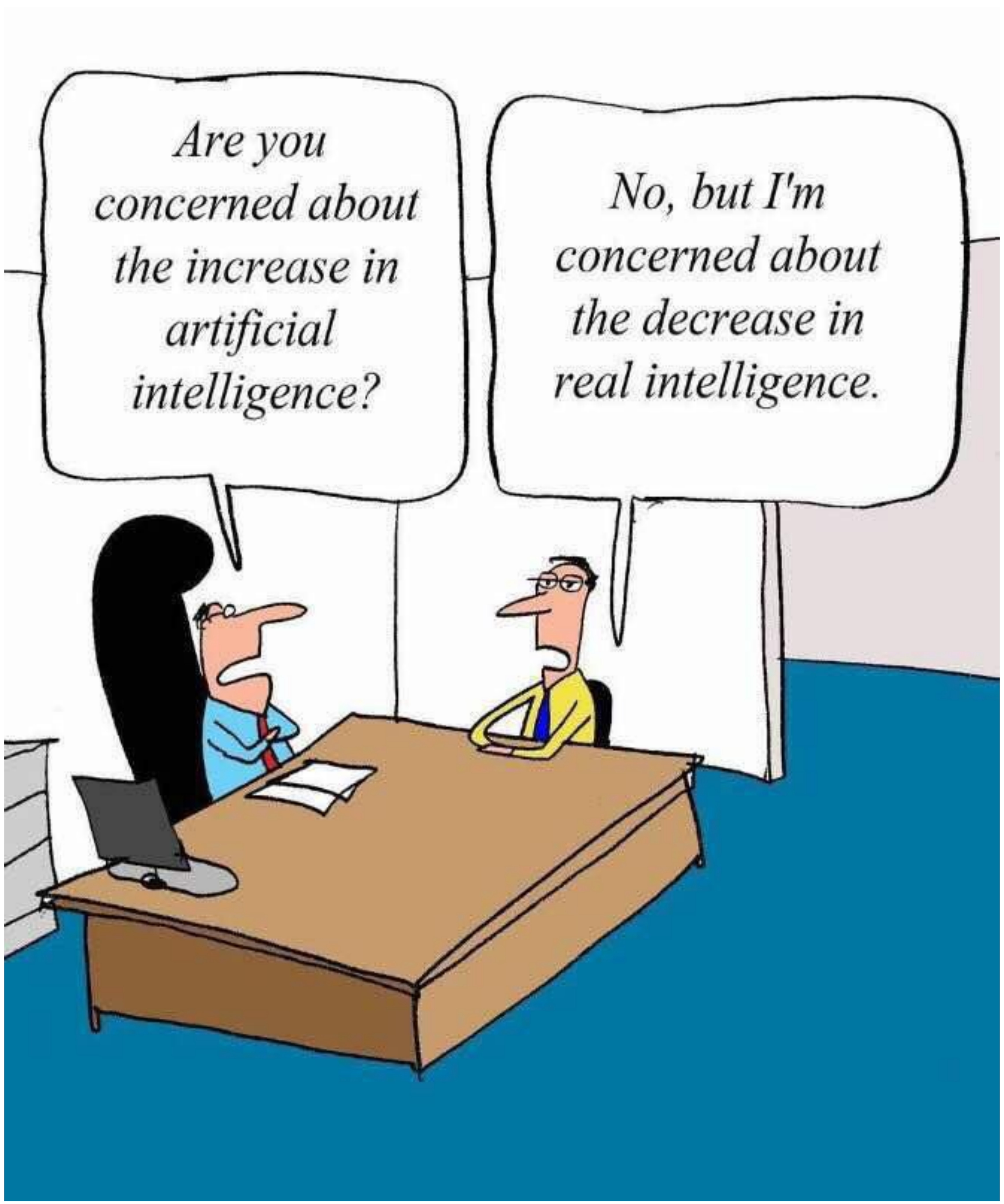

Are you concerned about the increase in artificial intelligence?
No, but I'm concerned about the decrease in real intelligence.

Deep Learning as a model of the brain

@kordinglab, CIS 522, 2020

CIS 522: Lecture 14T

Neuroscience & Deep Learning
4/28/20

Today

- Why the brain must approximate BP
- Proposals of how the brain may approximate BP

So far in this course

- Look at neuroscience
- and cognitive science
- Be (weakly) inspired to try new cost functions, architectures, and optimizers

Could the brain do something like backpropagation of error?

What is a learning algorithm?

$$\Delta W$$

What is a learning algorithm?

$$\Delta \mathcal{L} = \Delta W \frac{\partial \mathcal{L}}{\partial W}$$

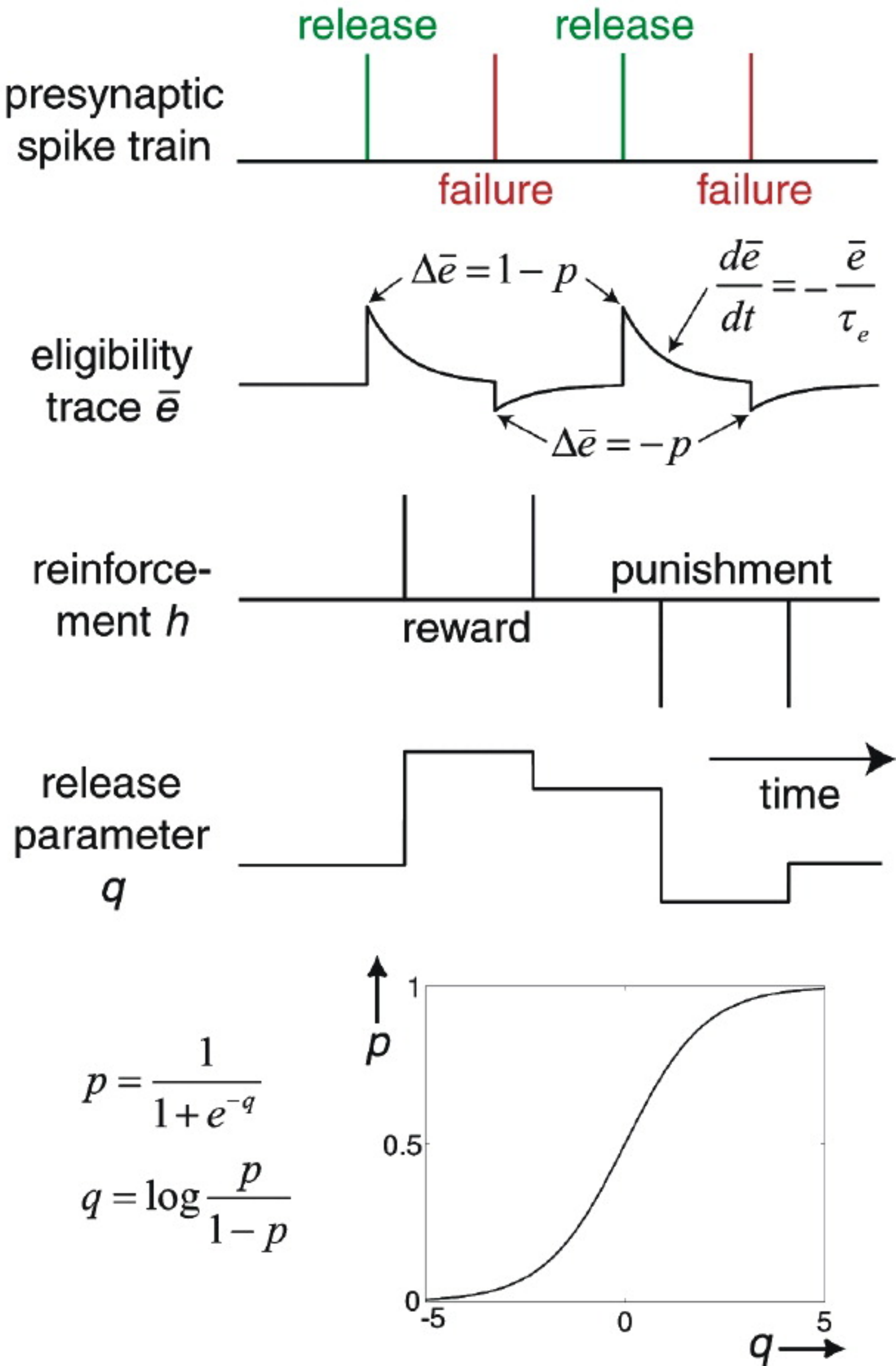
Why backpropagation works

$$\Delta \mathcal{L} = \Delta W \frac{\partial \mathcal{L}}{\partial W} \qquad \Delta W = -\eta \frac{\partial L}{\partial W}$$

$$\Rightarrow \qquad \Delta L = -\eta \left| \frac{\partial L}{\partial W} \right|^2$$

Simple solutions can not work

Weight perturbation



Sebastian Seung, Ila Fiete and others

Variance of the estimator?

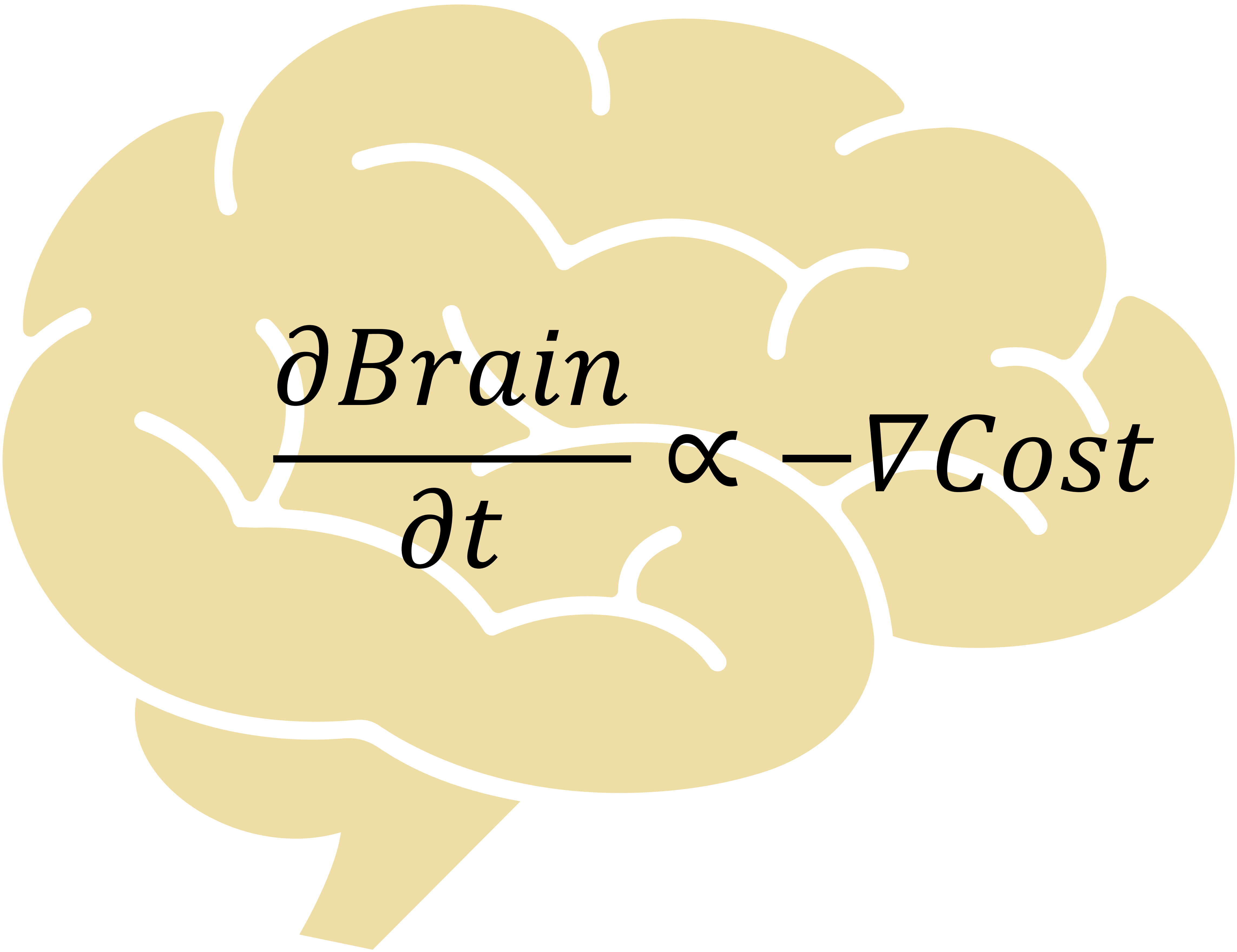
$$\Delta W \sim N(0, \epsilon)$$

$$f(W + \Delta W) = f(W) + \sum_i \frac{\partial f}{\partial W_i} \Delta W_i$$

$$var f(W + \Delta W) = N\epsilon$$

$$var \frac{\Delta f}{\Delta W_i} \approx N$$

An increasingly popular hypothesis



$$\frac{\partial Brain}{\partial t} \propto -\nabla Cost$$

Why now?

- Hydraulic Brains
- Electric Brains
- Computer Brains
- **Deep learning Brains**

Cynical answer:
Whatever works is
our best hypothesis

Could the brain implement backdrop?

- Problem 1: Weight transport
- Problem 2: Multiplexing

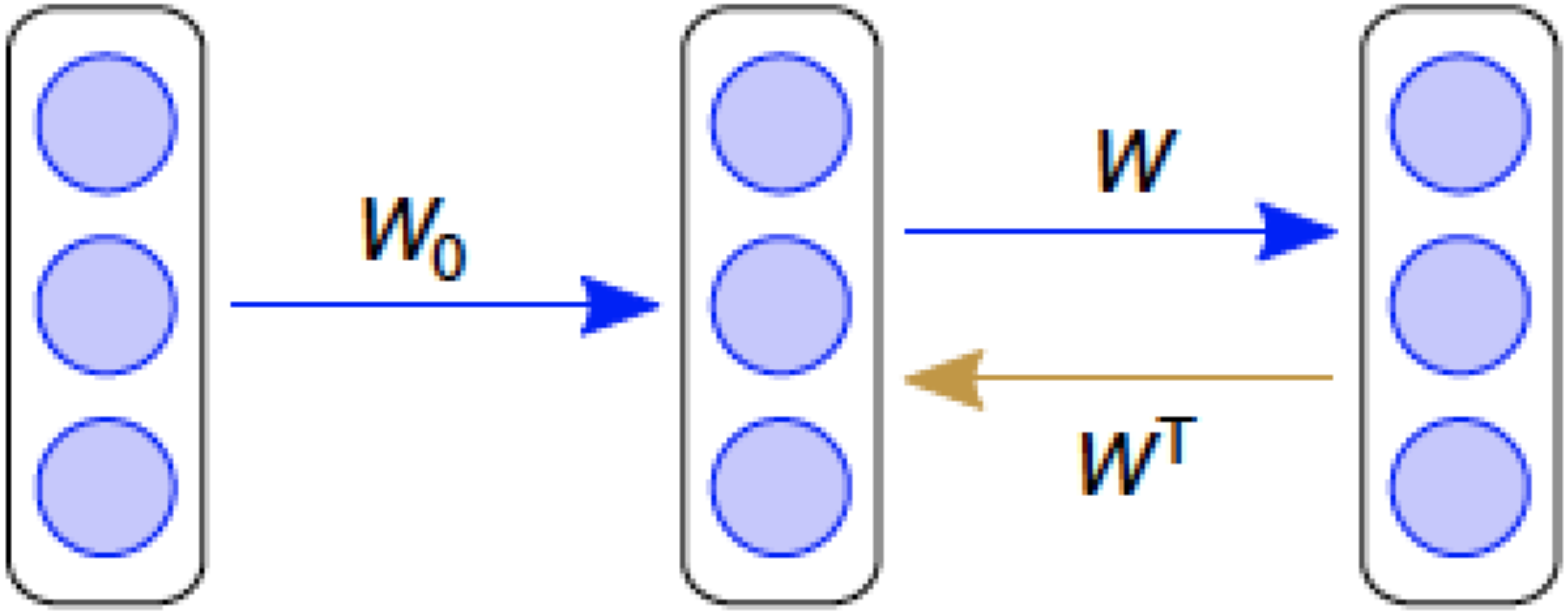
Realistic BP is about as old as BP

COGNITIVE SCIENCE 9, 147–169 (1985)

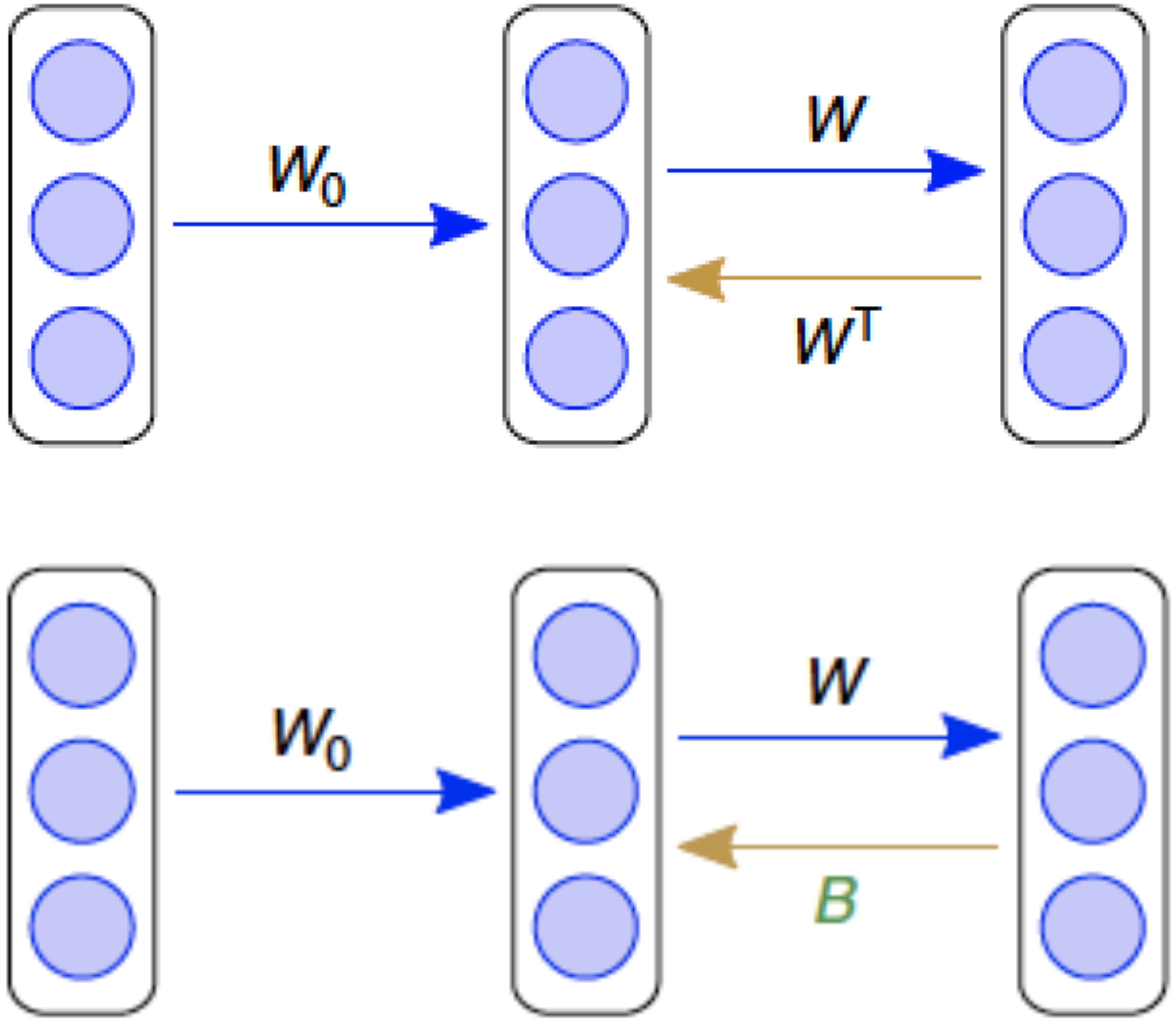
A Learning Algorithm for Boltzmann Machines*

DAVID H. ACKLEY
GEOFFREY E. HINTON
Computer Science Department
Carnegie-Mellon University
TERRENCE J. SEJNOWSKI
Biophysics Department
The Johns Hopkins University

Problem 1: weight transport



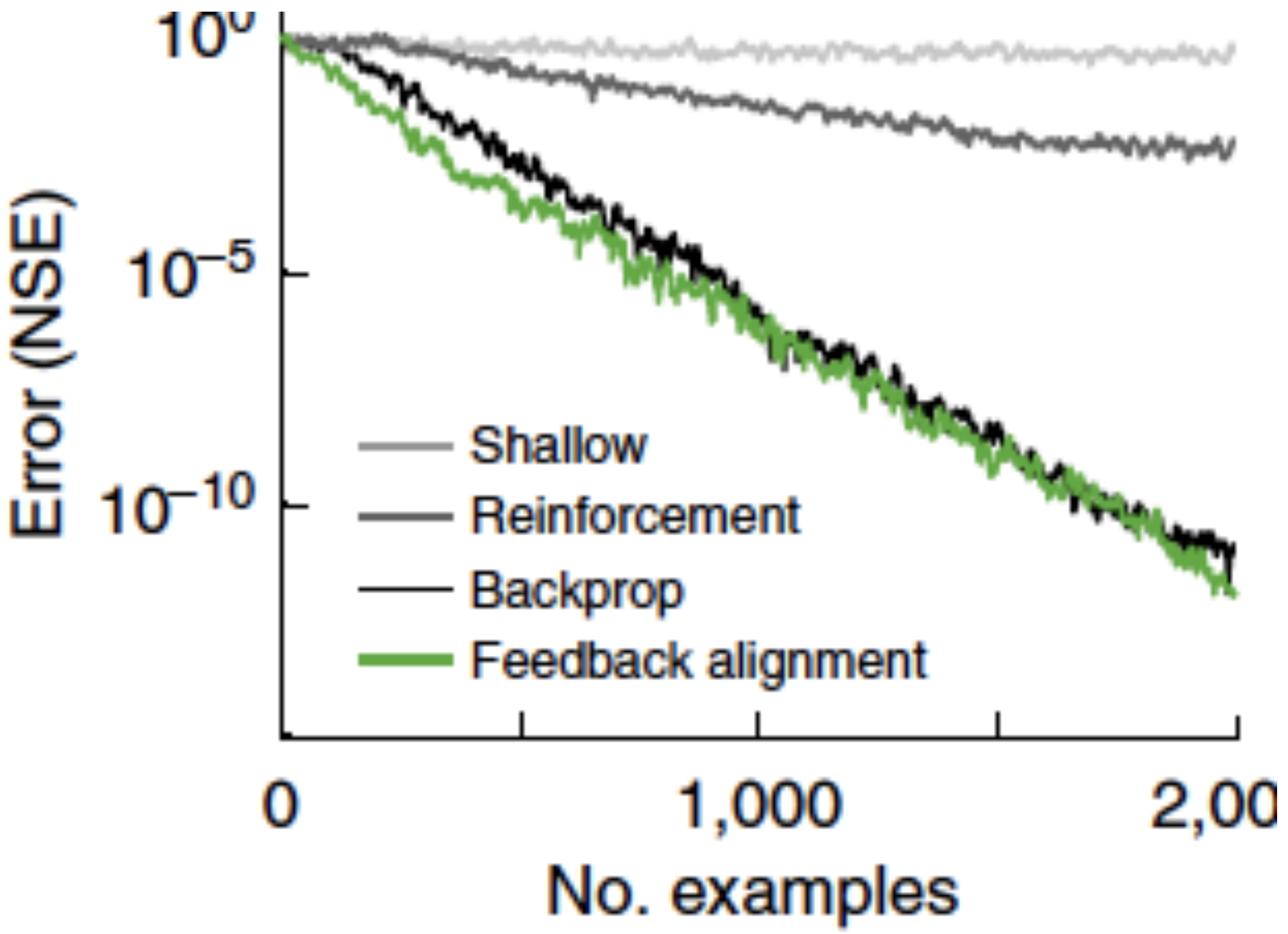
Feedback alignment

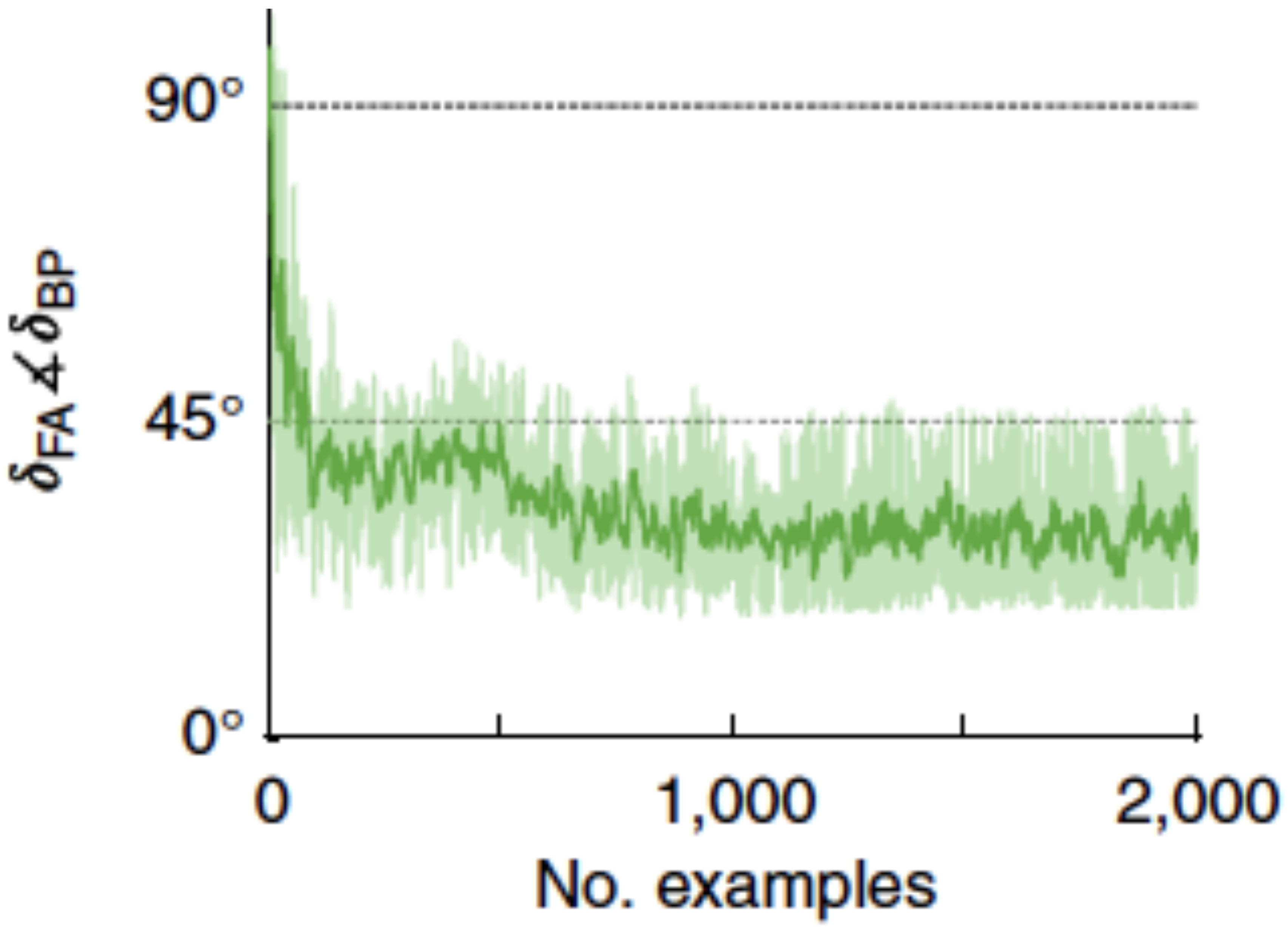


$$\Delta W_0 \propto B\mathbf{e}\mathbf{x}^T$$

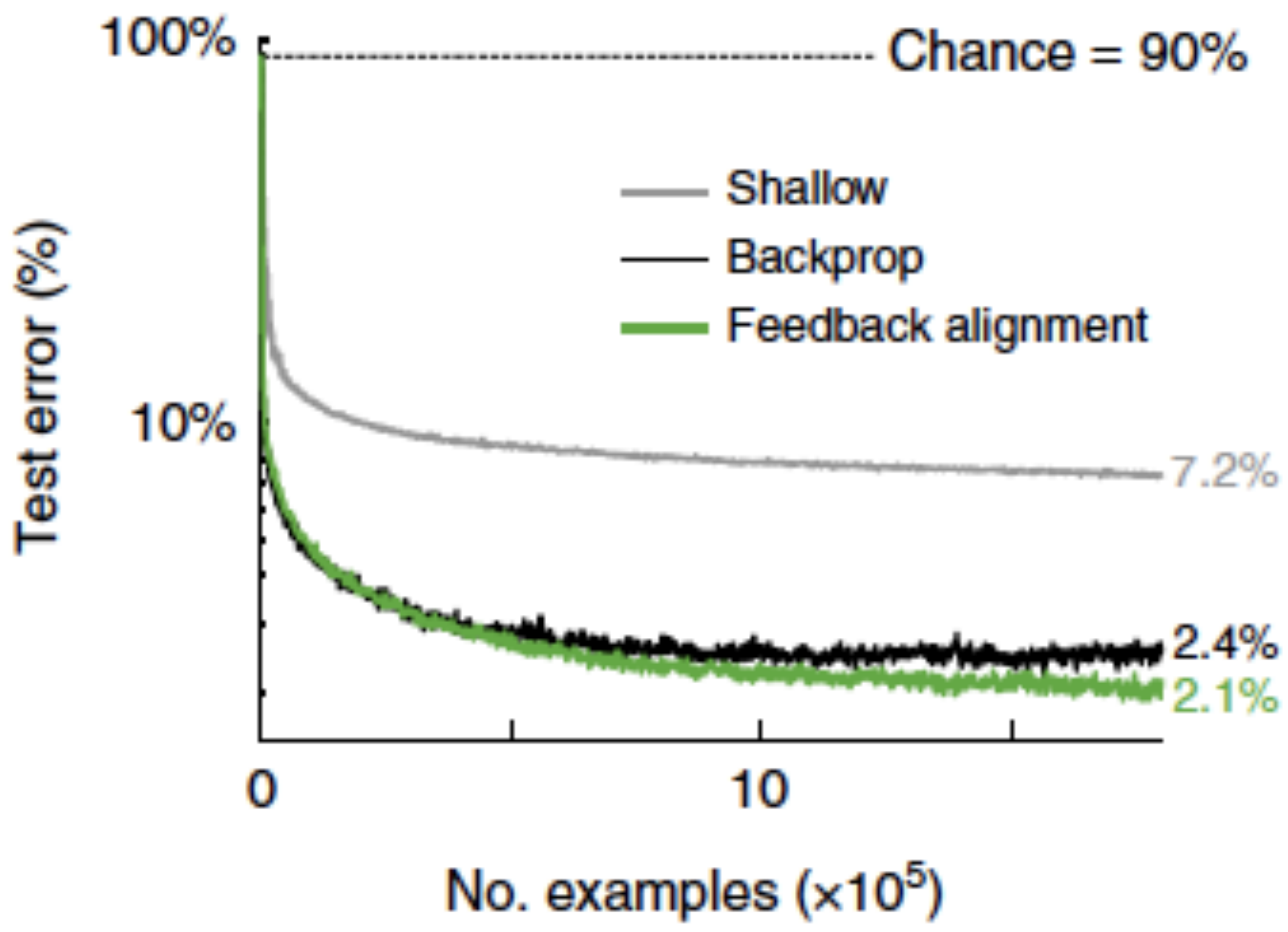
Lillicrap Tweed ...

Works on linear problems



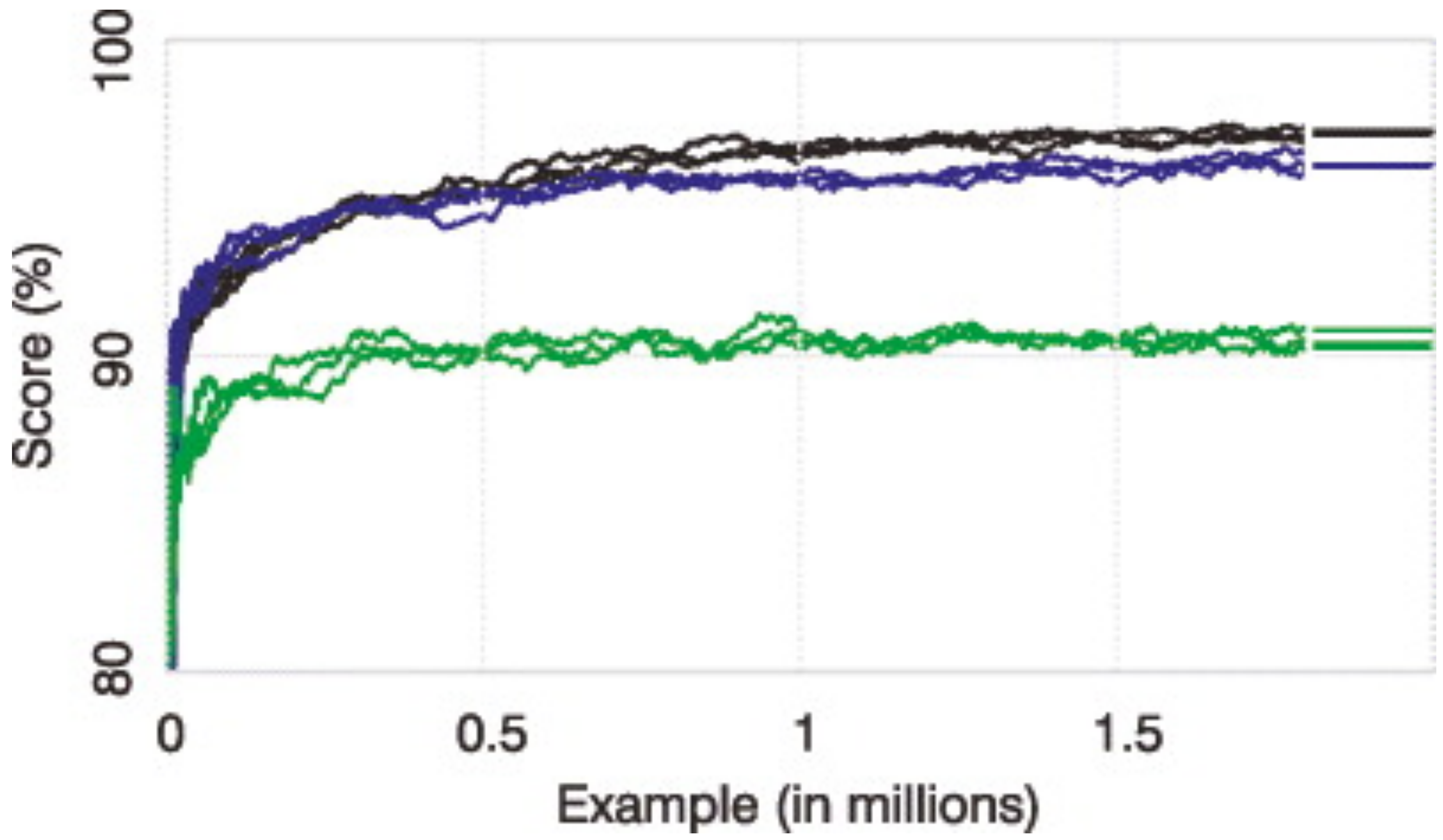


MNIST



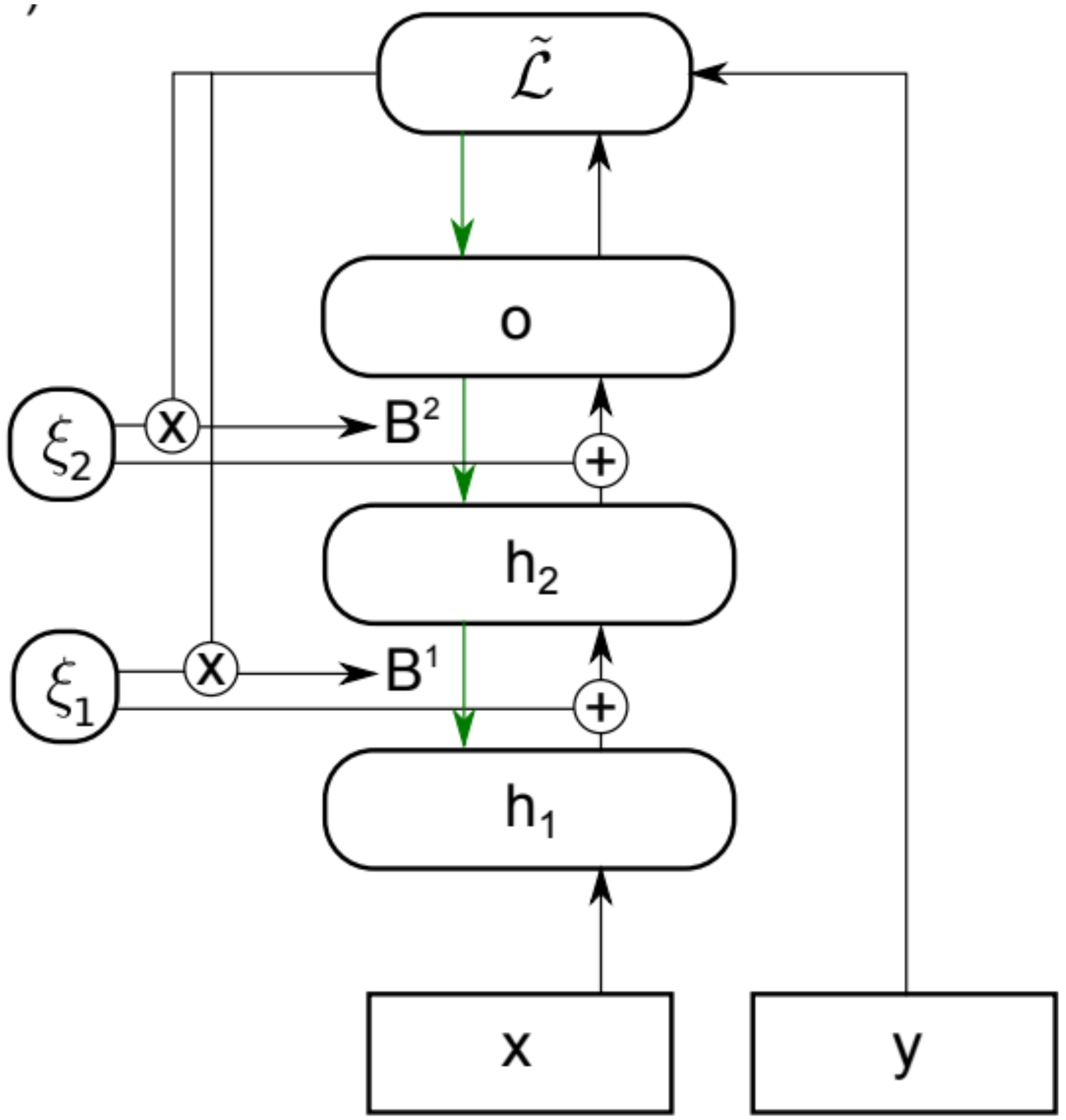
100%
Chance = 90%
Shallow
Backprop
Feedback alignment
10%
Test error (%)
7.2%
2.4%
2.1%
0
10
No. examples (×10^5)

Spiking MNIST



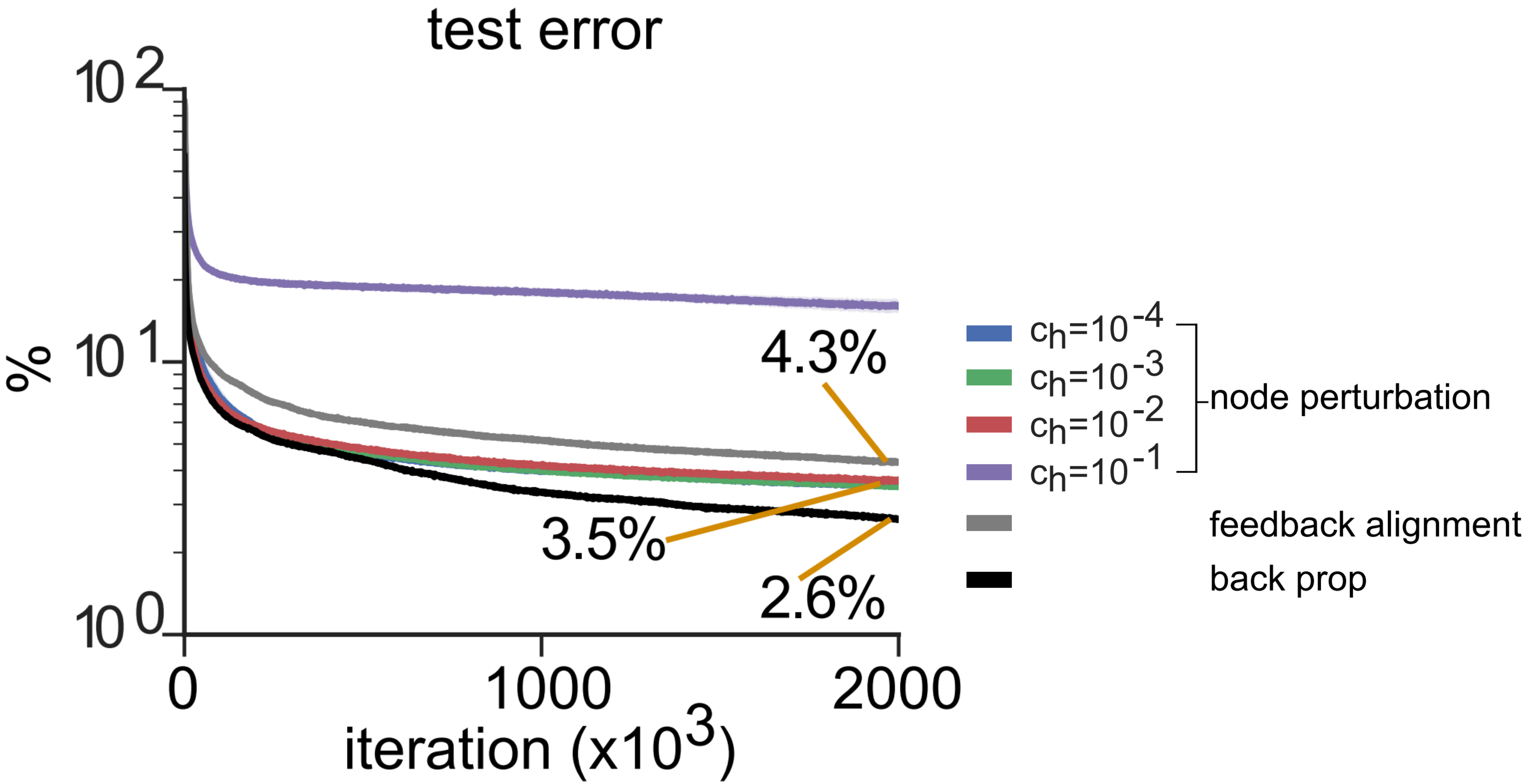
Samadi, Lillicrap, Tweed

But then, why not improve the weights?

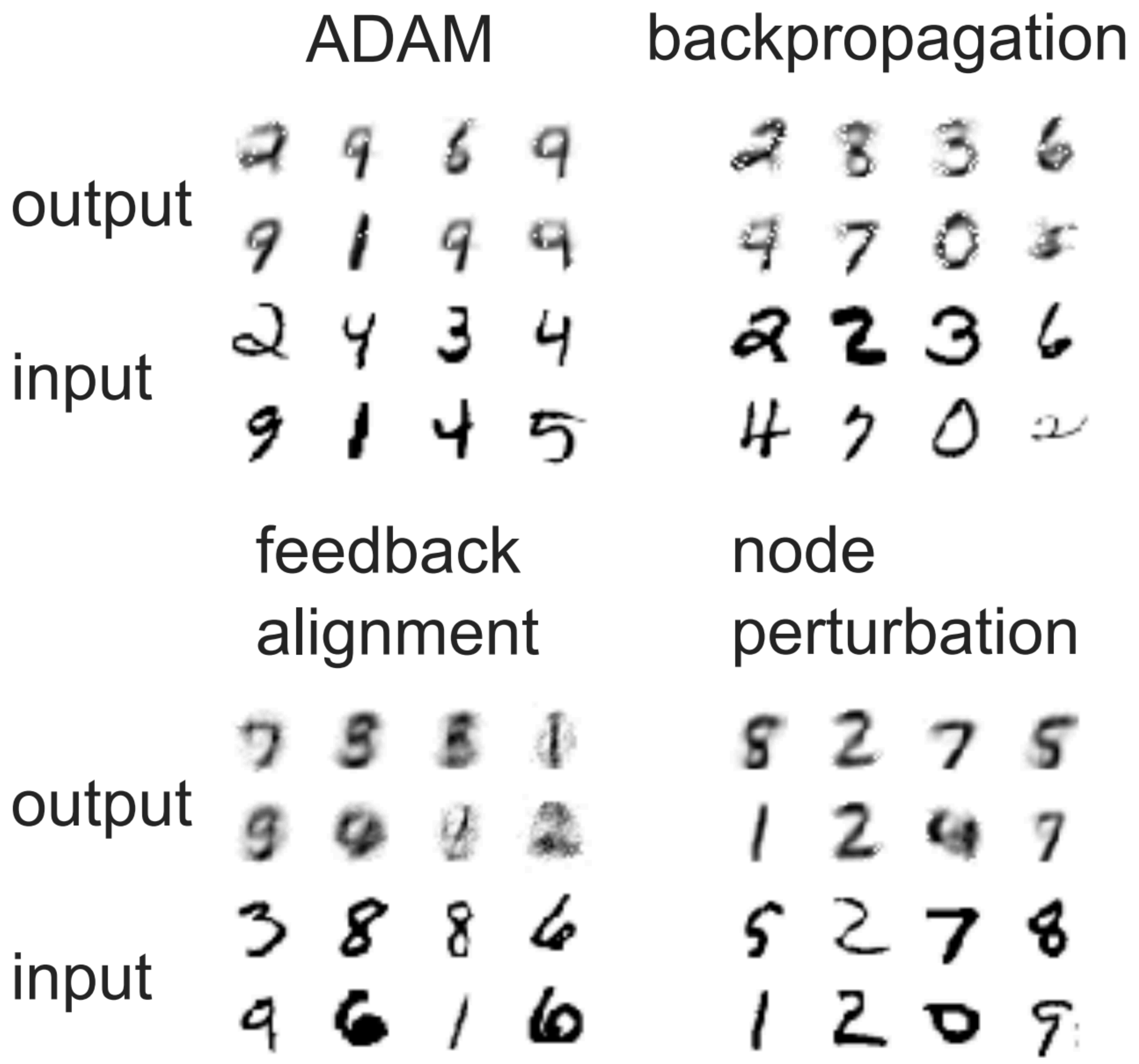


With Lansdell

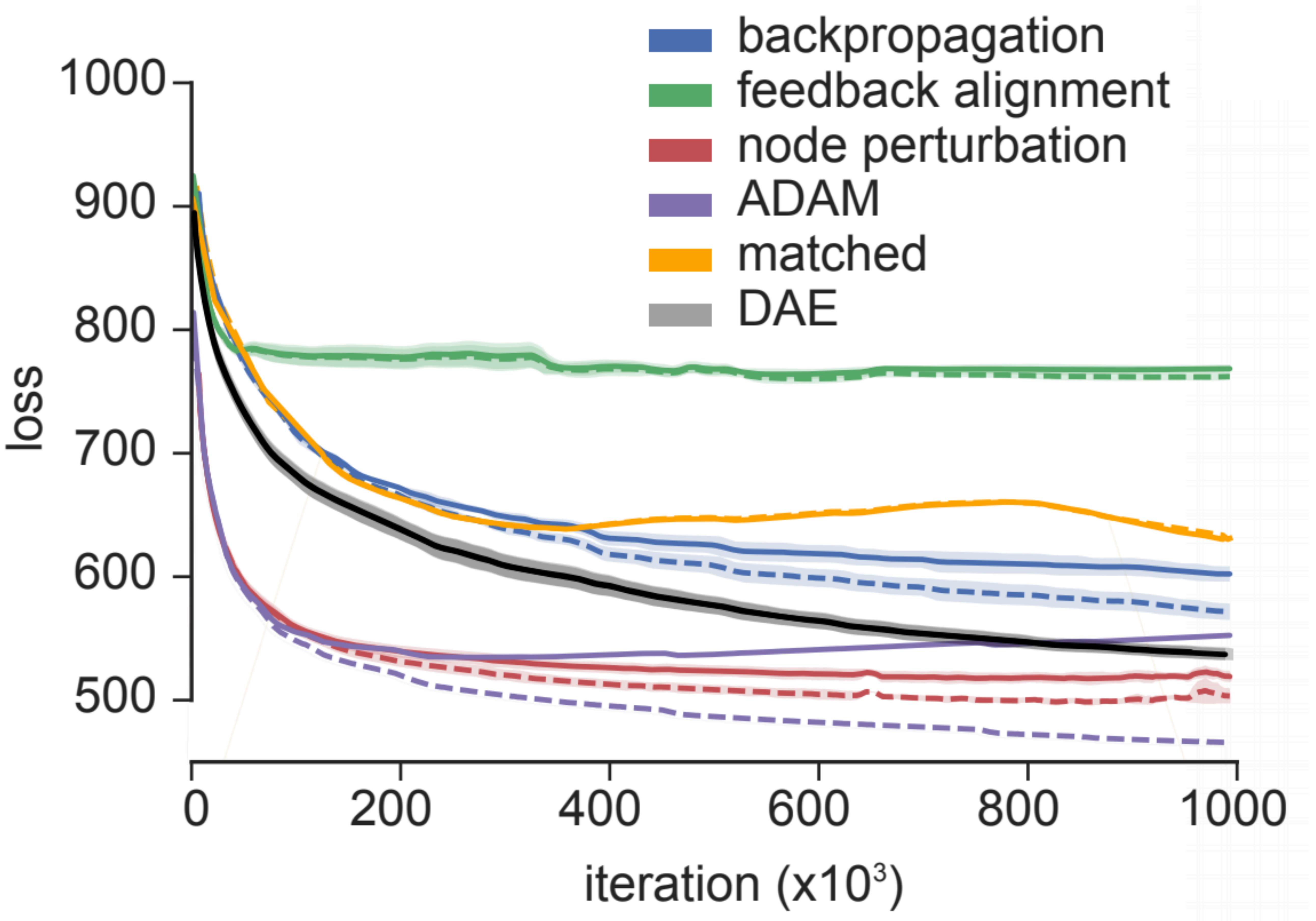
Helps on MNIST



MNIST autoencoder



Helps a lot


backpropagation
feedback alignment
node perturbation
ADAM
matched
DAE
1000
900
800
700
600
500
loss
0
200
400
600
800
1000
iteration (x10³)

Also helps for nontrivial problems

dataset	backpropagation	node perturbation	DFA
CIFAR10	77	75	72
CIFAR100	51	48	47

Proofs

- See paper for proofs
- for some cases we will converge to correct weights

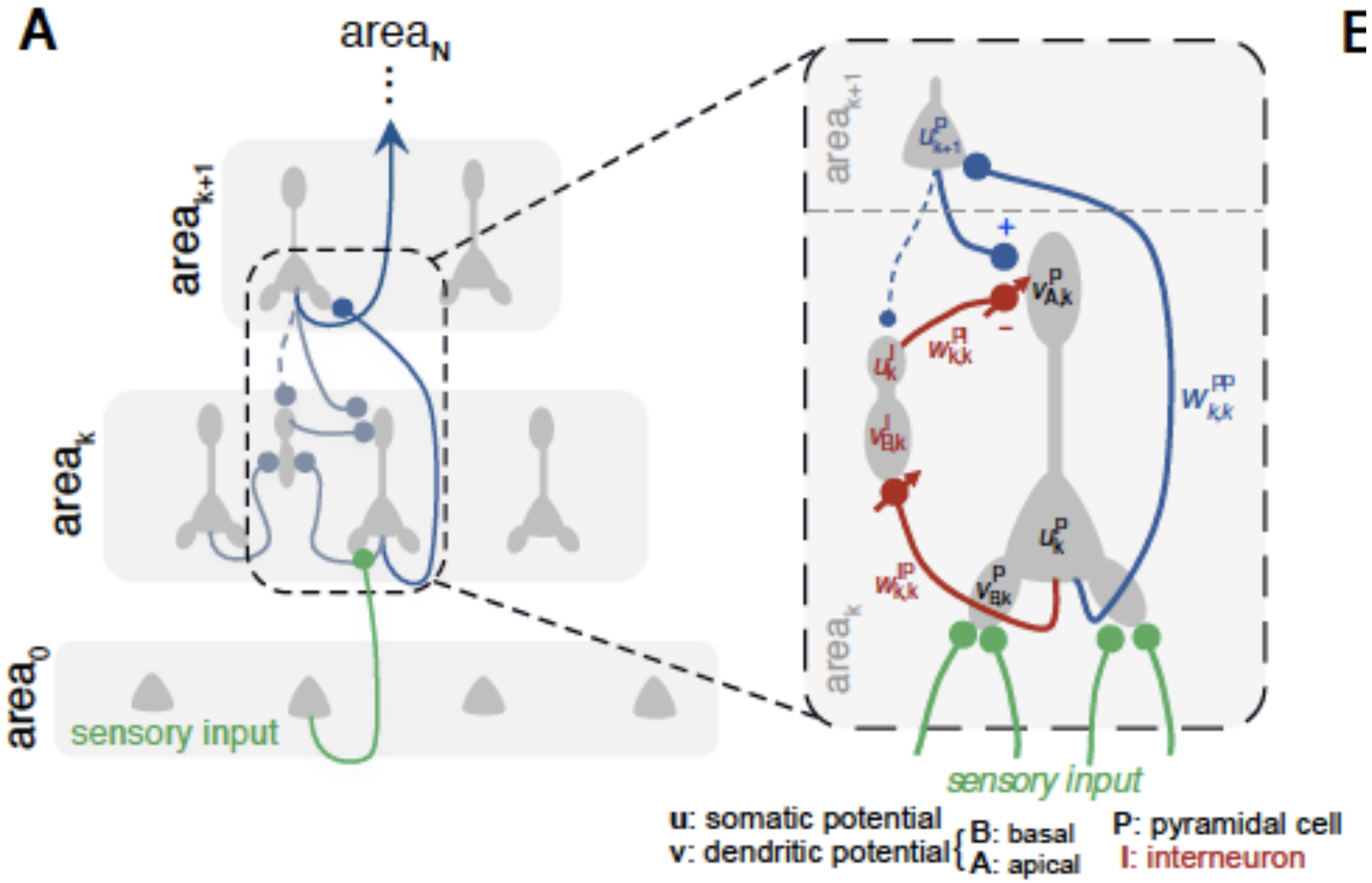
Problem 2: Multiplexing

- Spikes are supposed to transmit the signals
- But then, how could they *also* transmit the gradients
- “No-one has observed gradients being transmitted”

Ideas

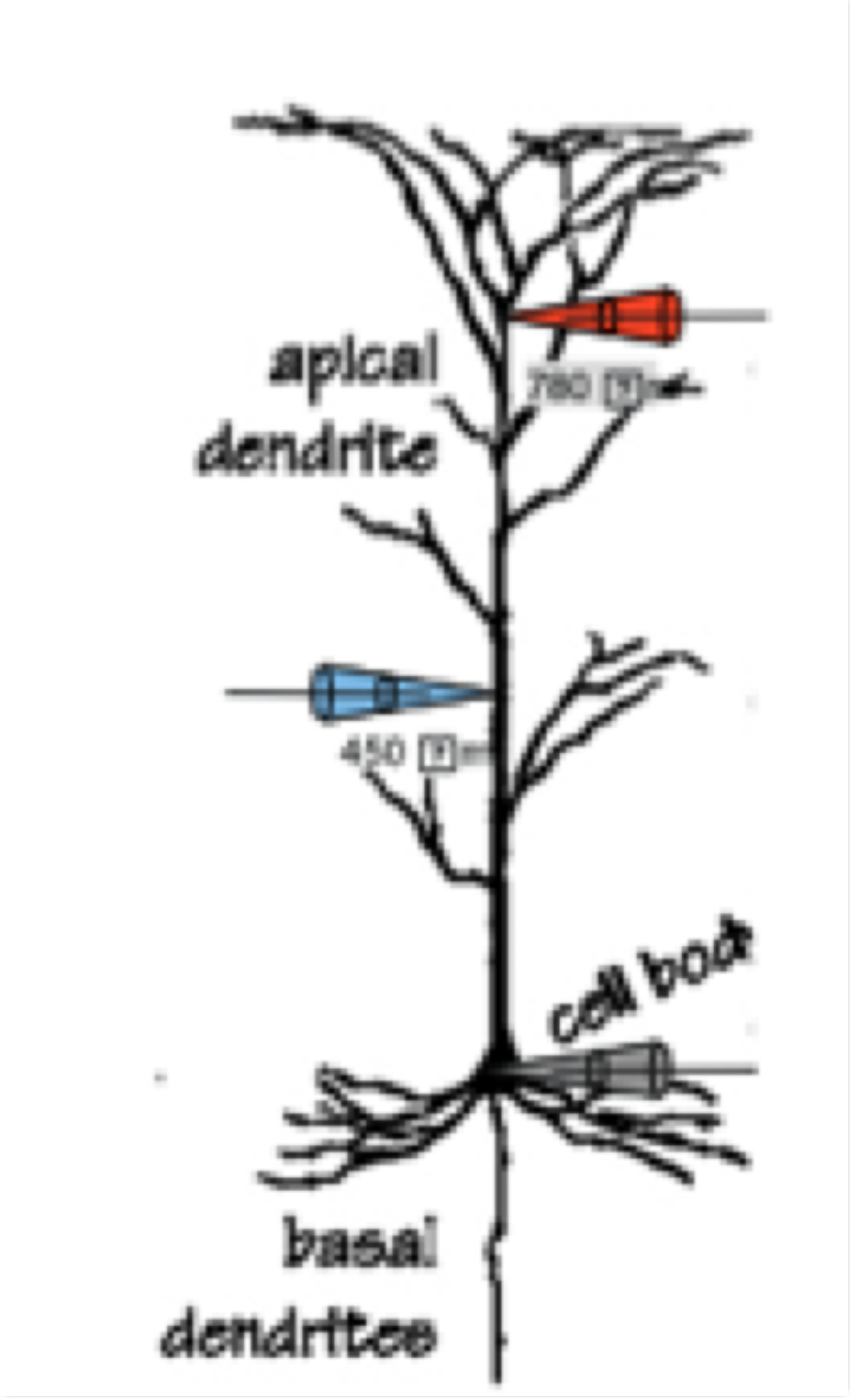
- Spikes vs Bursts
- Early vs Late
- Excitatory vs Inhibitory
- Metalearning

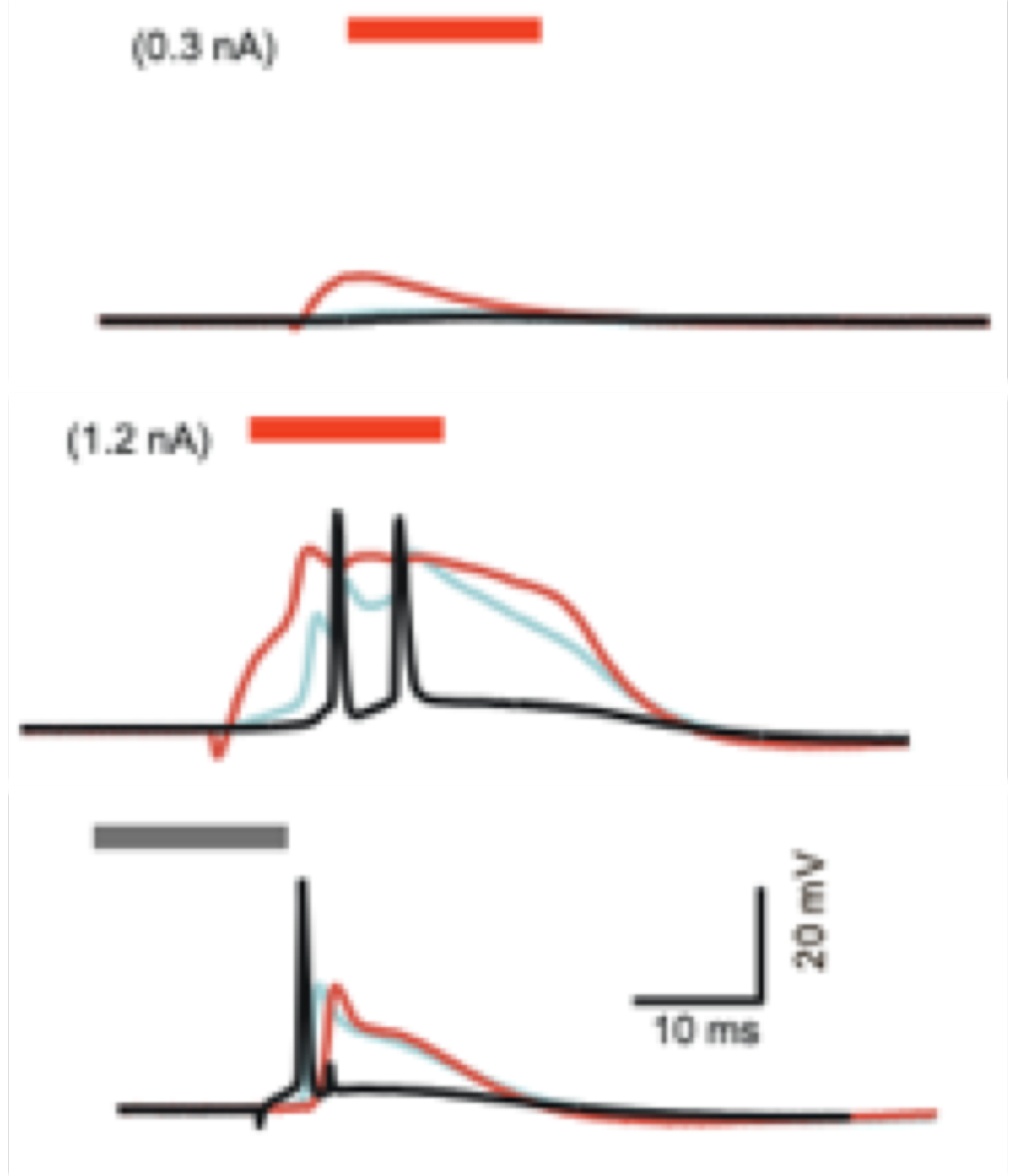
e.g. Inhibition

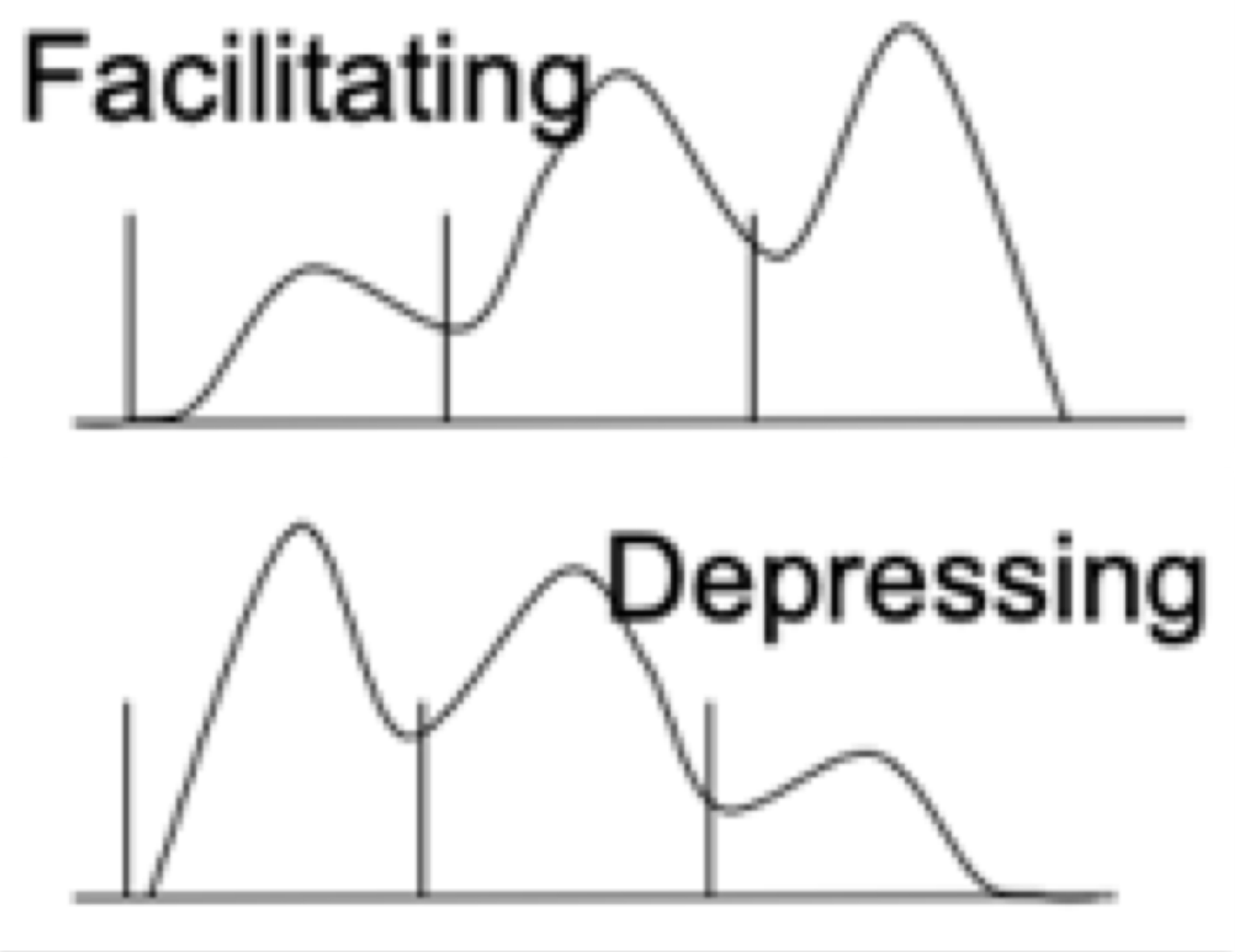


Senn and Bengio

e.g. Spikes vs Bursts



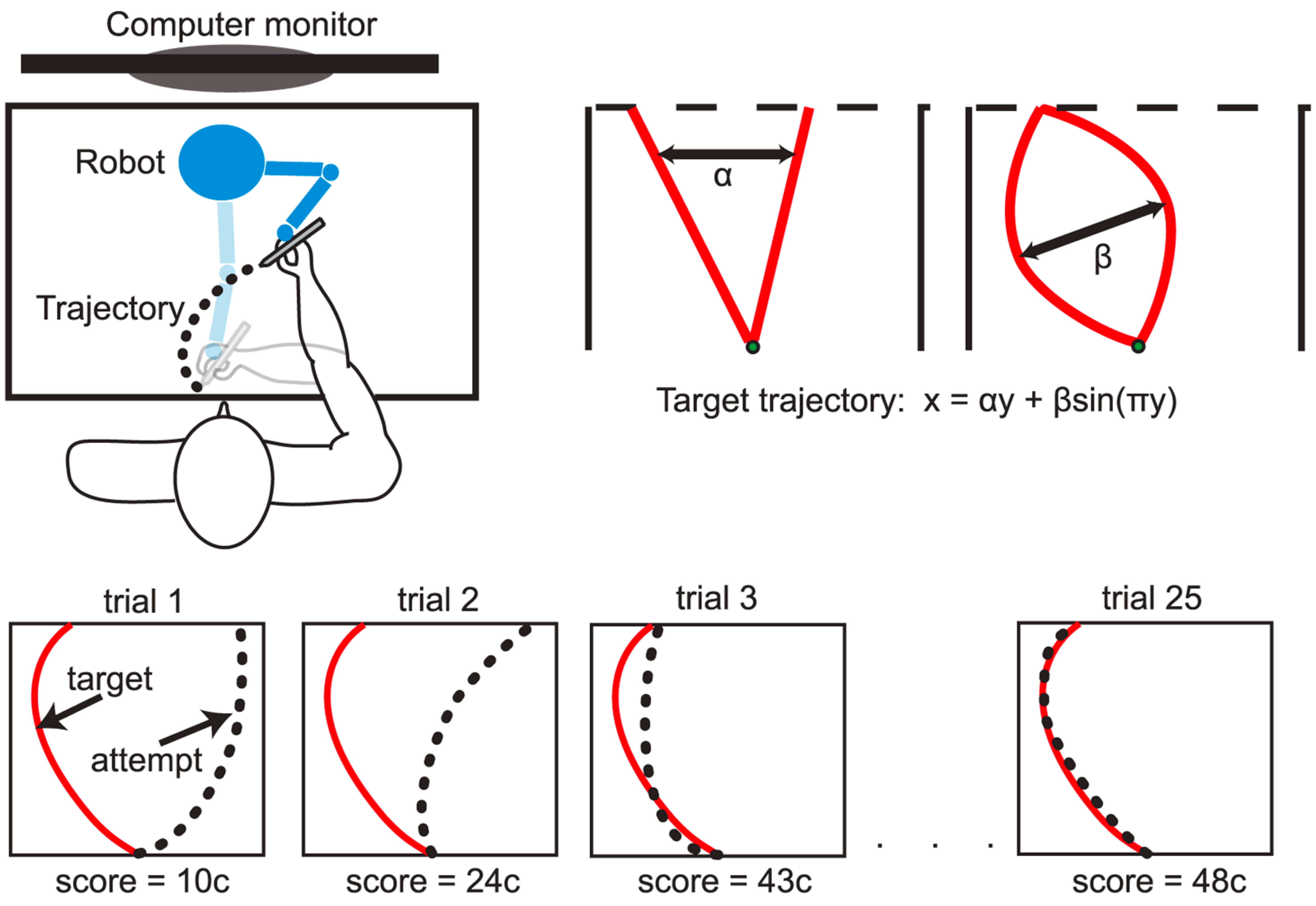




Larkum, My PhD thesis

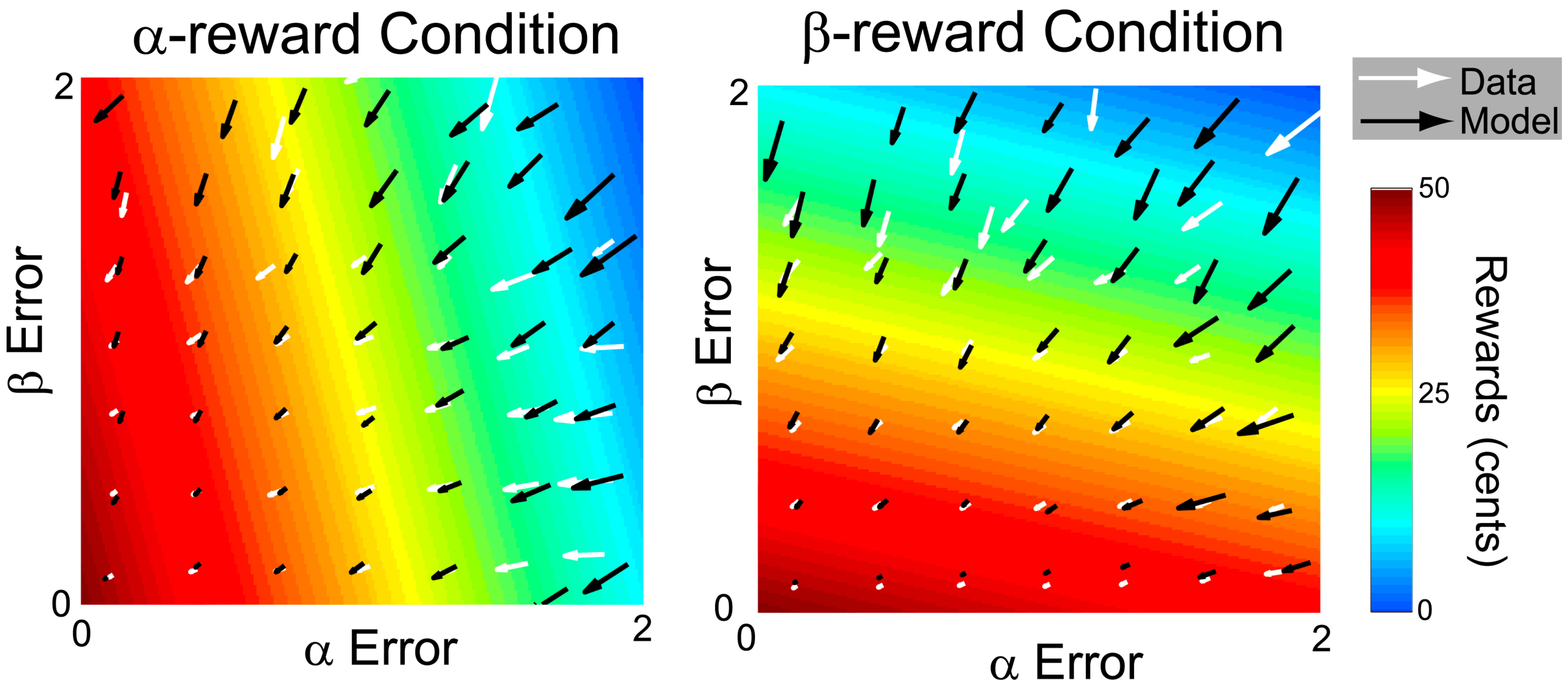
e.g. Learning to learn

People do it

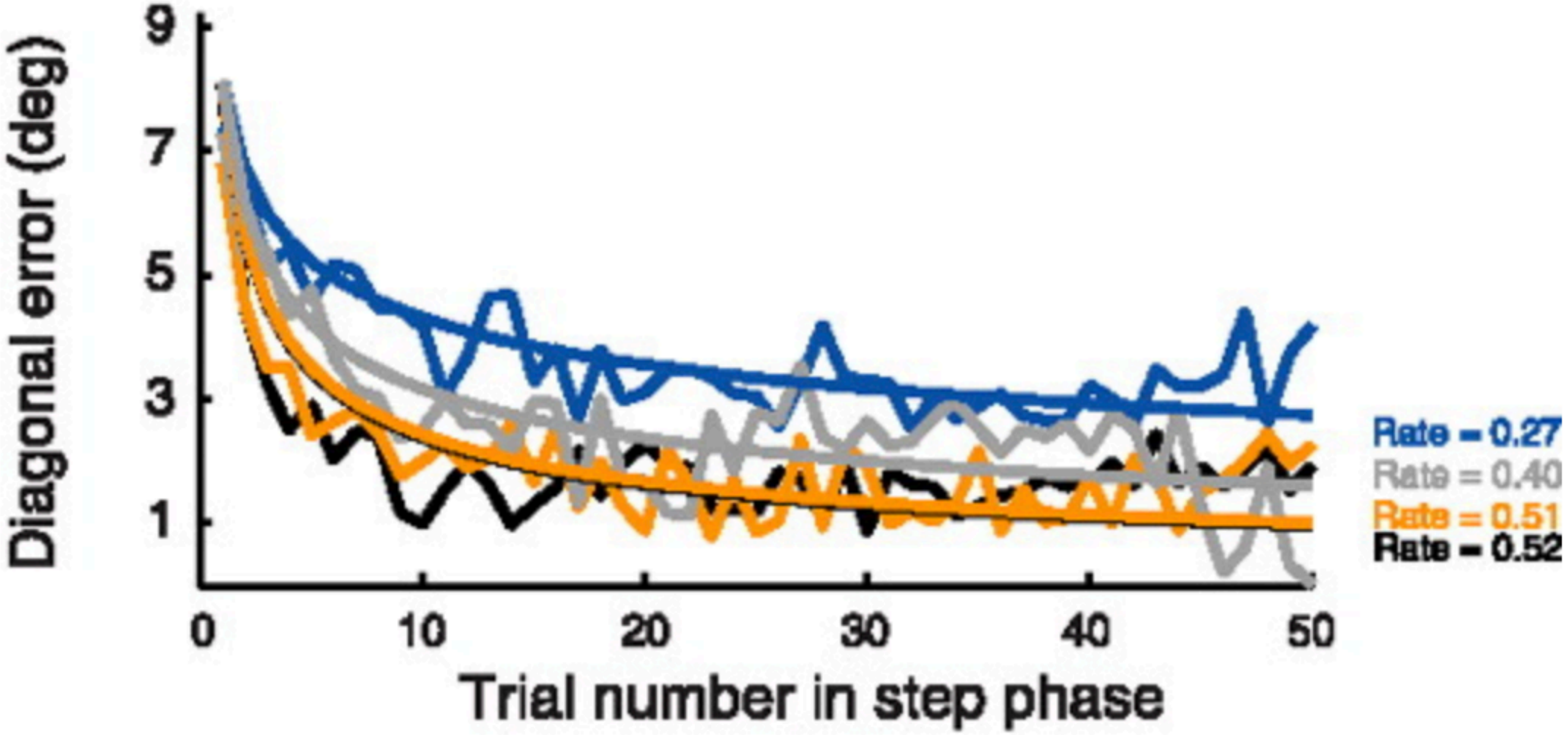


Dam, Kording, Wei

Fast learning on axis that matters

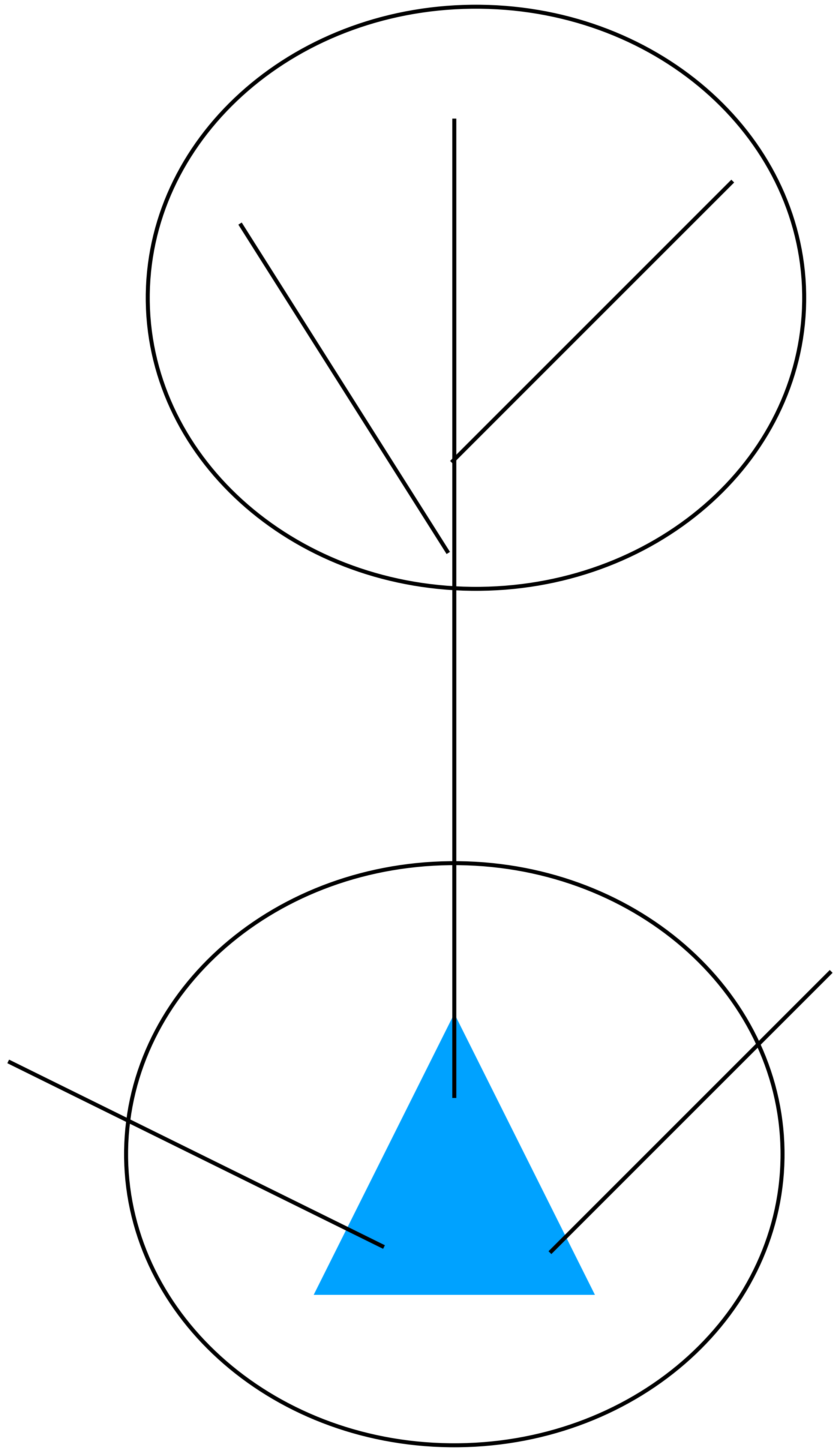


People do it: Learning to optimize



Burge, Ernst, Banks 2008

Unspecific signals



Teacher
(needs access
to truth)

Student

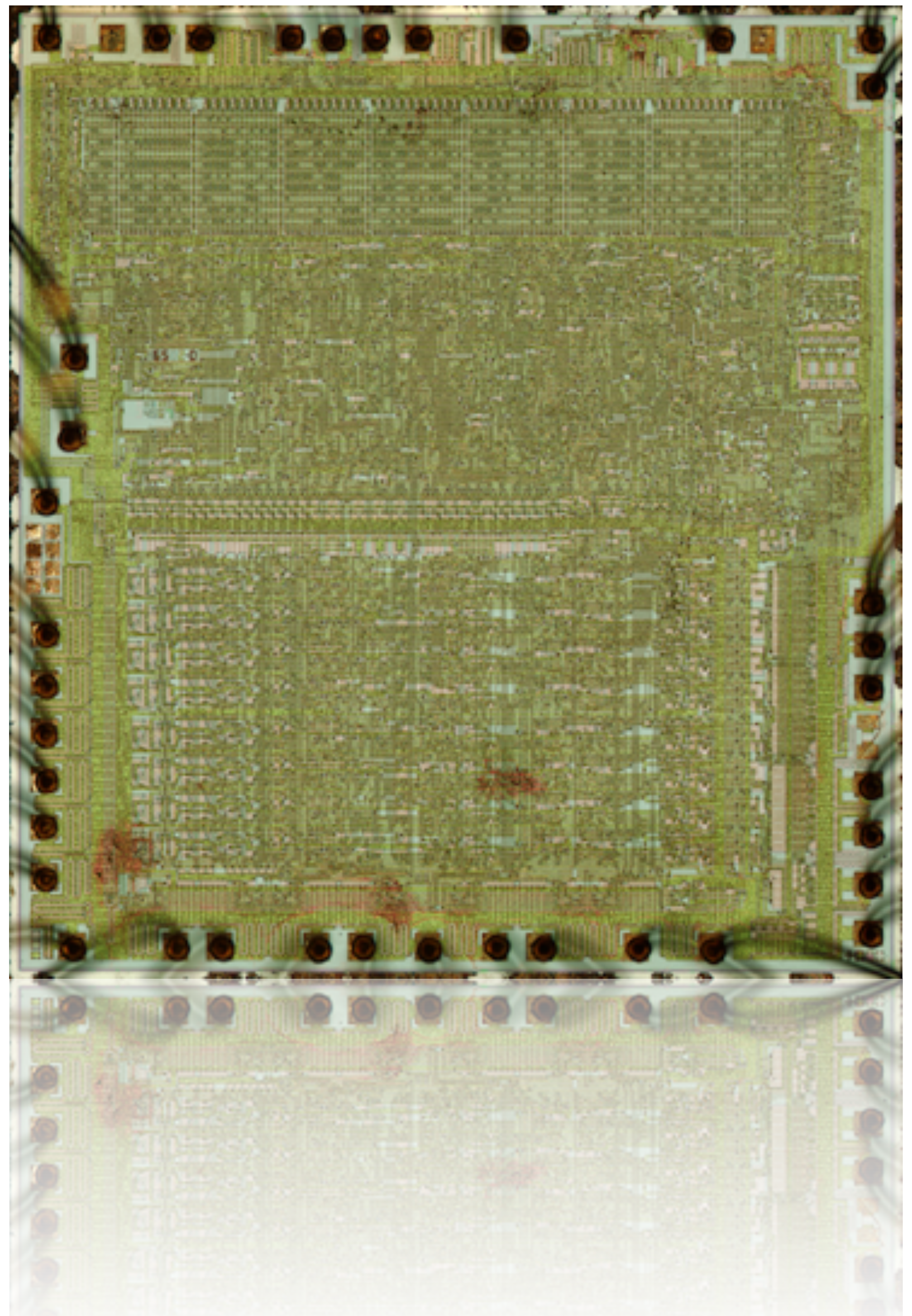
Compare synthetic gradients

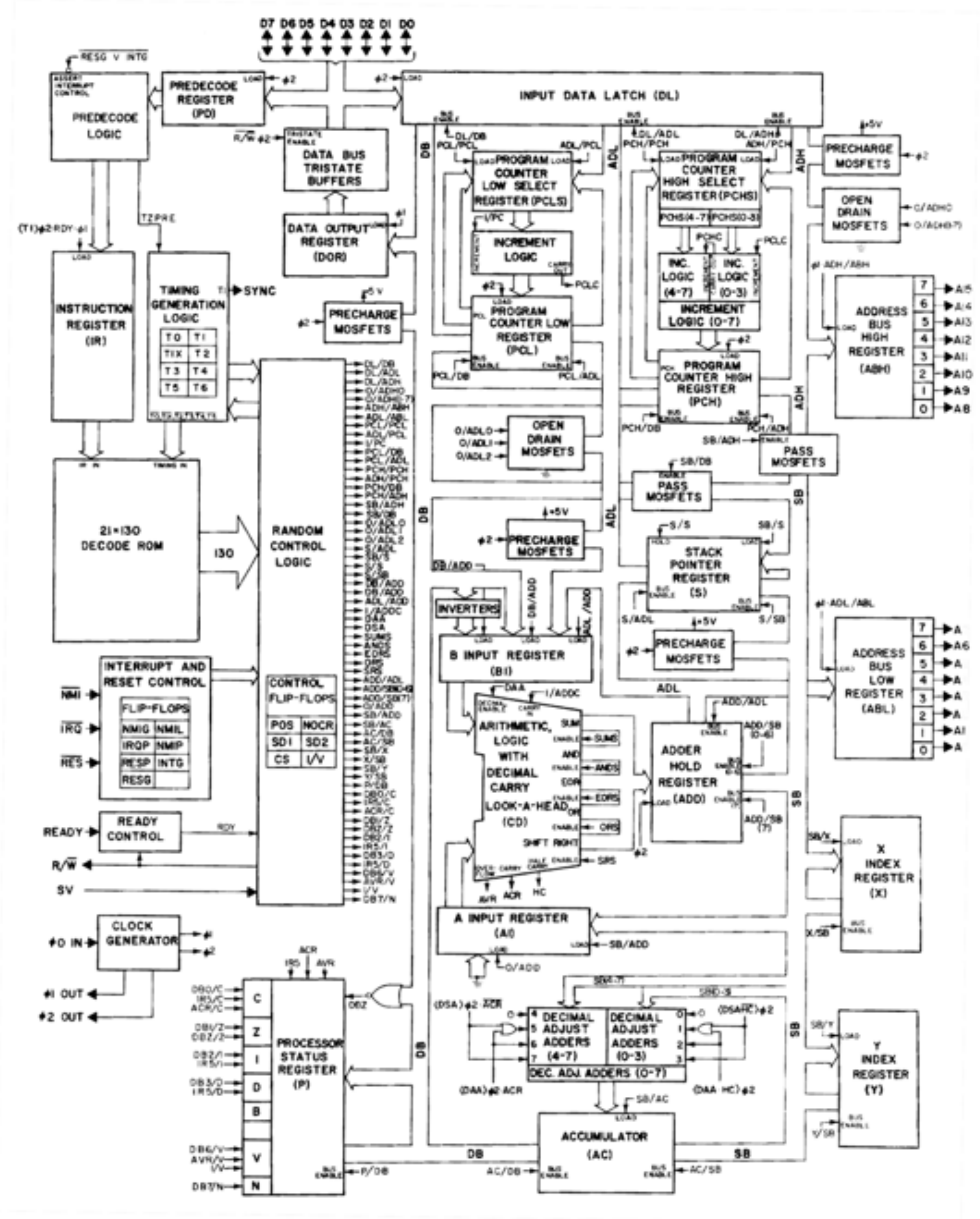
III: Neuroscience is in a conceptual crisis

Flavors of understanding

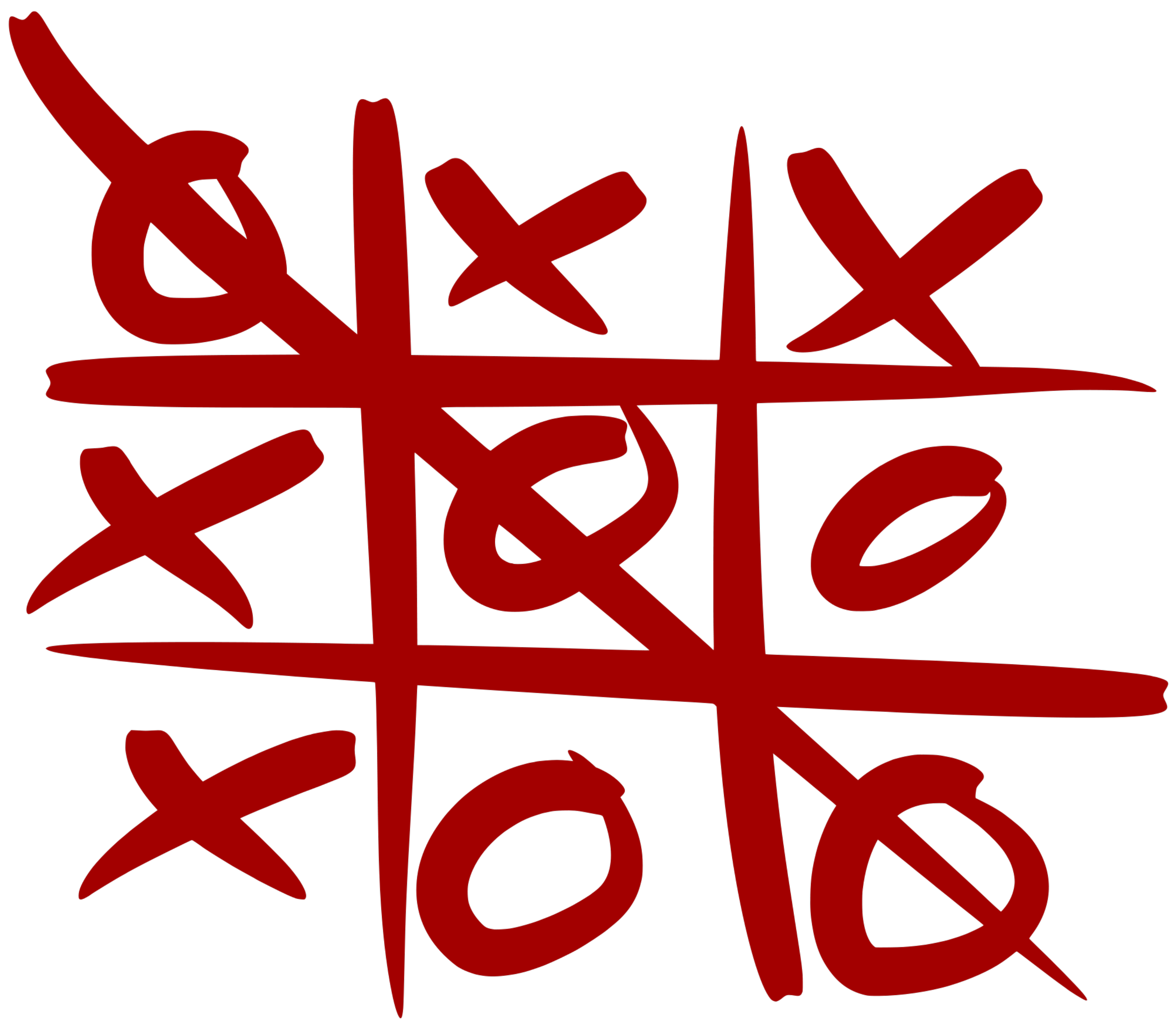
- Know some truths about it
- Fix it
- Simulate it
- Understand how it works

Microprocessor as a model



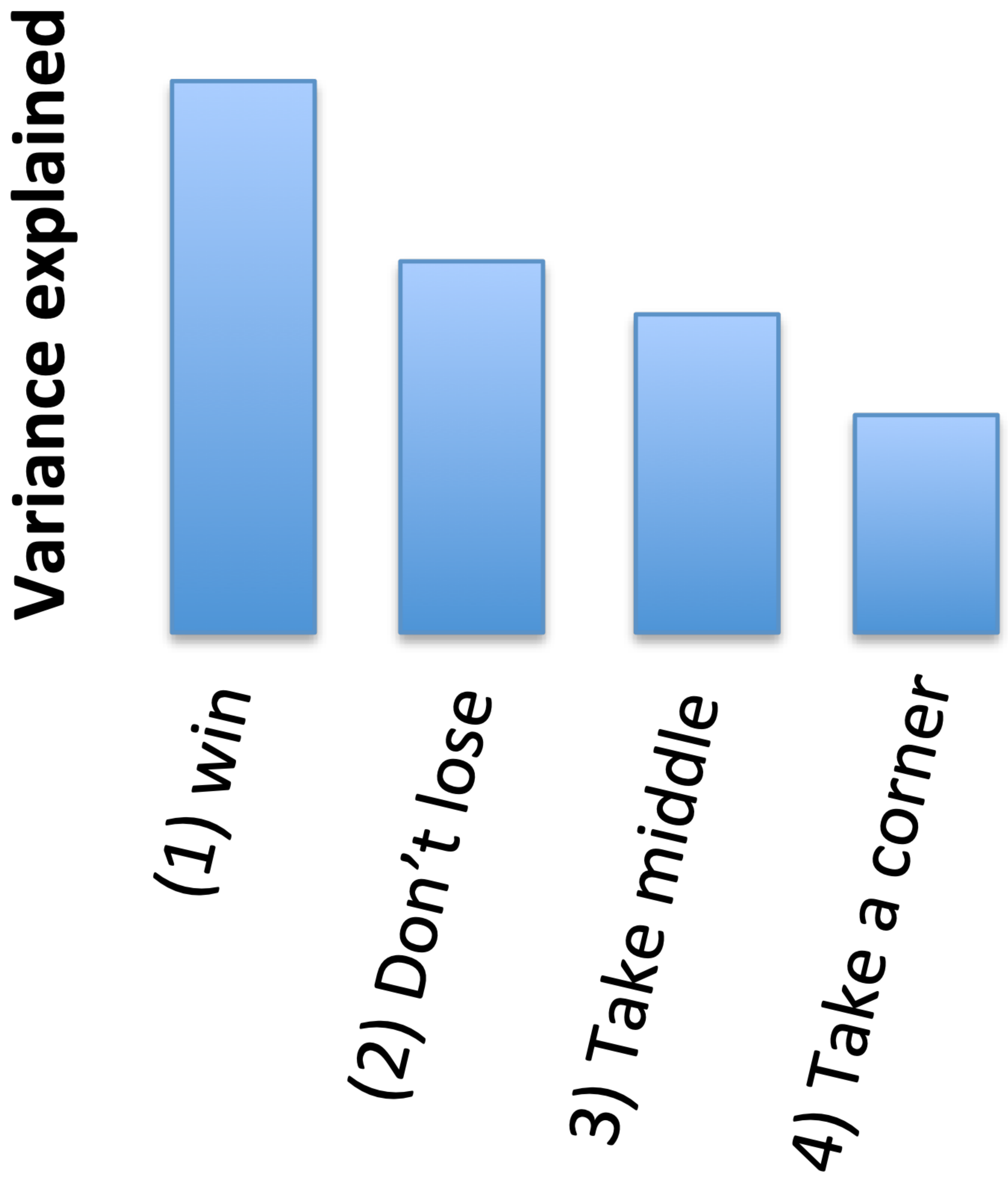


Tic Tac Toe



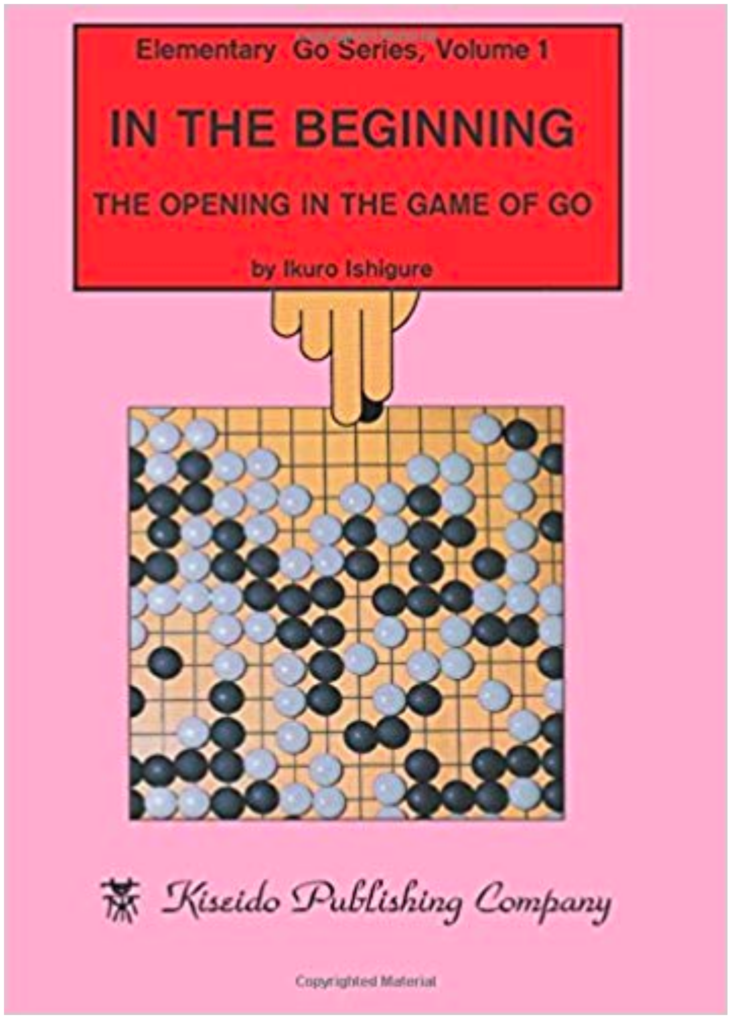
255,168 distinct games!

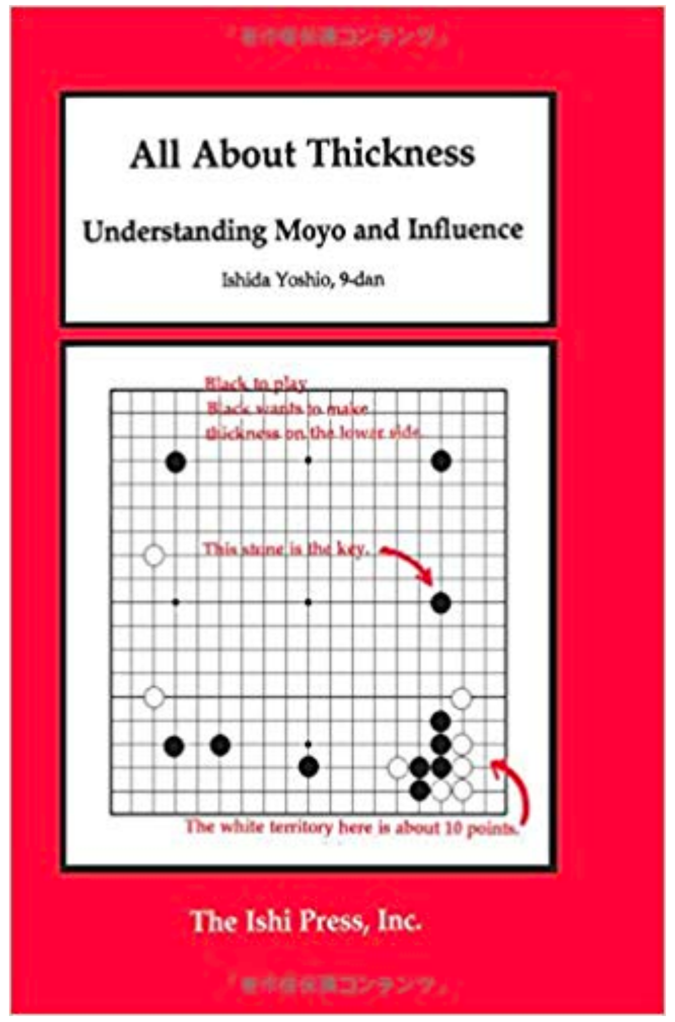
Compressable

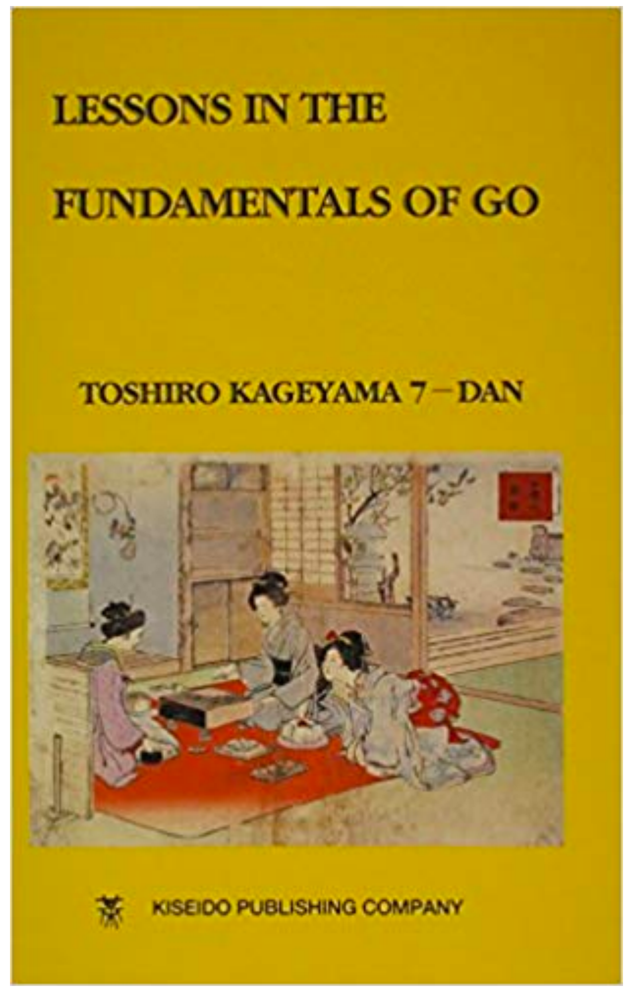


Go

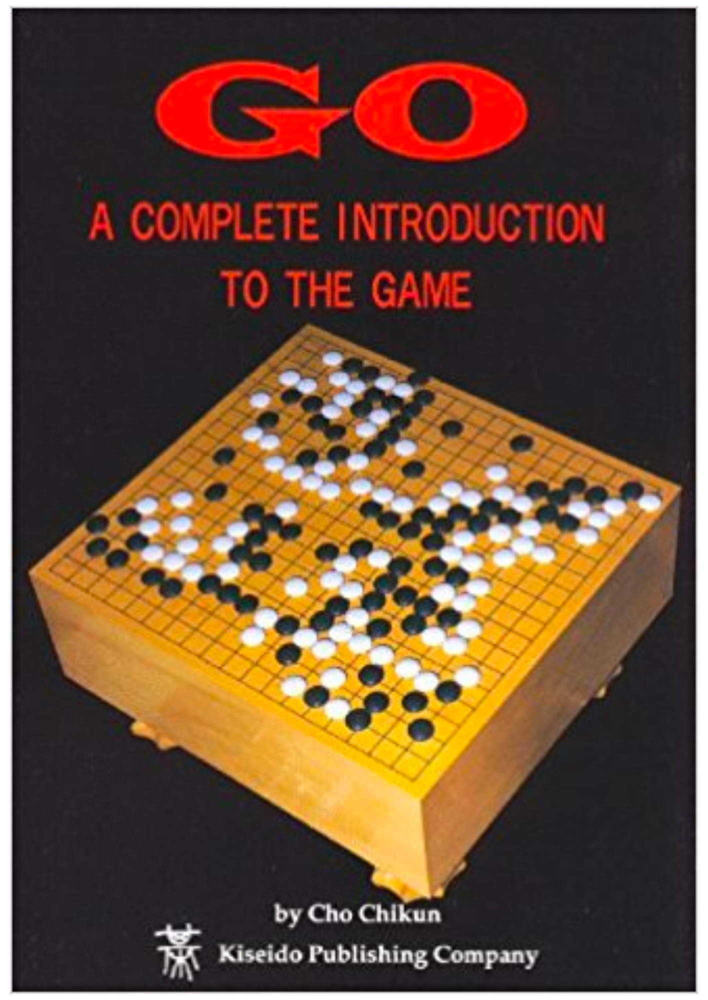
Probably no way to compactly describe it

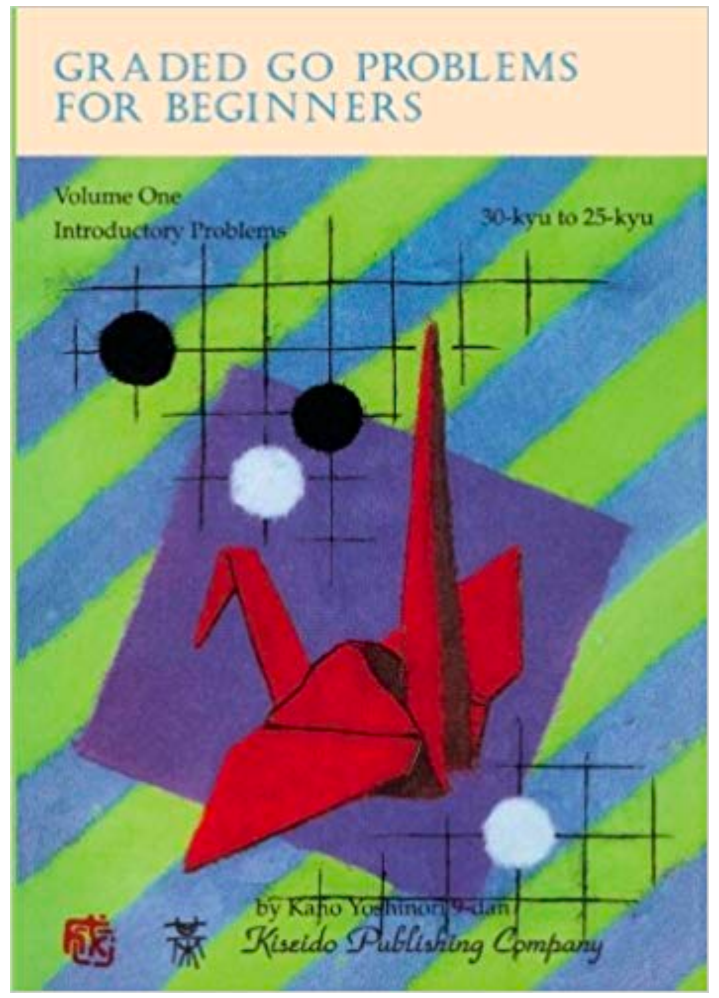


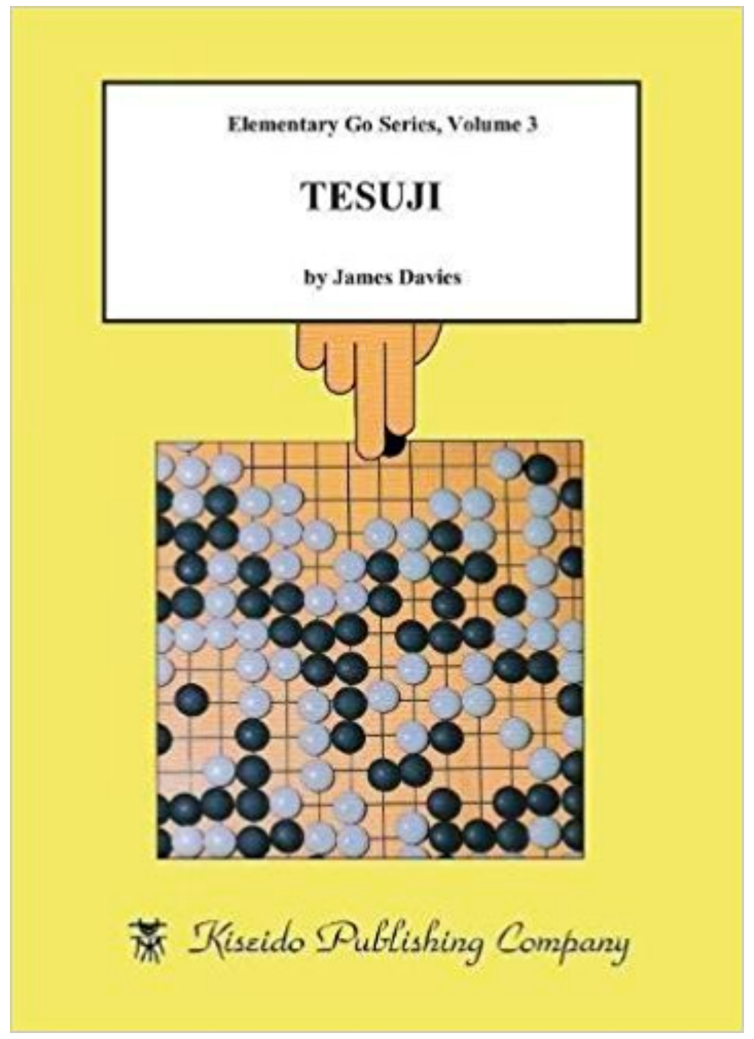


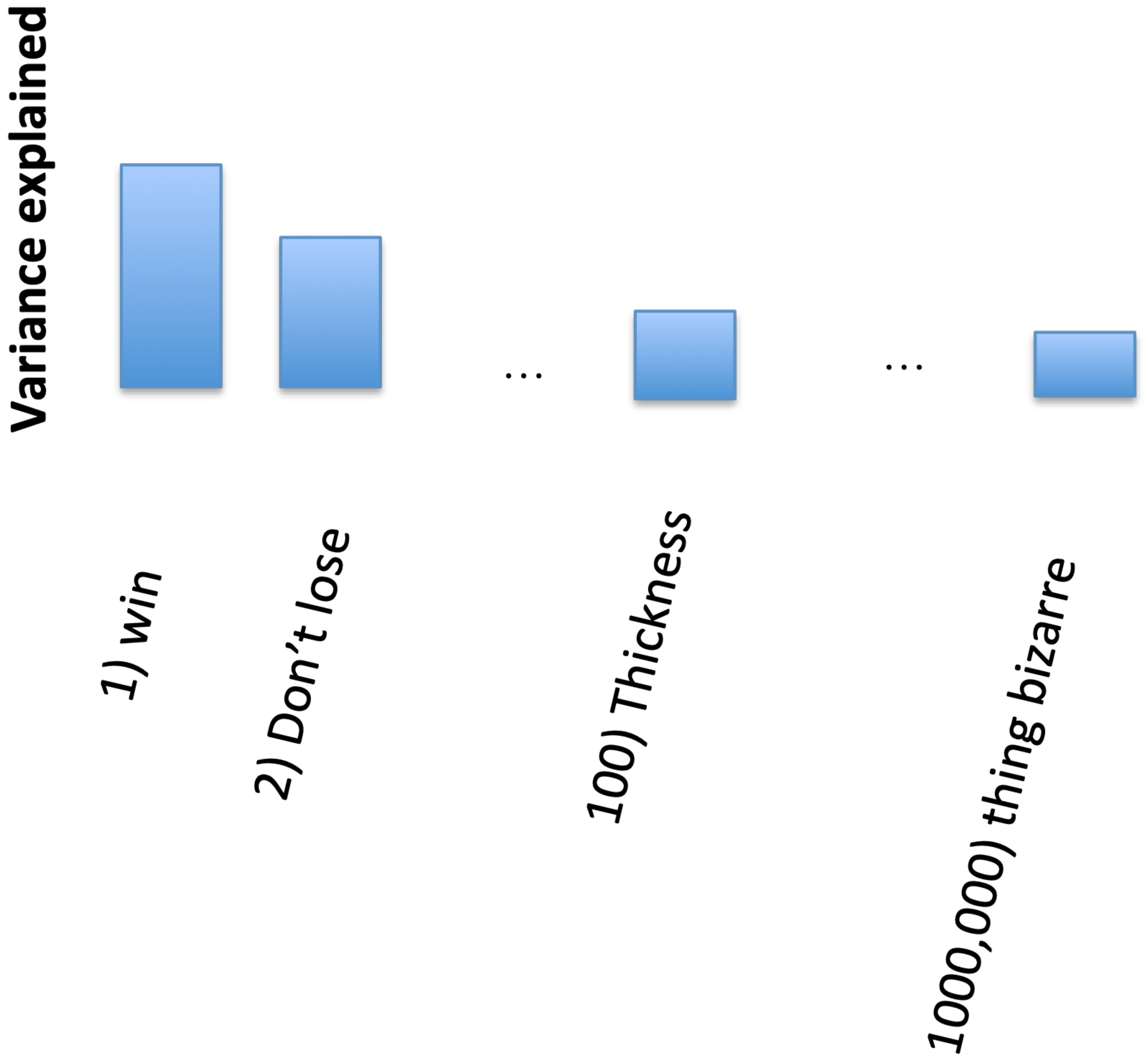


...









They are all real. Replicable from Go grand master to Go grandmaster.

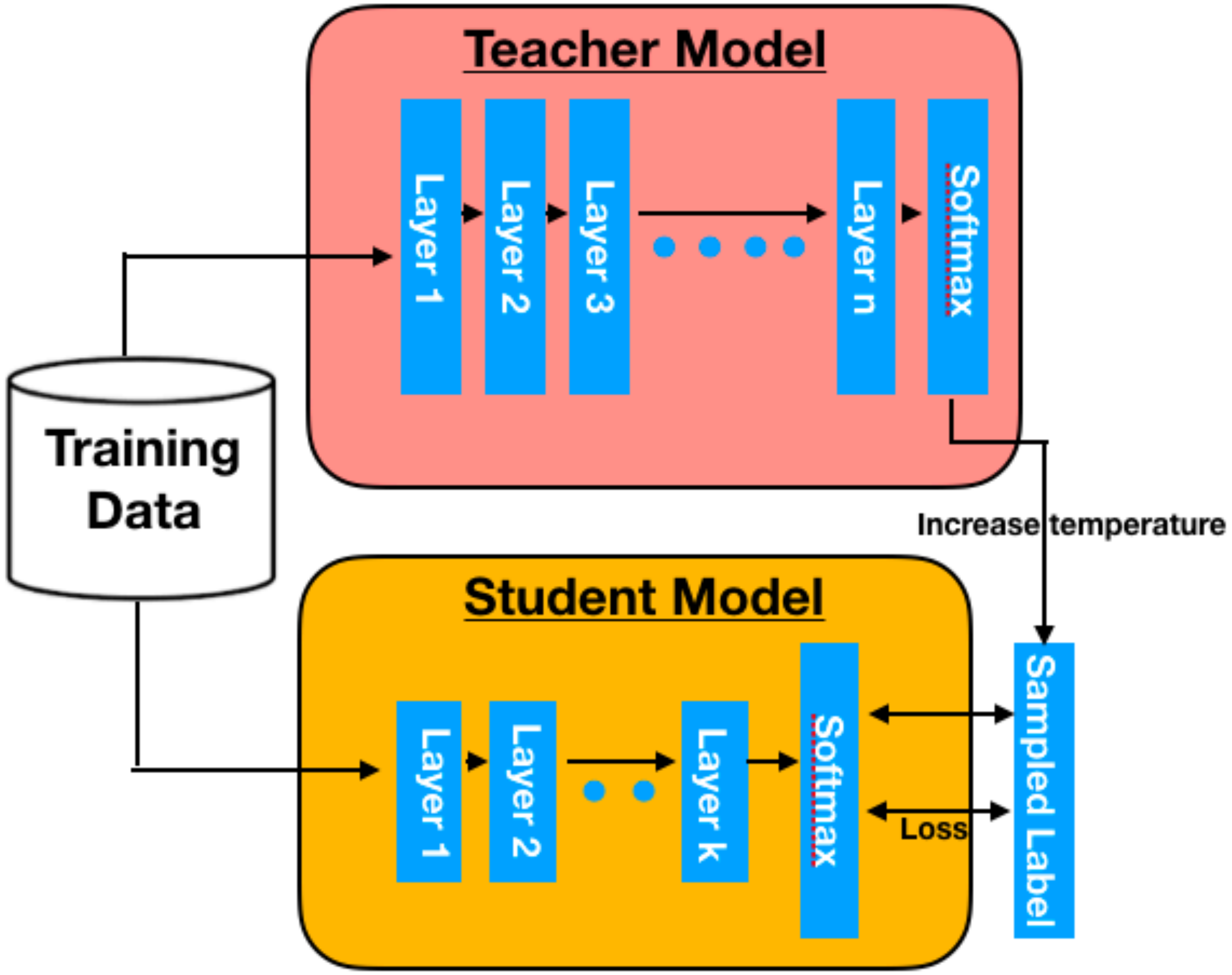
Understand a neural network

- **Pytorch code**
- Vs the **Weight dump**

Kolmogorov complexity of brain like computations

- Distillation
- Complexity calculations
- Back of the envelope calculations

Distillation



from mc.ai

Factor 10-100 on MNIST, imaginet

e.g. Ba and Caruana, Zhu et al 2018

Can we compress NNs?

- MNIST -> soft decision trees
    - BAD
- imagenet

Back of the envelope human

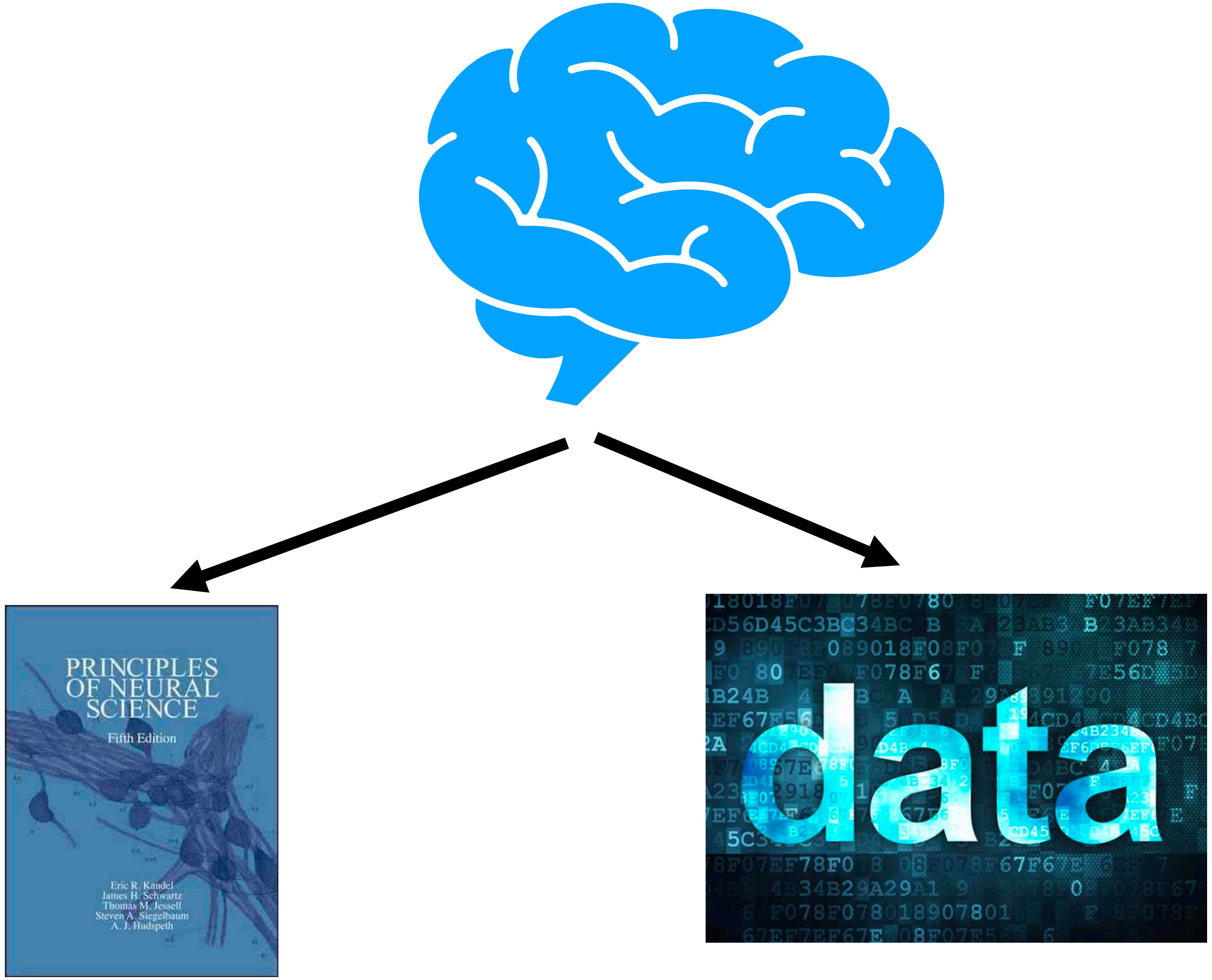
- 10 bits/s
- pi*10^7 seconds/a
- 30 years

- 10^10 bits
- 10^6 bits/book -> 10^4 books

H(DNA)<<H(World)

- DNA: 2*3*10^9 nucleotides
  - mostly non-nervous system
  - of nervous system possibly much non-computational
  - very non-compressed
- Nurture >> Nature

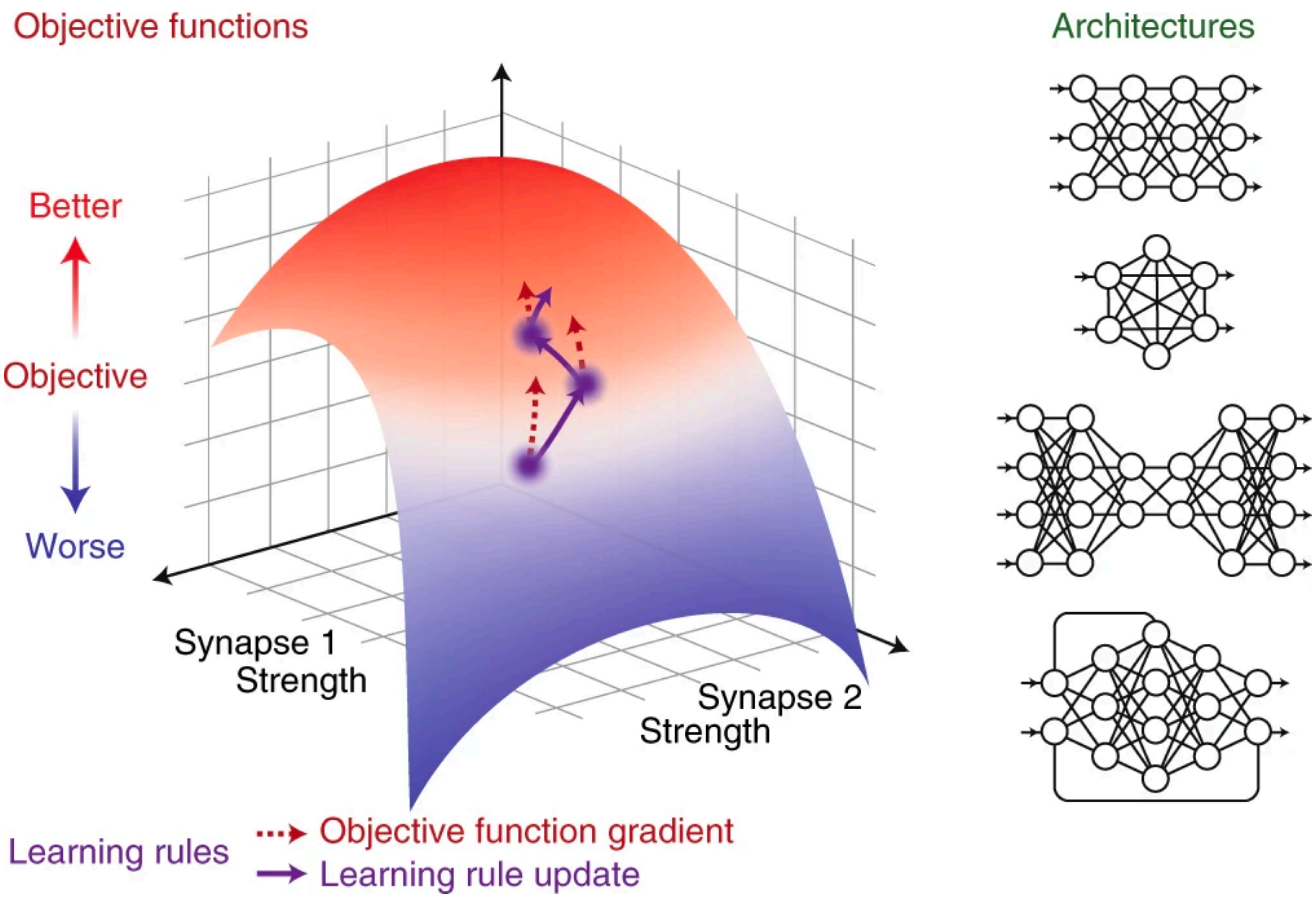
Ok. So what if the brain is not compressible?



Argument in a nutshell

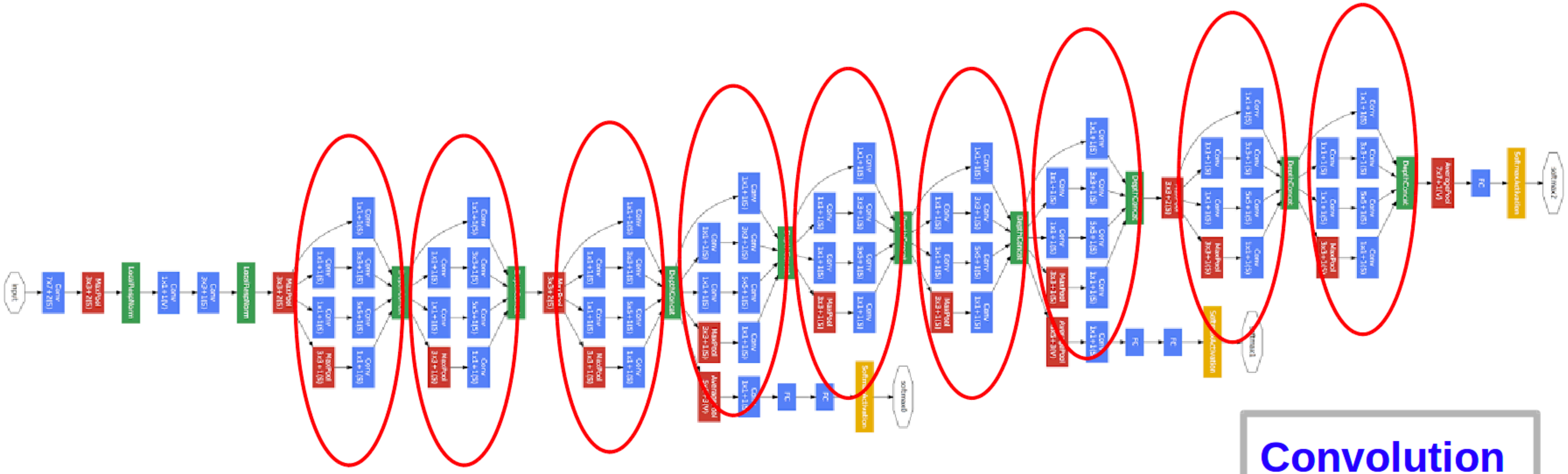
- Pytorch code to make an ANN is easy for us to understand
- Resulting network is (probably) impossible to understand
- So lets do the analogue of the first in neuroscience

So how may neuroscience have to look like?



Objective functions
Better
Objective
Worse
Synapse 1
Strength
Synapse 2
Strength
Architectures
Learning rules
Objective function gradient
Learning rule update

Anatomy (Googlenet)

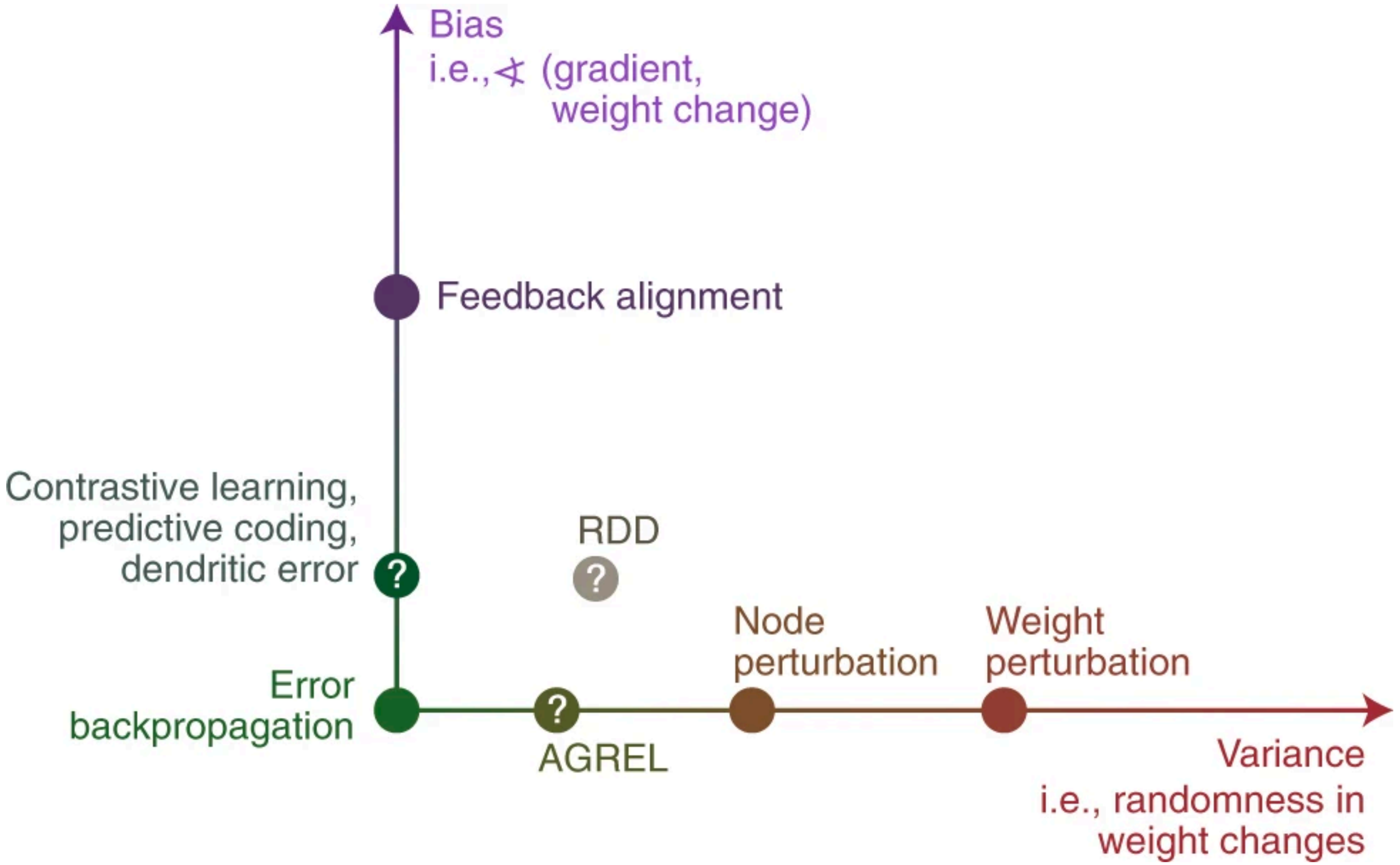


Convolution
Pooling
Softmax
Concat/Normalize

Objective function (softmax)

$$P(y = j \mid \mathbf{x}) = \frac{e^{\mathbf{x}^\top \mathbf{w}_j}}{\sum_{k=1}^{K} e^{\mathbf{x}^\top \mathbf{w}_k}}$$

We need computational ideas!



Summary

- Timely to think about DL as a model of the brain
- Weight transport is one problem
- Another problem is multiplexing
- And variance is always an issue
- DL may bring a new hope to neuroscience

Last lecture, let us look back!

- This was a great experience for me
- I learned a great deal
- We have outstanding TAs
- Sadat was an amazing lead TA
- I hope you are and stay well