Ising Machines and Intelligence

Yasser Roudi
Department of Mathematics
King's College London

A physicist’s approach to practically everything

- Milk production at a dairy farm was low, so the farmer wrote to the local university, asking for help from academia.

- Physicist: "I have the solution, but it works only in the case of spherical cows in a vacuum.”

  T. Lee, Washington Post 2013

But this approach has its merits, as evidenced by the success of physics

- It is complemented by a key observation: simple but not too simple.

Outline of the talk

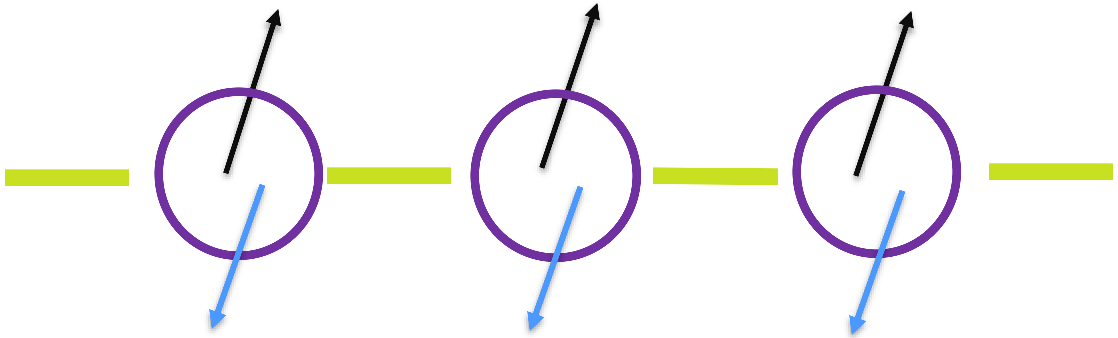
- Spherical Cow Models in vaccum: Ising model, Hopfield model and Restricted Boltzmann Machines

- Less spherical Cow Models

- Taking the Cow out of vaccum

Ferr/Ferri-magnets

Material	Curie temp. (K)
Co	1388
Fe	1043
Fe_2O_3[a]	948
$NiOFe_2O_3$[a]	858
$CuOFe_2O_3$[a]	728
$MgOFe_2O_3$[a]	713
MnBi	630
Ni	627
$Nd_2Fe_{14}B$	593
MnSb	587
$MnOFe_2O_3$[a]	573
$Y_3Fe_5O_{12}$[a]	560
CrO_2	386
MnAs	318
Gd	292
Tb	219
Dy	88
EuO	69



$$\sigma_i = \pm 1 \qquad H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

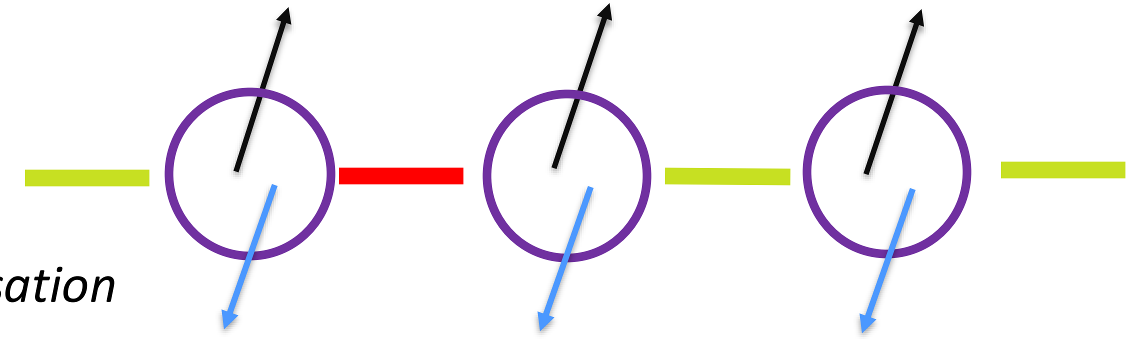
$$p(\sigma_1, \cdots, \sigma_N) \propto e^{-H(\sigma_1, \cdots, \sigma_N)/T}$$

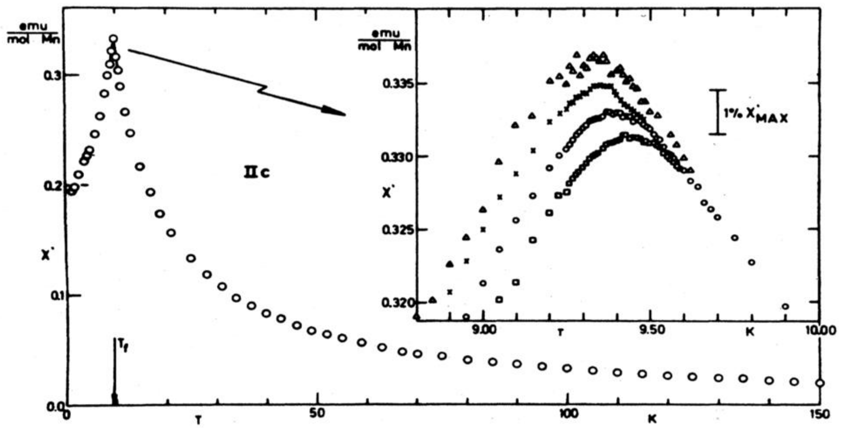
1D chain: analytical solution, no spontaneous magnetisation
2D: analytical solution, spontaneous magnetisation
3D: no analytical solution, spontaneous magnetisation

Spin Glasses: Quenched Disorder

Transition metal impurities in noble metal hosts

- *Local non-zero magnetisation, zero global magnetisation*
- *Strange heat capacity and susceptibility temperature*
- *Long relaxation time*



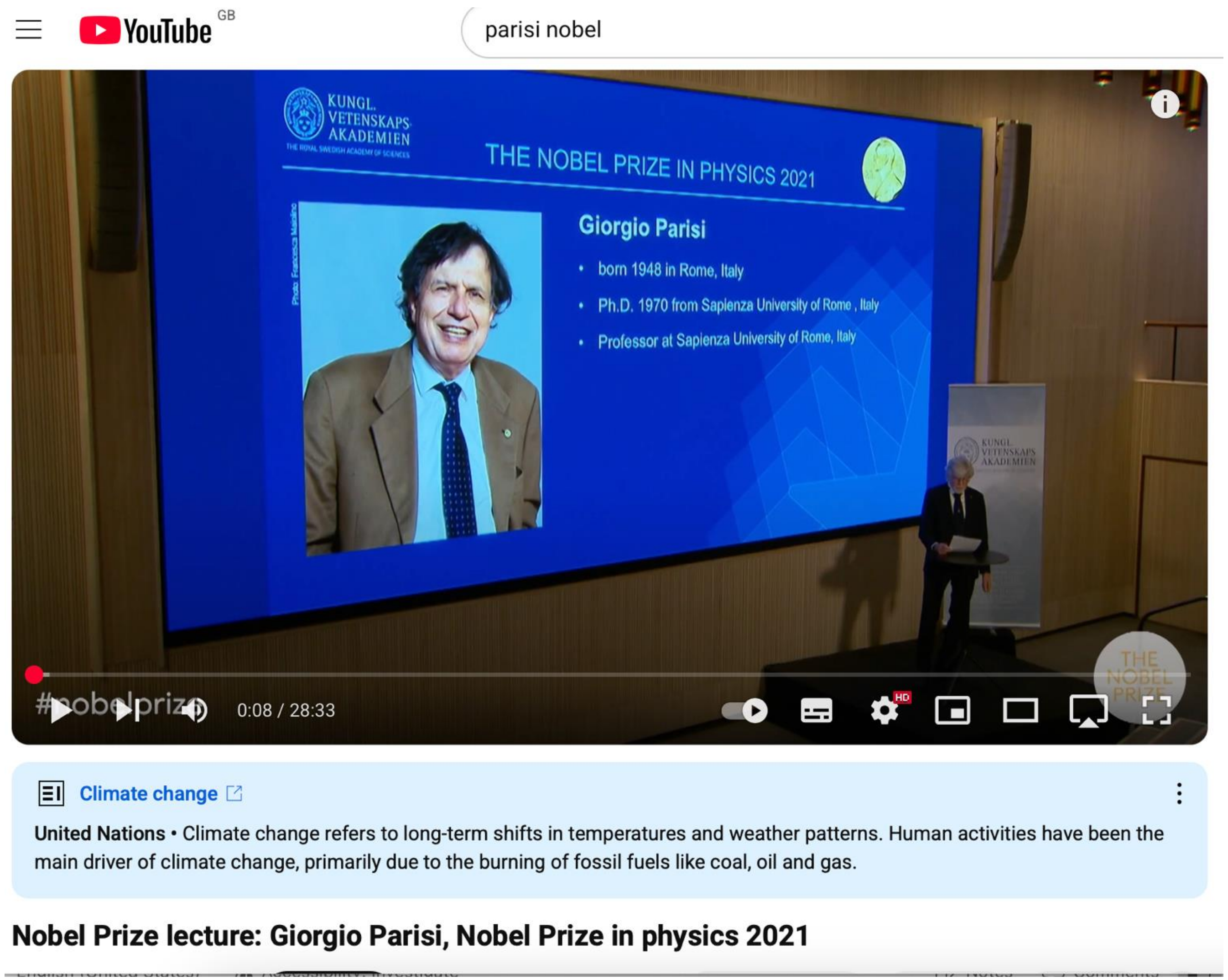


$$\sigma_i = \pm 1 \qquad H = -\sum_{\langle ij \rangle} J_{ij}\sigma_i\sigma_j$$

$$p(\sigma_1, \cdots, \sigma_N) \propto e^{-H(\sigma_1, \cdots, \sigma_N)/T}$$

$$p(J_{ij}) = \frac{1}{2}\delta(J_{ij} - J) + \frac{1}{2}\delta(J_{ij} + J)$$

Difficulty in disordered systems



Hebb, Hopfield, Amit and Gardner

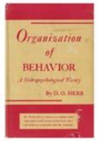
Plasticity and Learning

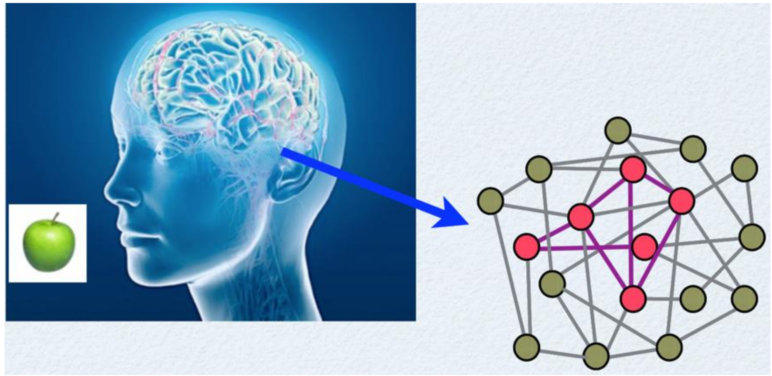
Synaptic Plasticity

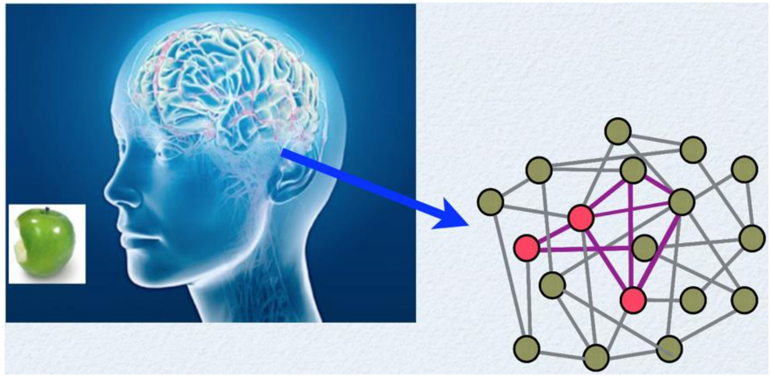
training and experience

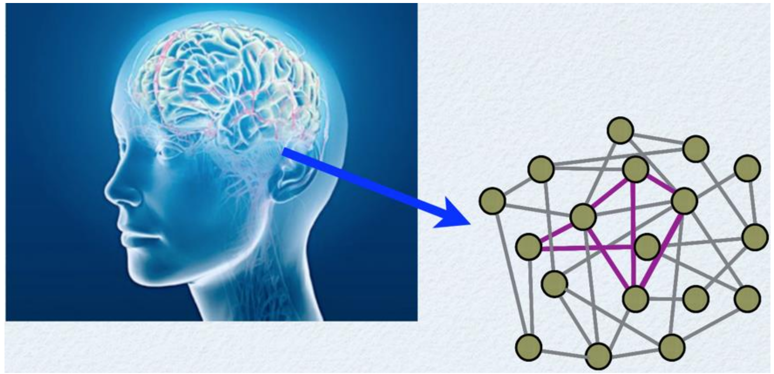
learning, memory and development of neural circuits

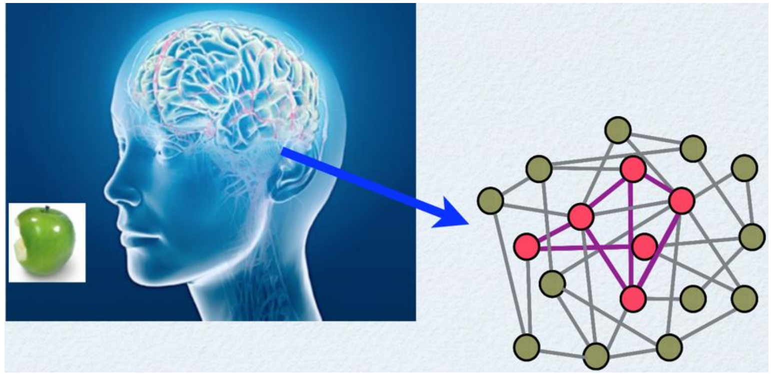
“Cells that fire together wire together”











Hebb, Hopfield, Amit (et al) and Gardner (et al)

John Hopfield

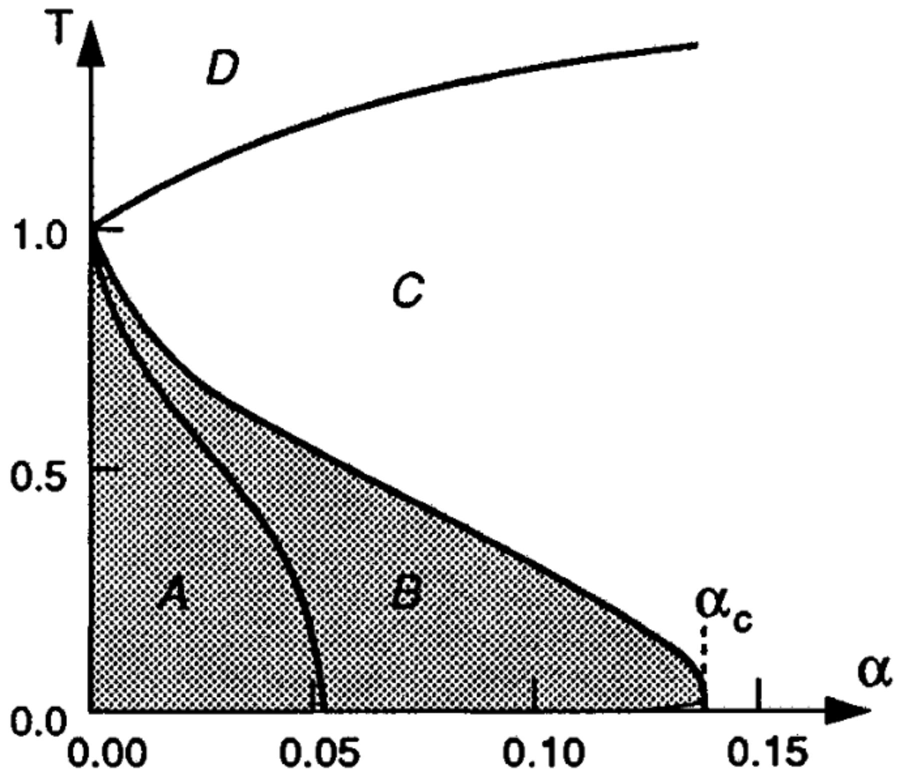
$$\sigma_i = \pm 1 \quad H = -\sum_{\langle ij \rangle} J_{ij}\sigma_i\sigma_j$$

$$p(\sigma_1, \cdots, \sigma_N) \propto e^{-H(\sigma_1, \cdots, \sigma_N)/T}$$

$$J_{ij} \propto \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu}$$

$\xi_i^{\mu} = 1$ If neuron *i* is active in pattern μ

Amit, Gutfreund and Sompolinsky



Hebb, Hopfield, Amit and Gardner

Fully connected (Hopfield Model) with Hebbian Learning

Hopfield 82, Amit et al 87

$$C = N - 1$$

$$N \to \infty$$

$$\alpha_c = \frac{P}{N} = 0.14$$

Extremely diluted network with Hebbian Learning

Derrida, Gardner, Zippelius 87

$$N \to \infty$$

$$C \to \infty$$

$$\frac{C}{N} \to 0$$

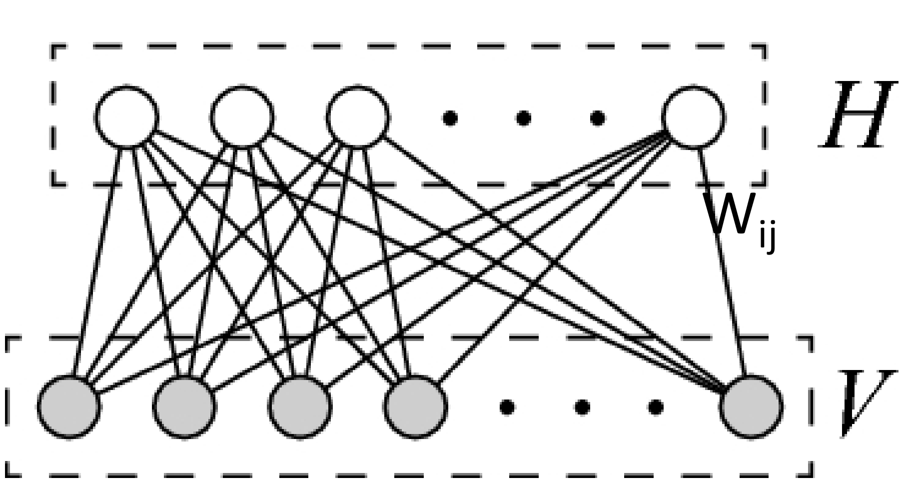
$$\alpha_c = \frac{P}{C} = \frac{2}{\pi}$$

Machines learn statistics: Restricted Boltzmann Machines

Hidden units $\boldsymbol{z} = (z_1, \ldots, z_M) \in \{0,1\}^M$

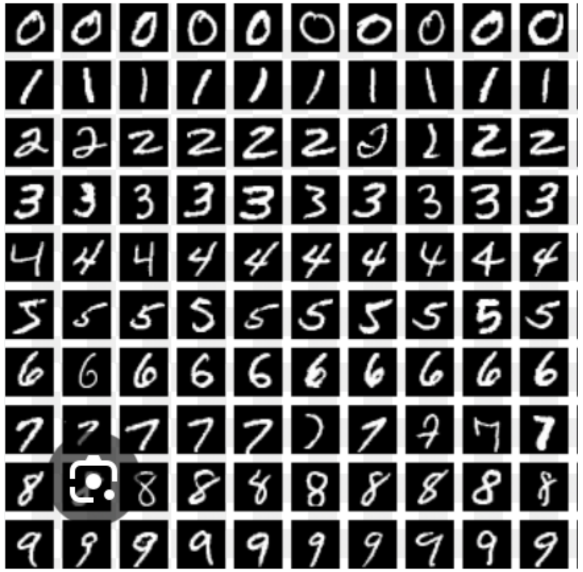
Visible units $\boldsymbol{v} = (v_1, \ldots, v_N) \in \{0,1\}^N$

$P(v_i|\boldsymbol{z}) = \psi \left(\sum_j W_{i\mu} z_\mu + b_i \right)$ $P(z_\mu|\boldsymbol{v}) = \sigma \, \psi \left(\sum_j W_{\mu j} v_j + c_j \right)$



$$P(\boldsymbol{v}, \boldsymbol{z}) = \frac{1}{Z} \exp \left[\boldsymbol{b}^\intercal \boldsymbol{v} + \boldsymbol{v}^\intercal \boldsymbol{W} \boldsymbol{z} + \boldsymbol{c}^\intercal \boldsymbol{z} \right]$$

Given a set of examples of v, find the best parameters that describe the statistical properties of the examples

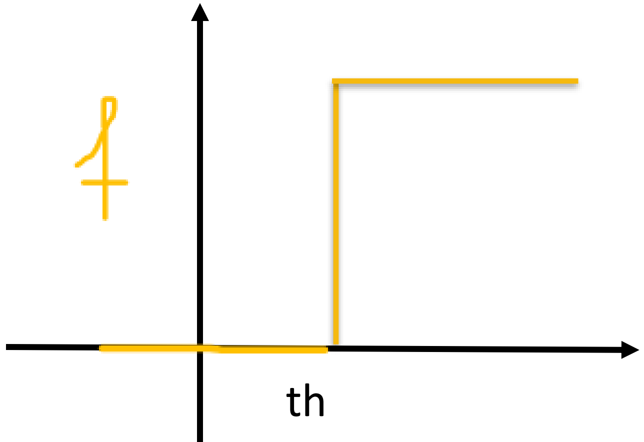


$$\Delta w_{i\mu} = \eta \left(\langle v_i z_\mu \rangle_{\text{data}} - \langle v_i z_\mu \rangle_{\text{model}} \right)$$

Spherical Cow Aspects

- Neurons do not have only high firing/low firing states (non-binary activation function)
- Neurons are either excitatory or inhibitory (Dale's law)
- The dynamics is not in equilibrium

The role of activation function



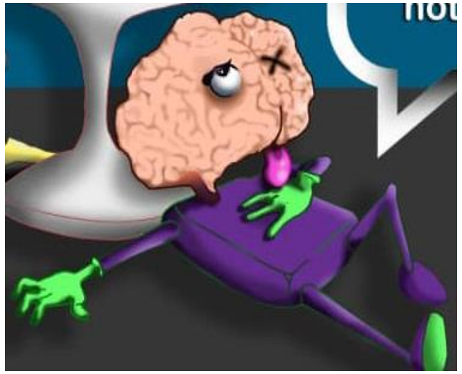
The Role of Activation Function on Capactiy

Hebb, Hopfield, Amit and Gardner

Fully connected (Hopfield Model) with Hebbian Learning

$$\alpha_c = \frac{P}{N} = 0.14$$

Hopfield 82, Amit et al 87



Extremely diluted network with Hebbian Learning

$$\alpha_c = \frac{P}{C} = \frac{2}{\pi}$$

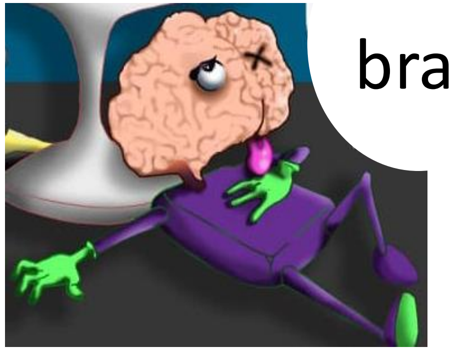
Derrida, Gardner, Zippelius 87

Optimal Capacity

$$\alpha_c = 2$$

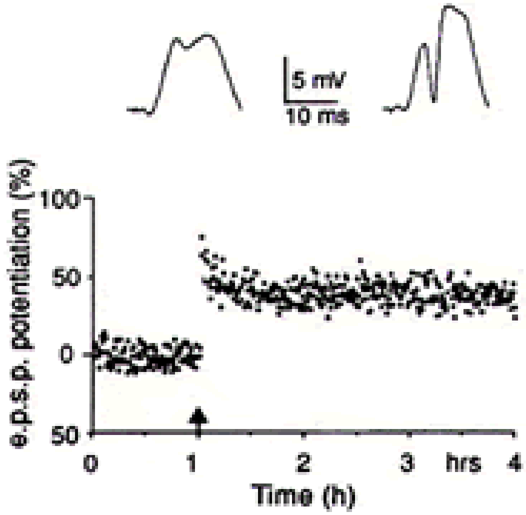
Back Propagation can reach this

Hebbian v Back Propagation

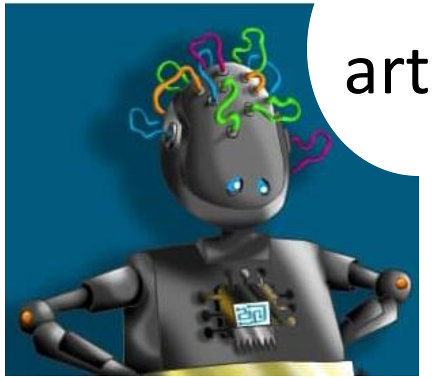


brain: synaptic plasticity mechanisms, e.g. LTP and LTD

depend only on pre and post synaptic firing; so they are <u>local</u>



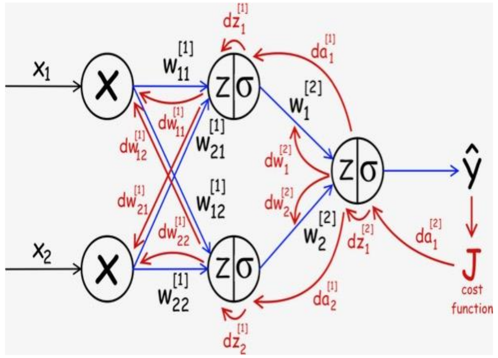
Bliss and Lømo (1973)



artificial NN: iterative algorithms based on Back Propagation

depend in complex ways on the activity of neurons other than the pre and post, so they are <u>non-local</u>

<u>not biologically plausible</u>



credit: http://www.adeveloperdiary.com

a general consensus

non-local iterative algorithms

≫

local biologically plausible ones

ppl argue BP can be/is implemented in the brain

Backpropagation and the brain

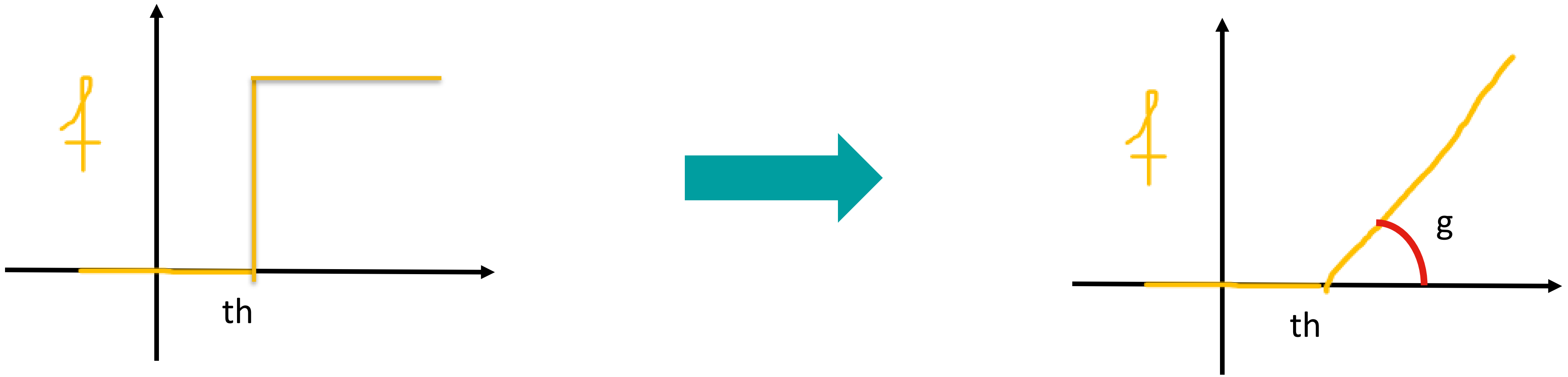
Timothy P. Lillicrap, Adam Santoro, Luke Marris, Colin J. Akerman and Geoffrey Hinton

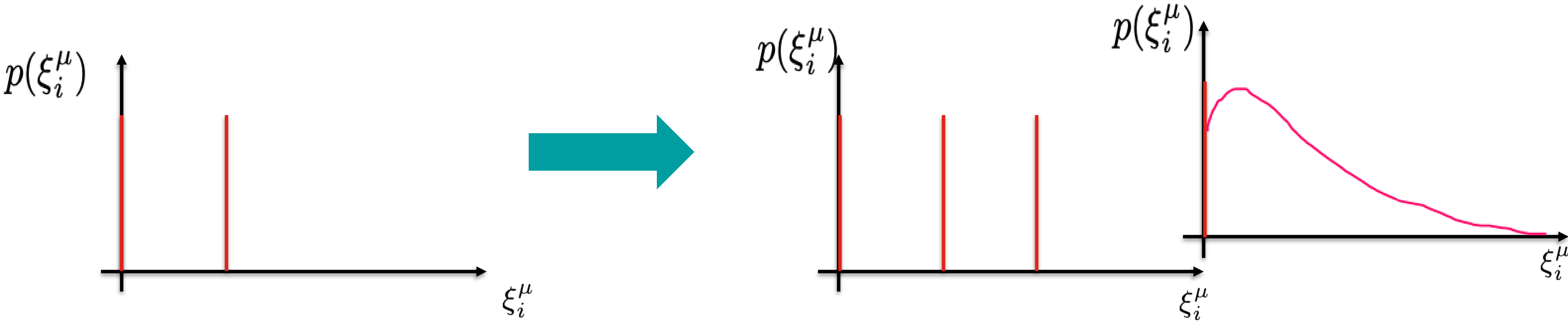
Abstract | During learning, the brain modifies synapses to improve behaviour. In the cortex, synapses are embedded within multilayered networks, making it difficult to determine the effect of an individual synaptic modification on the behaviour of the system. The backpropagation algorithm solves this problem in deep artificial neural networks, but historically it has been viewed as biologically problematic.

formulations of backpropagation. Here we build on past and recent developments to argue that feedback connections may instead induce neural activities whose differences can be used to locally approximate these signals and hence drive effective learning in deep networks in the brain.

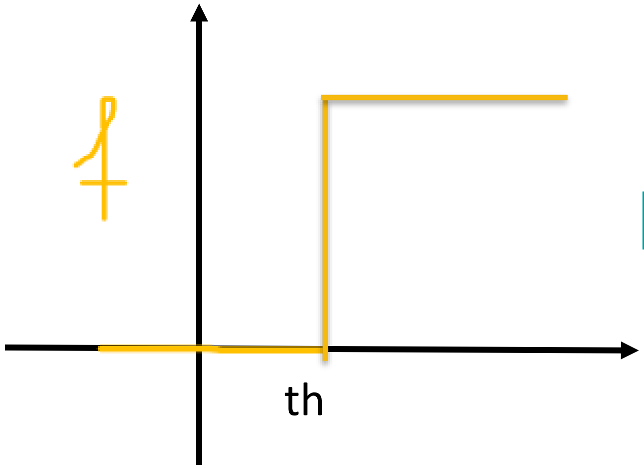
Nat. Rev. Neuro. (2020)

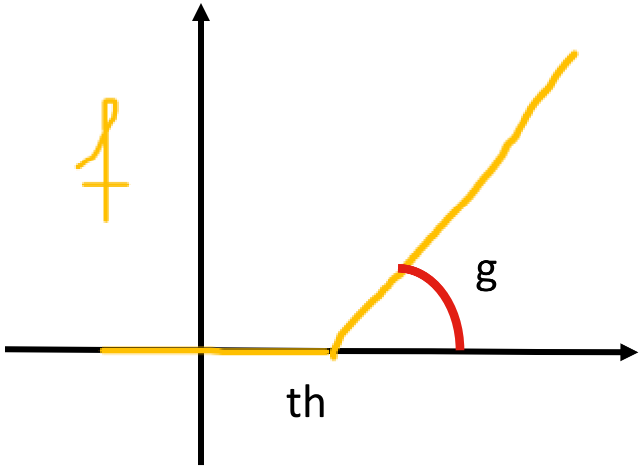
Role of Activation Function on Capacity

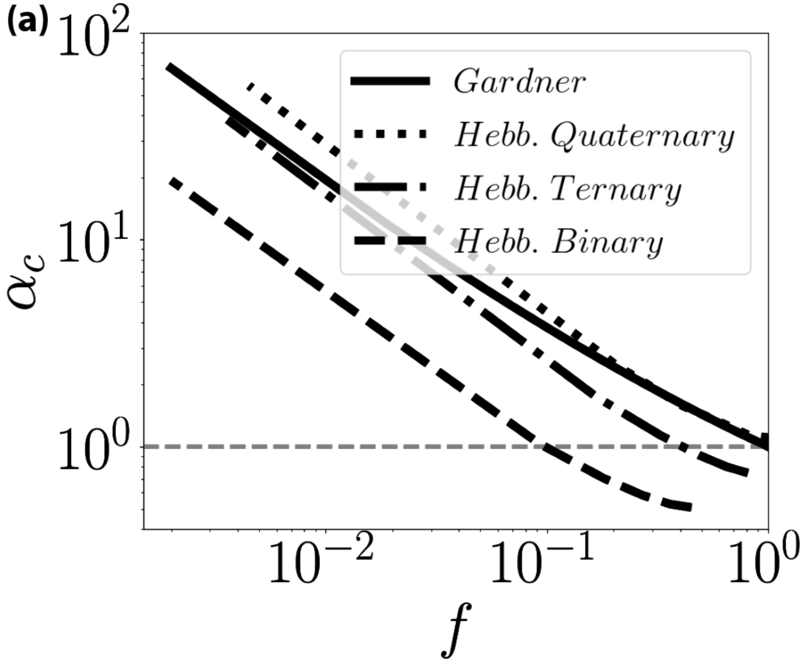


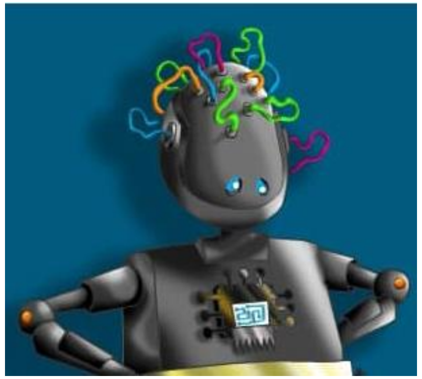


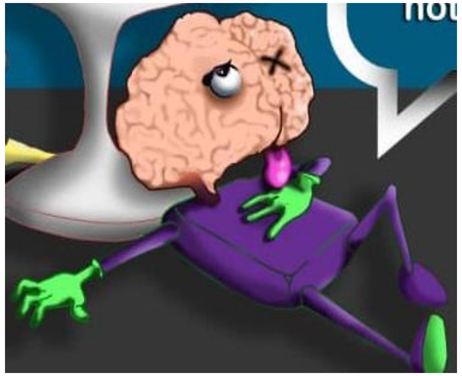
Role of Activation Function on Capacity









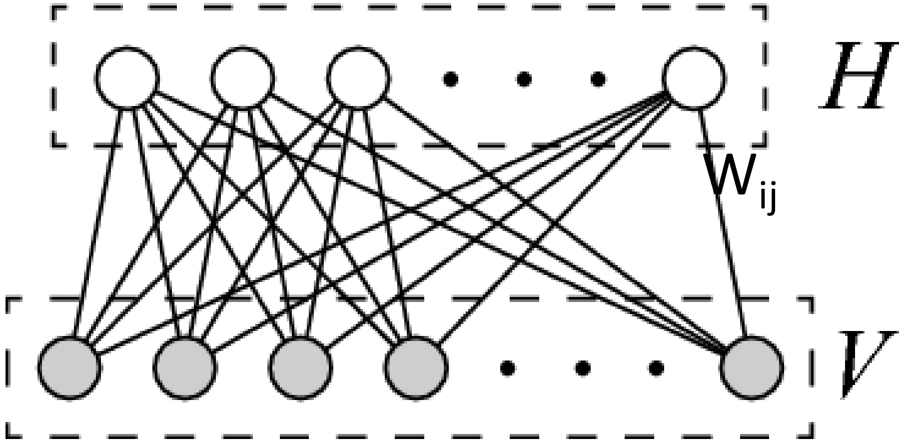


The Role of Activation Function in RBM

The Role of Activation Function in RBMs

Hidden units $\boldsymbol{z} = (z_1, \ldots, z_M) \in \{0,1\}^M$

Visible units $\boldsymbol{v} = (v_1, \ldots, v_N) \in \{0,1\}^N$



$$P(\boldsymbol{v}, \boldsymbol{z}) = \frac{1}{Z} \exp\left[\boldsymbol{b}^\top \boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{W} \quad + \boldsymbol{c}^\top \boldsymbol{z}\right]$$

$$P(\boldsymbol{v}, \boldsymbol{z}) = \frac{1}{Z} \exp\left[\boldsymbol{b}^\top \boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{W} \boldsymbol{z} - U(\boldsymbol{z})\right]$$

$$U(\boldsymbol{z}) = \sum_\mu U_\mu(z_\mu)$$

Role of Activation Function in RBMs

Main Mathematical Result

$$\rho(\boldsymbol{z}) = \frac{\exp[-U(\boldsymbol{z})]}{\int d\boldsymbol{z} \exp[-U(\boldsymbol{z})]}. \qquad K(\boldsymbol{q}) \equiv \log \int d\boldsymbol{z} \exp(\boldsymbol{q}^\intercal \boldsymbol{z})\, \rho(\boldsymbol{z}).$$

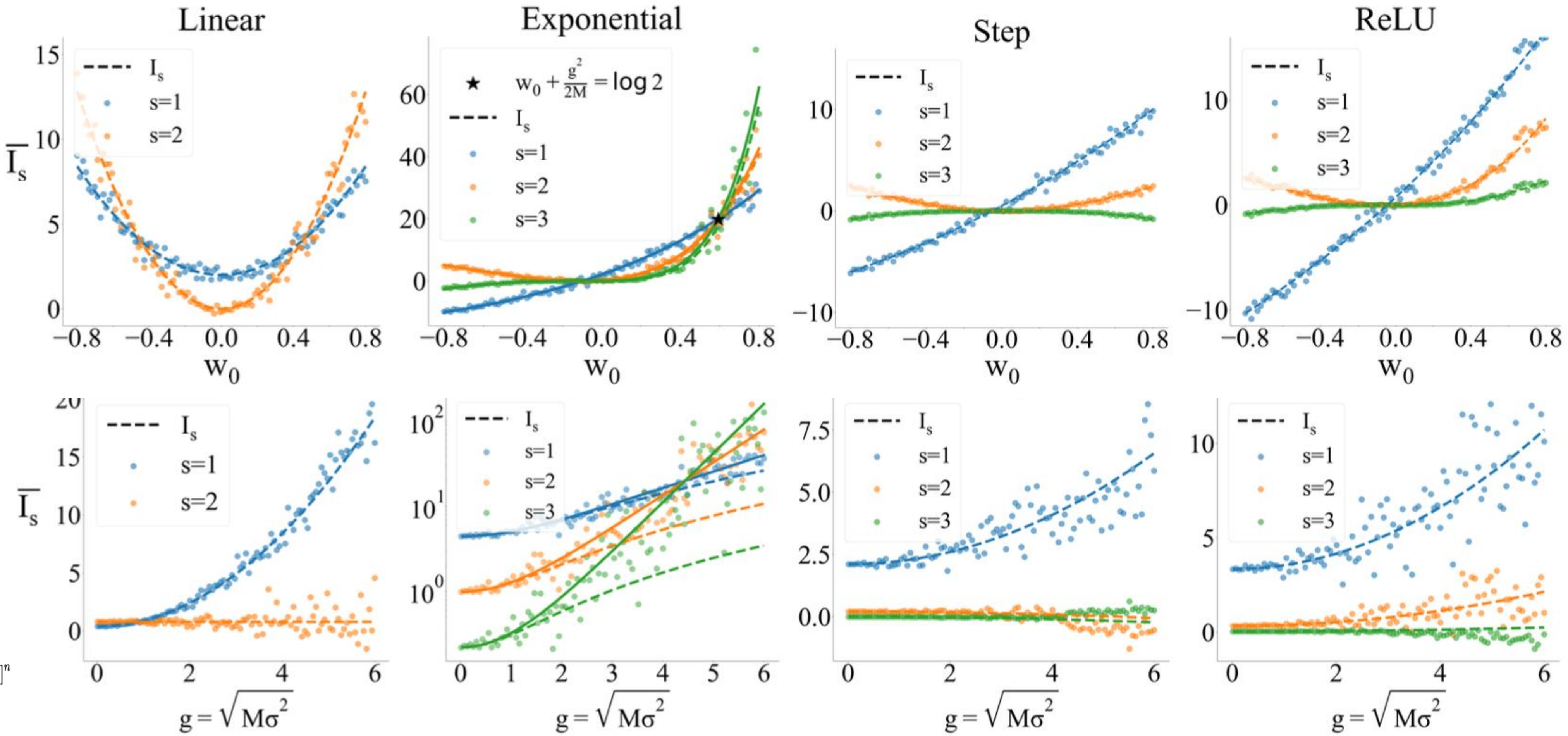
$$P(\boldsymbol{v}) = \frac{1}{Z'} \exp\left(\sum_{k_1} I^{(1)}_{k_1} v_{k_1} + \sum_{k_1<k_2} I^{(2)}_{k_1,k_2} v_{k_1} v_{k_2} + \cdots + I^{(N)}_{1,2,\dots,N} \prod_{k=1}^{N} v_k \right) \qquad I^{(s)}_{k_1,k_2,\dots,k_s} = \sum_{\mu=1}^{M} \sum_{p=0}^{s-1} (-1)^p \sum_{j_1<j_2<\cdots<j_{s-p}=1}^{s} K\left(\sum_{l=1}^{s-p} w_{k_{j_l},\mu} \right)$$

Role of Activation Function in RBMs

Linear Activation
Only produces pairwise interactions between visible units.

Exponential Activation
Can produce stronger higher-order interactions than lower-order ones.

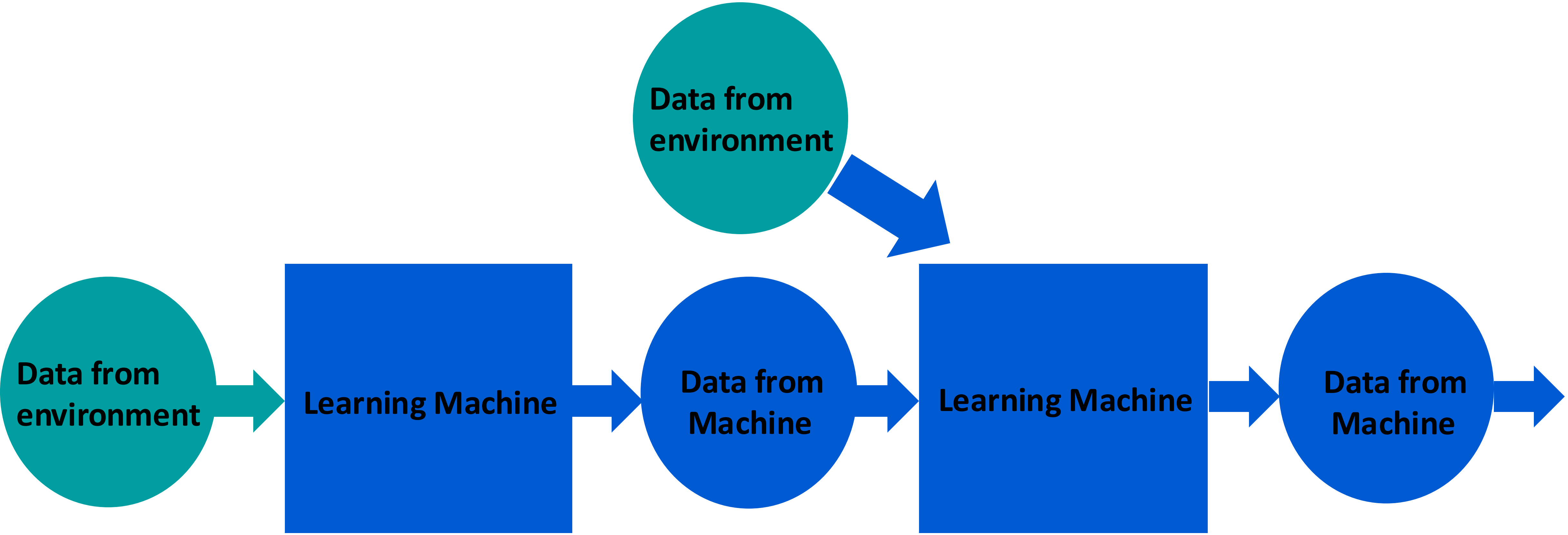
Step & ReLU
Produce interaction structures where lower-order interactions dominate



$$\overline{I_s^n} \equiv \binom{N}{s}^{-1} \sum_{k_1<\dots<k_s}^{N} [I_{k_1,\cdots,k_s}]^n$$

Taking the cow out of the vacuum

Two Schemes of Learning



A simplified model

$$f(s|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{\sum_a \theta_a \phi_a(s)}, \quad Z(\boldsymbol{\theta}) = \sum_s e^{\sum_a \theta_a \phi_a(s)}$$

1: draw M independent samples, $\hat{s}^1(t), \ldots, \hat{s}^M(t)$, from Eq. (1) with $\boldsymbol{\theta} = \boldsymbol{\theta}_t$ and m samples $\hat{s}^{M+1}(t), \ldots, \hat{s}^{M+m}(t)$ generated from a distribution $q(s)$. Denoting these samples as $\hat{\boldsymbol{s}}(t) \equiv (\hat{s}^1(t), \ldots, \hat{s}^{M+m}(t))$, their likelihood is

$$p(\hat{\boldsymbol{s}}(t)|\boldsymbol{\theta}_t) = \prod_{i=1}^{M} f(\hat{s}^i(t)|\boldsymbol{\theta}_t) \prod_{j=1}^{m} q(\hat{s}^j(t)).$$

2: Given $\hat{\boldsymbol{s}}(t)$ find the parameters

$$\boldsymbol{\theta}_{t+1} \equiv \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{s}}(t)) = \arg\max_{\boldsymbol{\theta}} \left[p(\hat{\boldsymbol{s}}(t)|\boldsymbol{\theta}) p_0^u(\boldsymbol{\theta}) \right].$$

loop: Set $t \to t+1$ and repeat.

Maximum Likelihood (ML) and no additional data: u=0 and m = 0

Maximum a Posteriori and no additional data: u=1 and m = 0

Data from environment

$q(\cdot)$

Machine

$f(\cdot|\theta)$

$p_0(\cdot)$

Closed Loop Learning and Model Collapse under ML

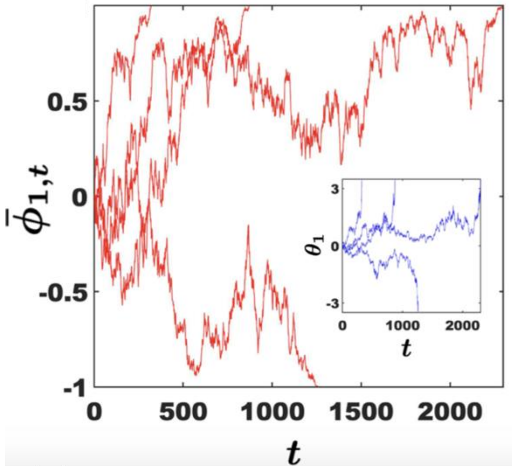
$$\lim_{t\to\infty} f(s|\boldsymbol{\theta}_t) = c\prod_a \delta(\phi_a(s) - \bar{\phi}_{a,\infty}) \qquad \overline{\phi}_{a,t} \equiv \frac{1}{M}\sum_l \phi_a(\hat{s}^l)$$

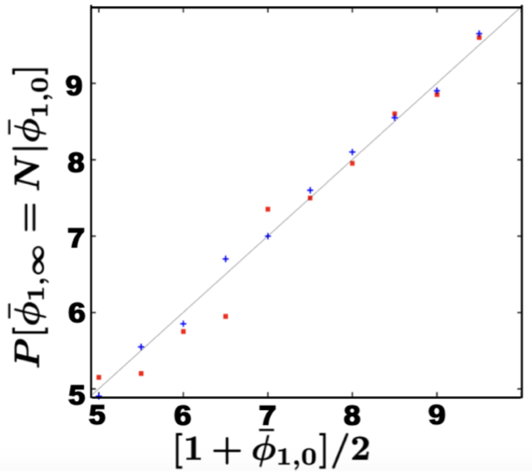
$$S(\boldsymbol{\varphi}) = \log Z - \sum_a \theta_a \varphi_a \qquad \langle S\rangle_{\boldsymbol{\varphi}_\tau} = \langle S\rangle_{\boldsymbol{\varphi}_0} - \tau\, D$$

Example: Ising Learning Machine

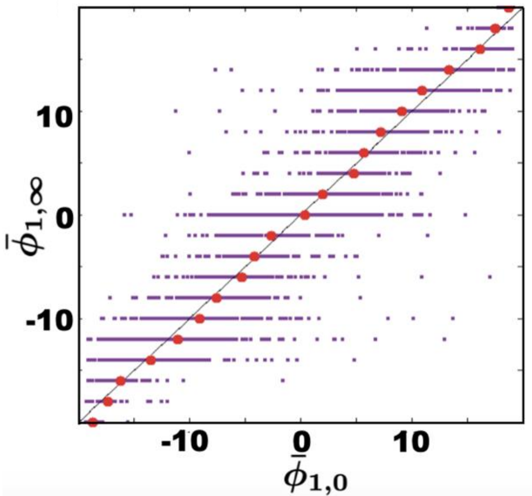
$$f(s|\theta_1,\theta_2) = \frac{1}{Z}\exp\left[\theta_1 \sum_i \sigma_i + \frac{\theta_2}{N}\sum_{i<j}\sigma_i\sigma_j\right].$$

$$\theta_2 = 0$$



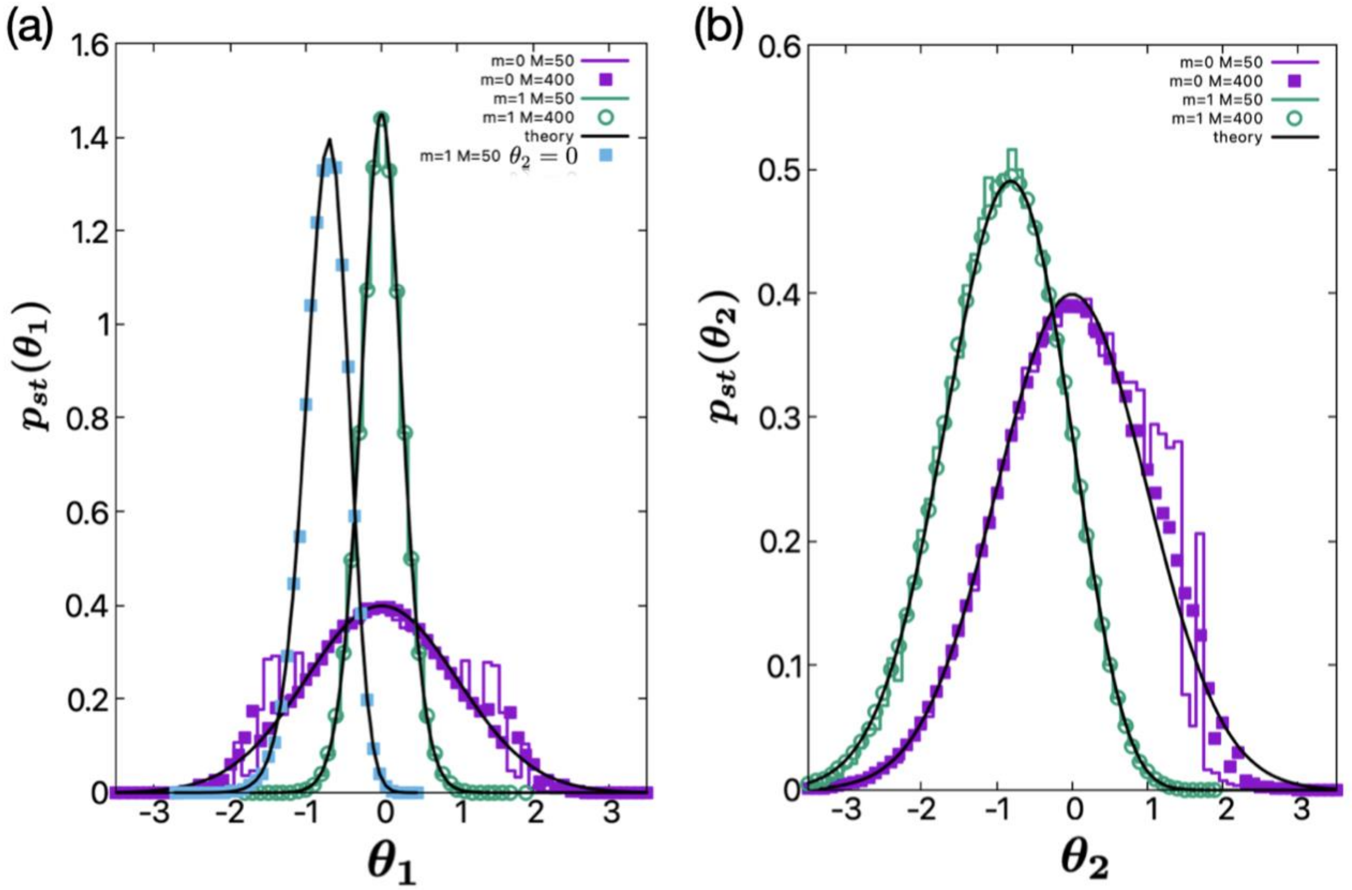


$$\theta_2 \neq 0$$



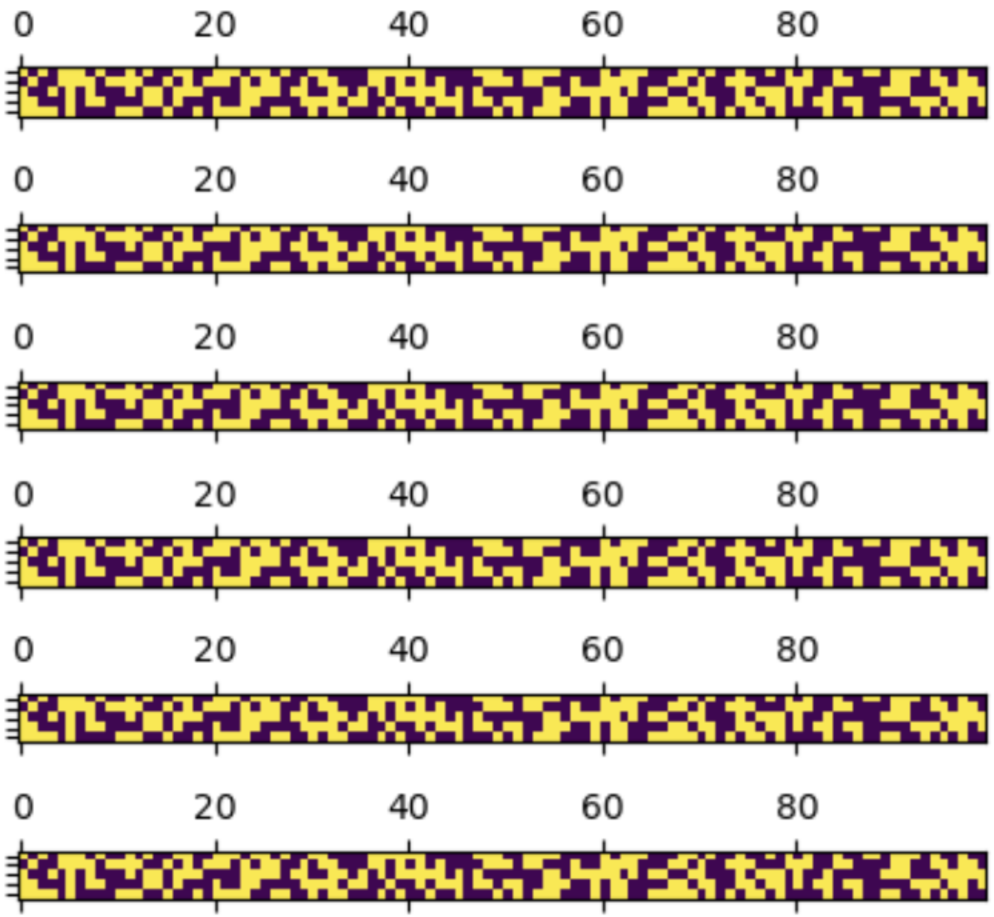
Prior or External Data can help avoiding Model Collapse

$$p_{st}(\boldsymbol{\theta}) = c\,p_0^{2u}(\boldsymbol{\theta})e^{-2m\mathrm{D}_{\mathrm{KL}}(q|\boldsymbol{\theta})}$$

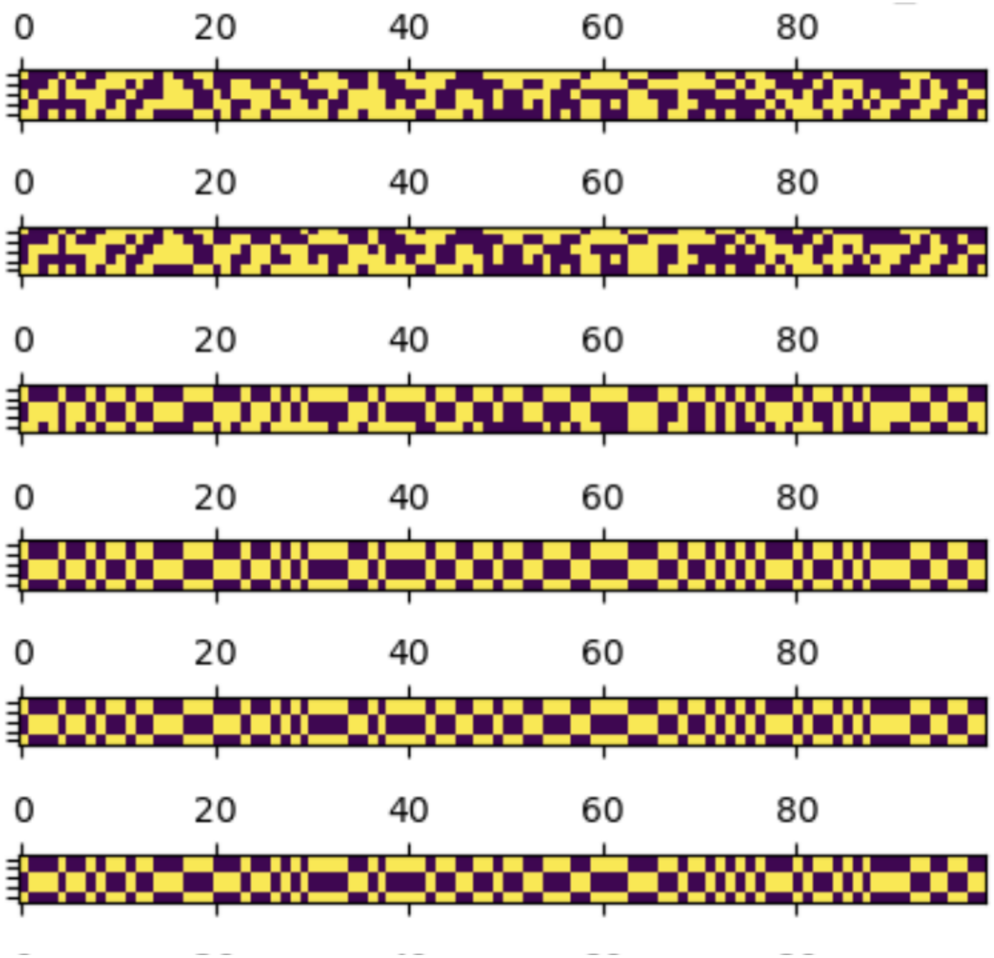


Closed Loop Learning in Hopfield Model

6 stored patterns



10 stored patterns



Summary

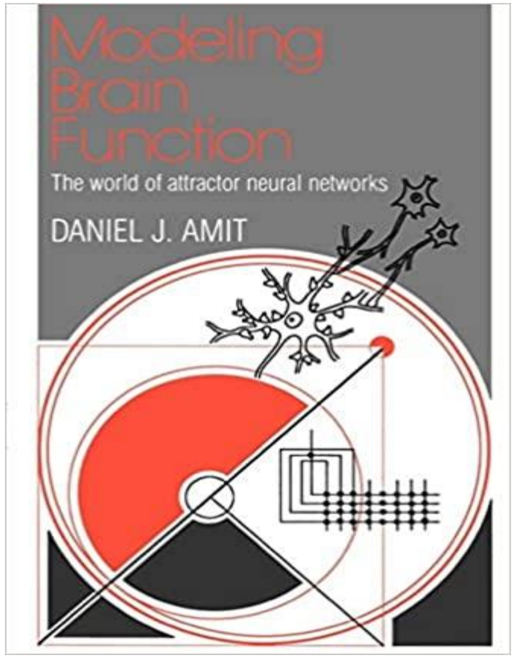
- Hopfield model, RBMs etc, are spherical cow models of the brain, useful but incomplete.

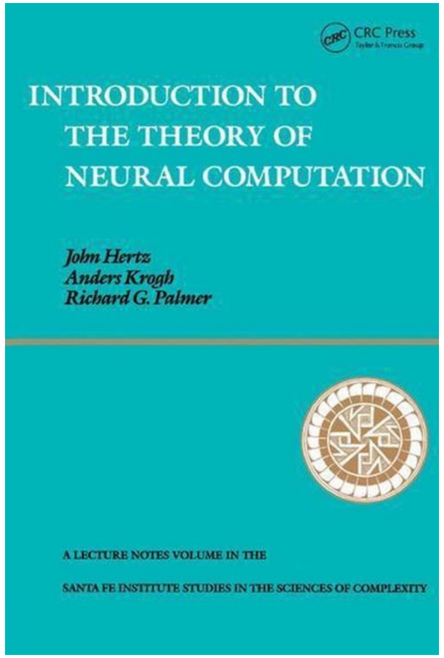
- They can be made more complete in at least in two ways:

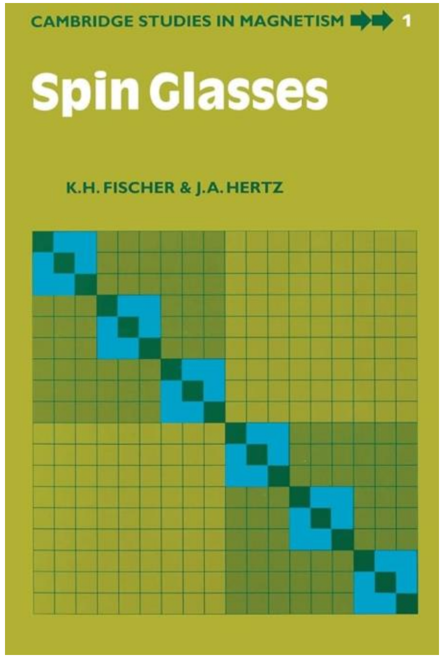
1. Making them less spherical: adding more internal biological realism
2. Taking them out of vacuum: interact with a world that itself is influenced by the machine

Which way is more fruitful for building models of intelligence?

Thank you!







F Schönsberg, Y Roudi, A Treves, PRL 2021
G di Sarra, B Bravi, Y Roudi, EPL, 2025
F Jangjoo, M Marsili, Y Roudi, arXiv preprint arXiv:2506.20623