The Cortex in the Code: How Neuroscience Makes AI Intelligent

A unique investigation into the theory behind machine learning, using the analogy of the human brain.

Dalton A R Sakthivadivel

Disclosures

- I have no relevant financial relationships with commercial interests to disclose.

Contents and aim

Table of contents:

- Conclusion: understanding how constructs from neuroscience allow machine learning to learn

Neuro 101

- The brain is made of neurones

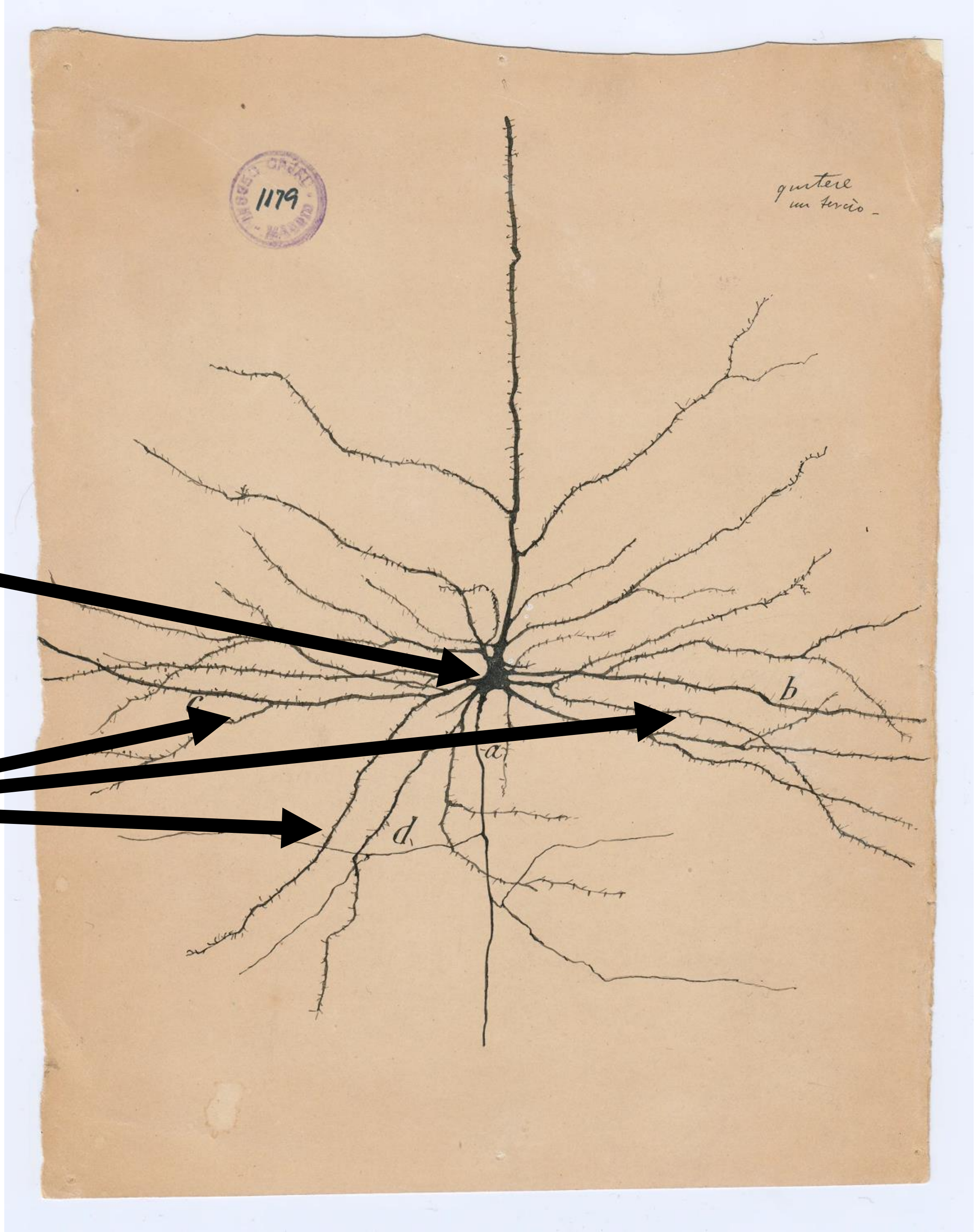
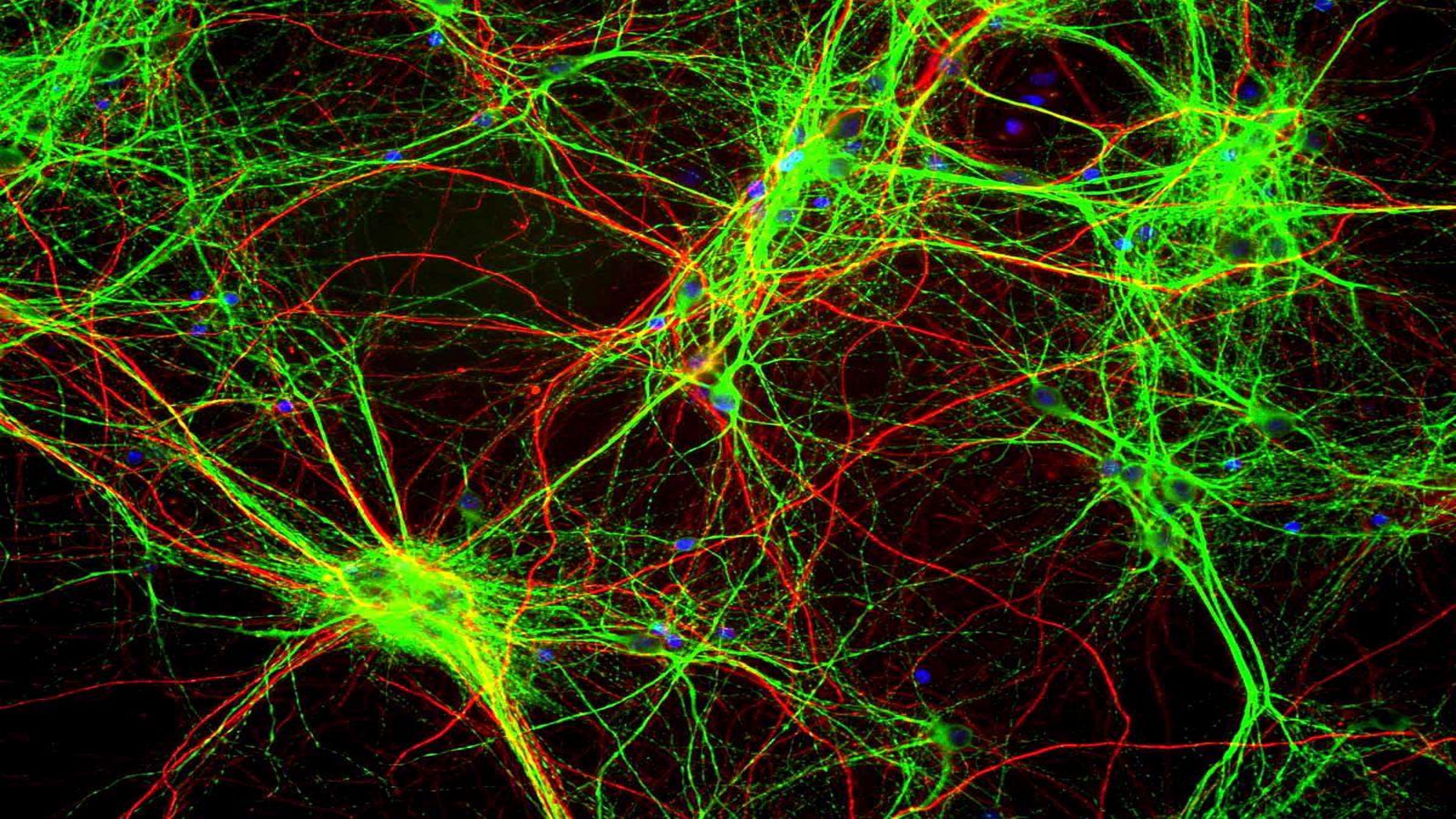


Figure 1. Santiago Ramón y Cajal, The pyramidal neuron of the cerebral cortex, 1904 Ink and pencil on paper, 8 5/8 x 6 7/8 in.

Figure 2 (next). Networks of neurones and dendritic arborisations. Credit: https://commons.wikimedia.org/wiki/File:Culture_of_rat_brain_cells_stained_with_antibody_to_MAP2_(green),_Neurofilament_(red)_and_DNA_(blue).jpg



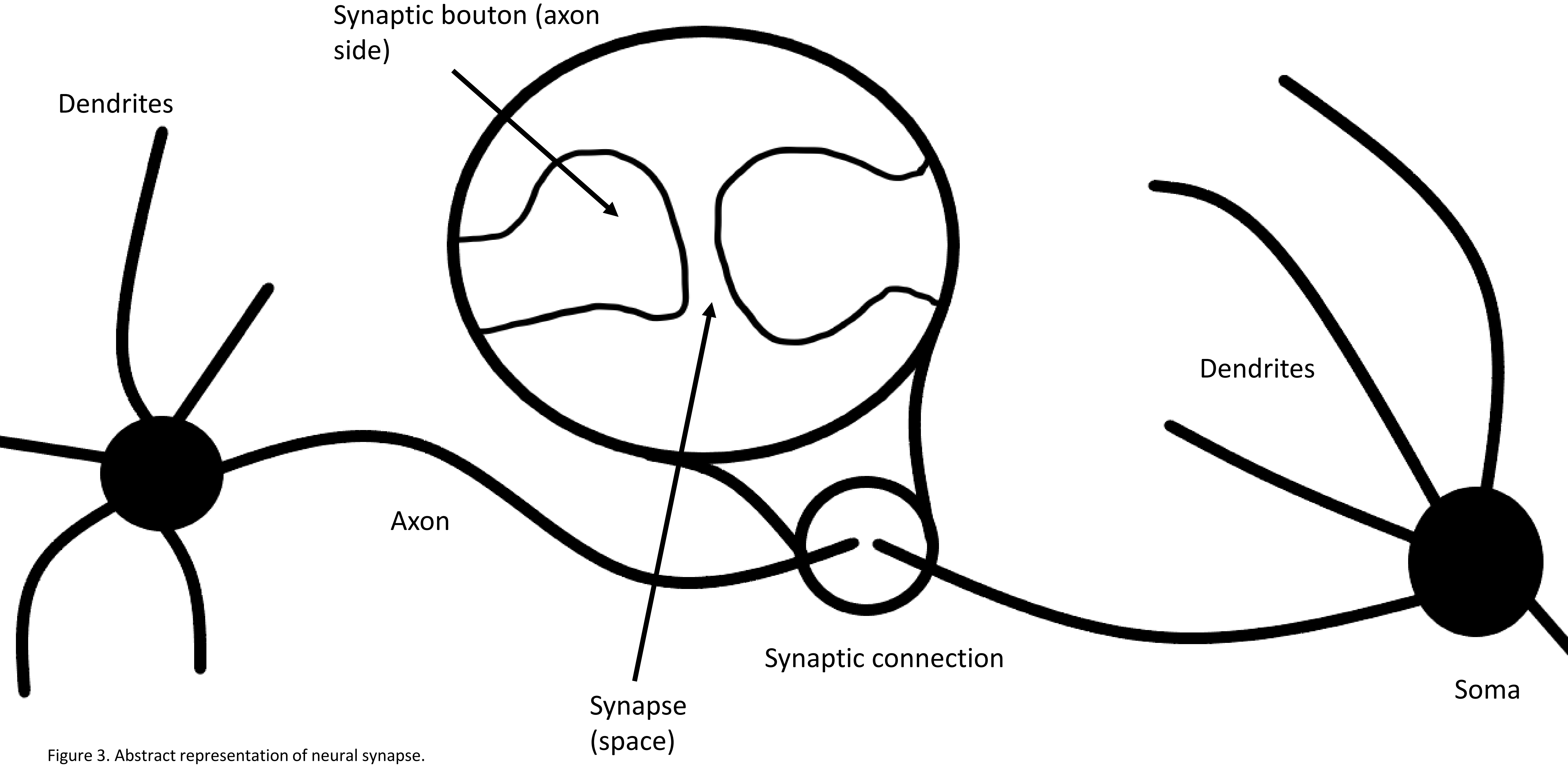


Figure 3. Abstract representation of neural synapse.

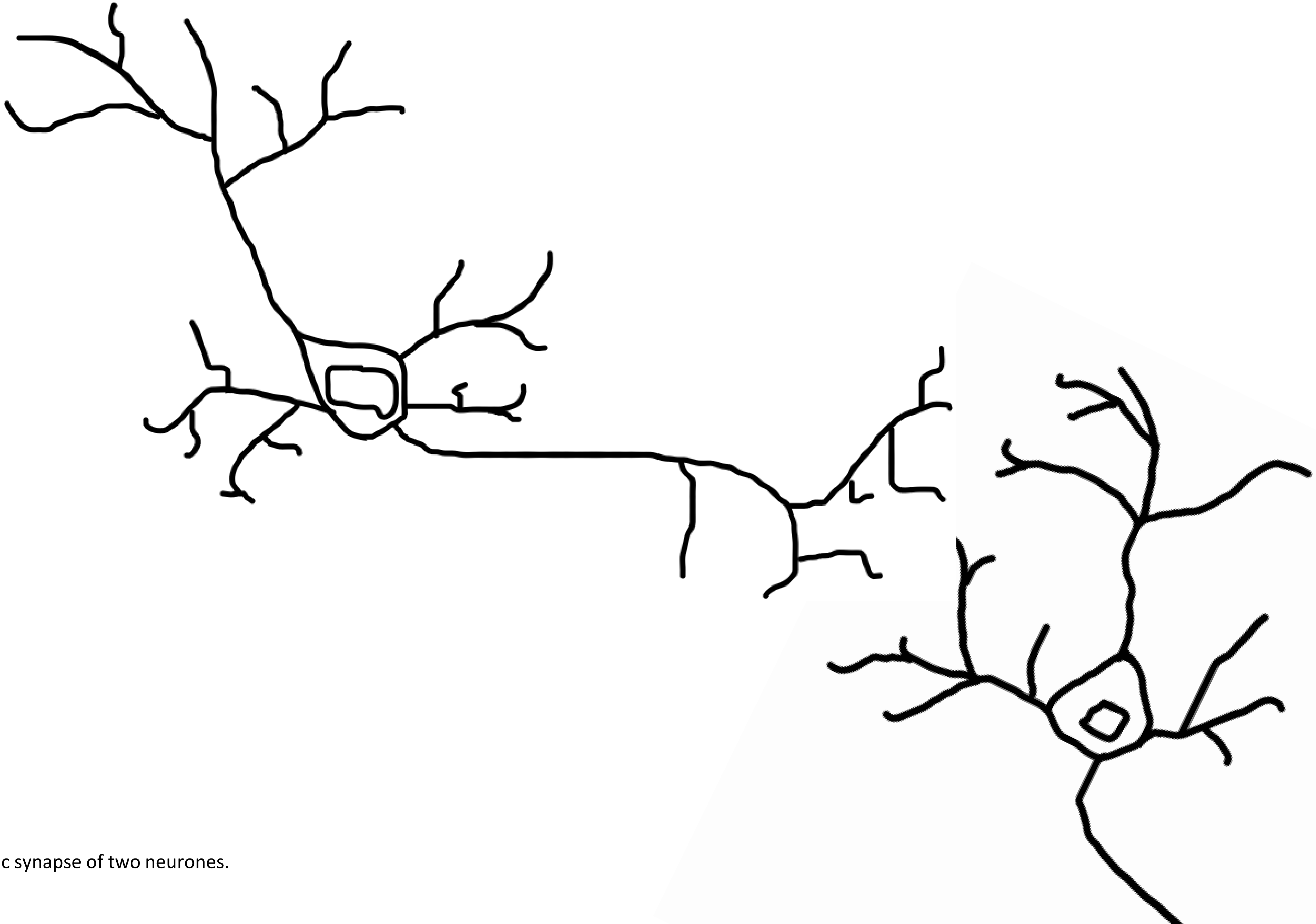


Figure 4. Axon - dendritic synapse of two neurones.

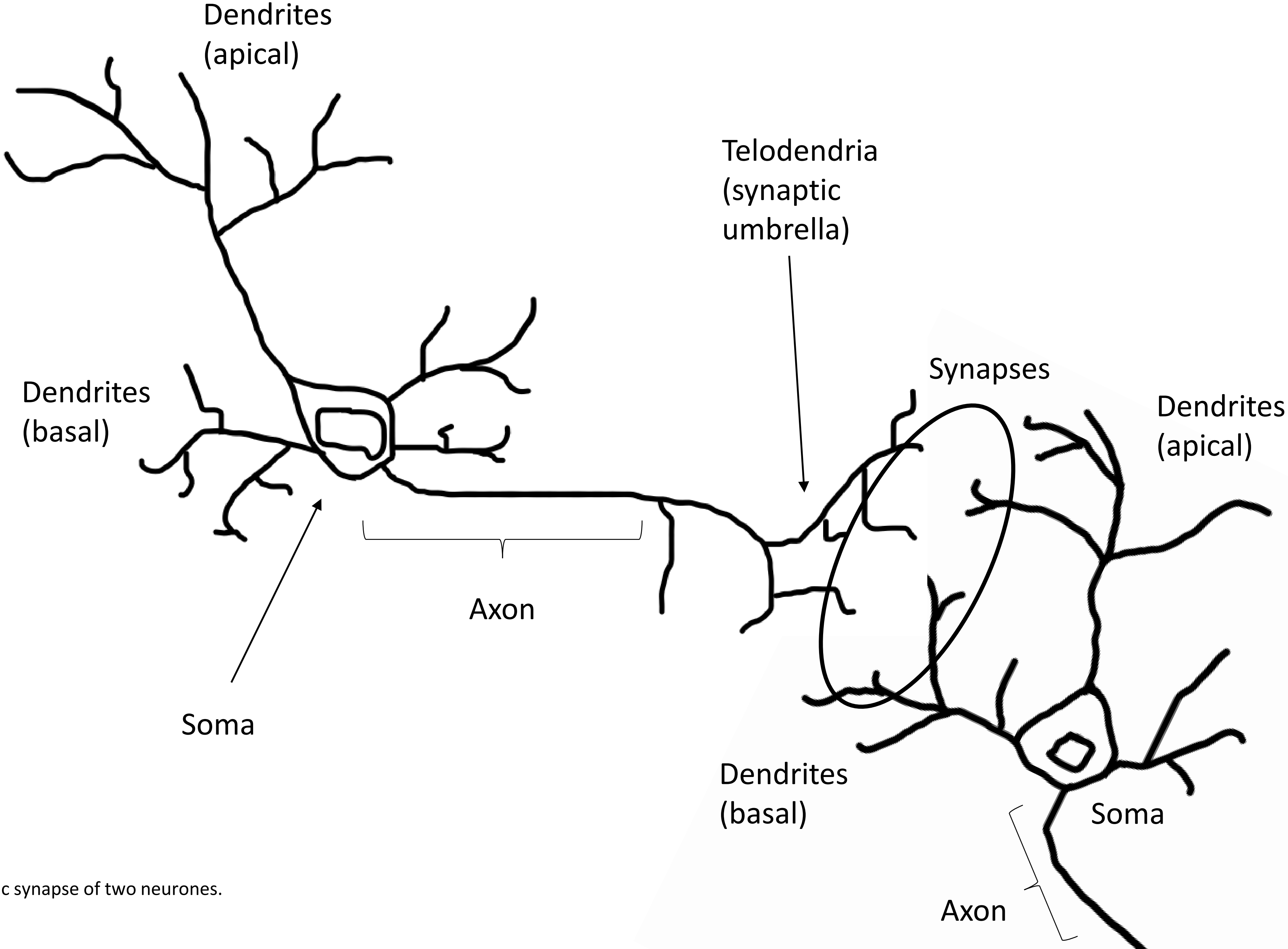


Figure 4. Axon - dendritic synapse of two neurones.

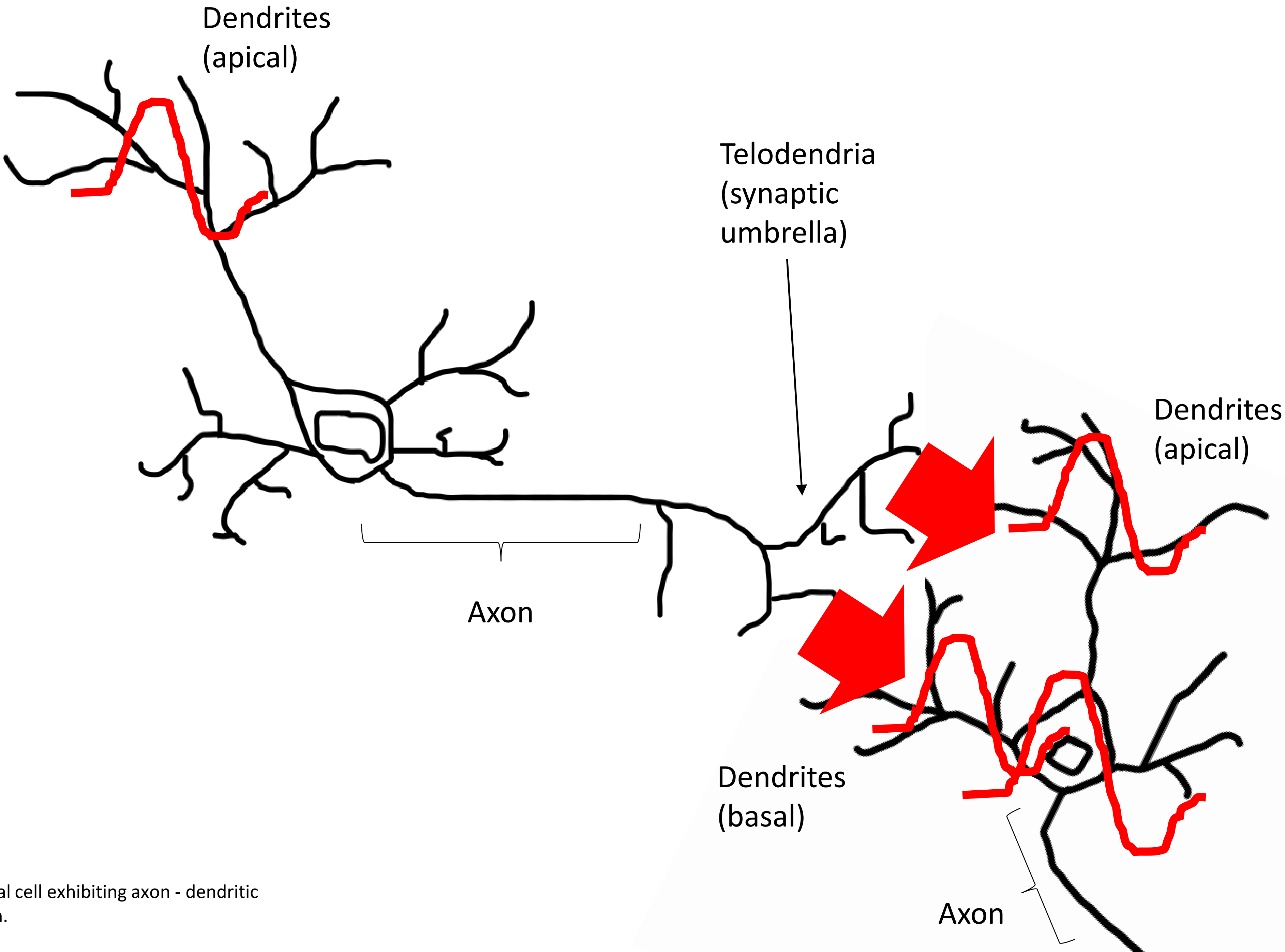
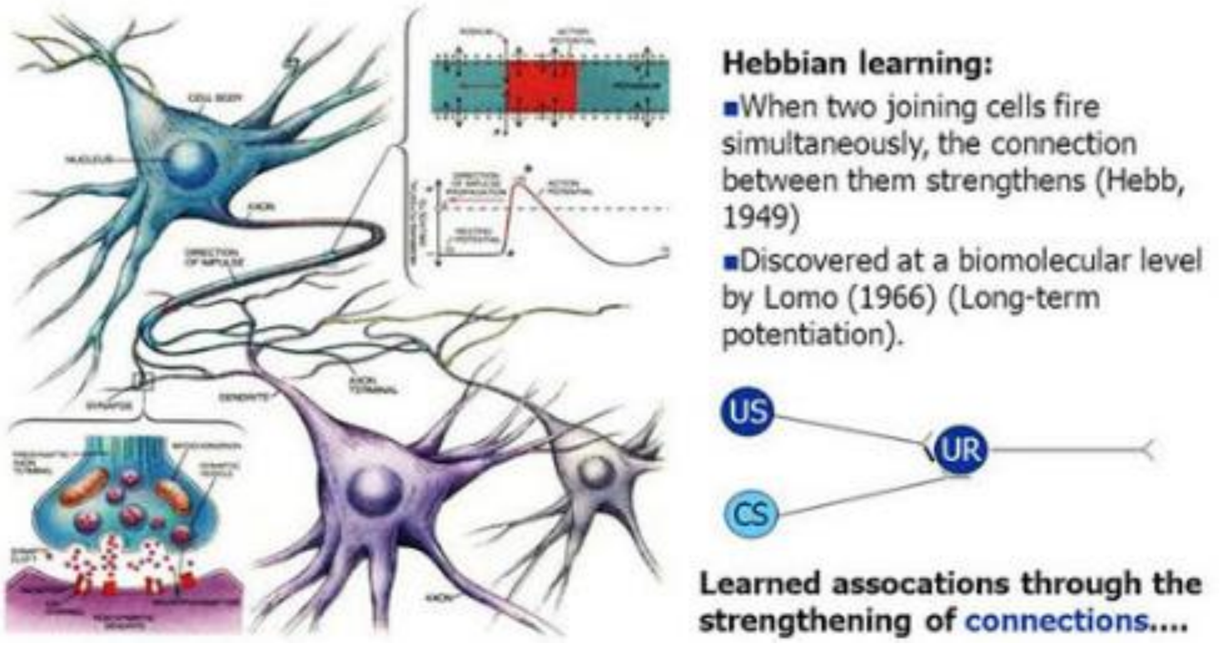


Figure 5. Firing pyramidal cell exhibiting axon - dendritic synaptic communication.

Neural synapses: Hebbian learning

- “When an axon of cell A is near enough to excite a cell B… A’s efficiency, as one of the cells firing B, is increased,” Hebb, 1949.
- Cells that fire together wire together
- Cells that communicate often become ‘stronger,’ those that don’t weaken



Hebbian learning:

- When two joining cells fire simultaneously, the connection between them strengthens (Hebb, 1949)
- Discovered at a biomolecular level by Lomo (1966) (Long-term potentiation).

Learned assocations through the strengthening of connections....

Figure 6. Hebbian synaptic adjustment. Credit: https://www.datasciencecentral.com/profiles/blogs/learning-rules-in-neural-network

Neural synapses: Hebbian learning

$$y(\mathbf{x}) = \sum_{i=1}^{l} x_i \cdot w_i \qquad \text{Eq. 1}$$

Output y depends on…

sum of all inputs * all weights

Neural synapses: Hebbian learning

Weight change:
Oja’s rule

$$w[n+1] = \boxed{w[n]} + \Delta w$$ Eq. 2

Next
weight

$$w_{ij}[n+1] = \boxed{w_{ij}} + \overbrace{\underline{\eta(x_i - w_{ij} \cdot y_j)}}$$ Eq. 3

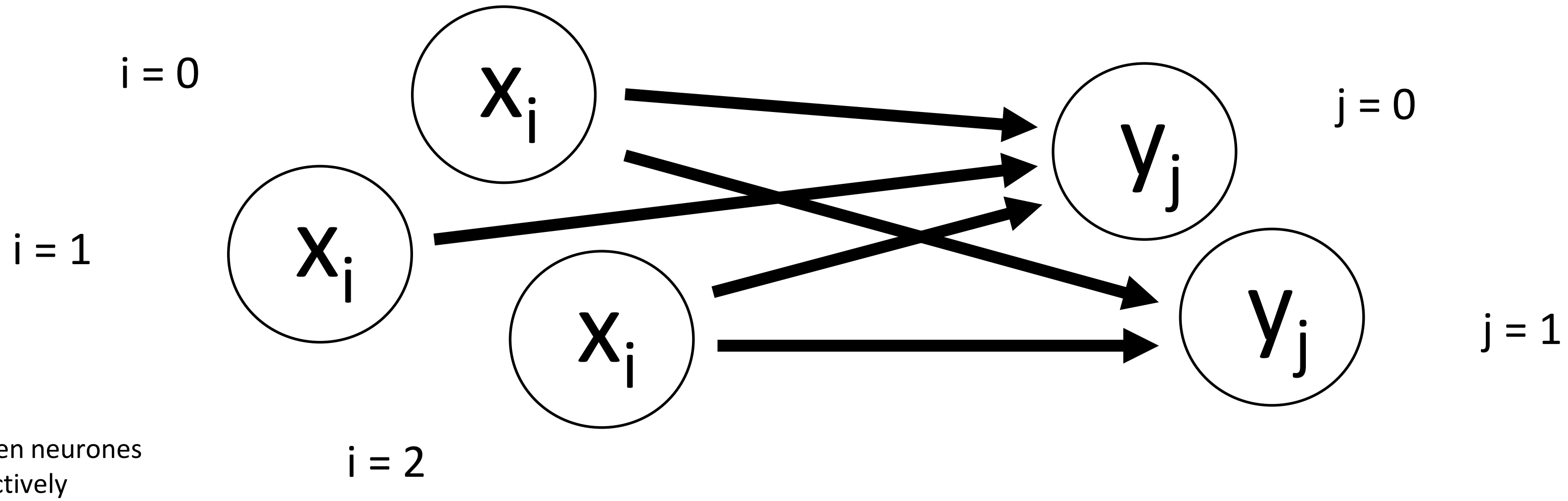
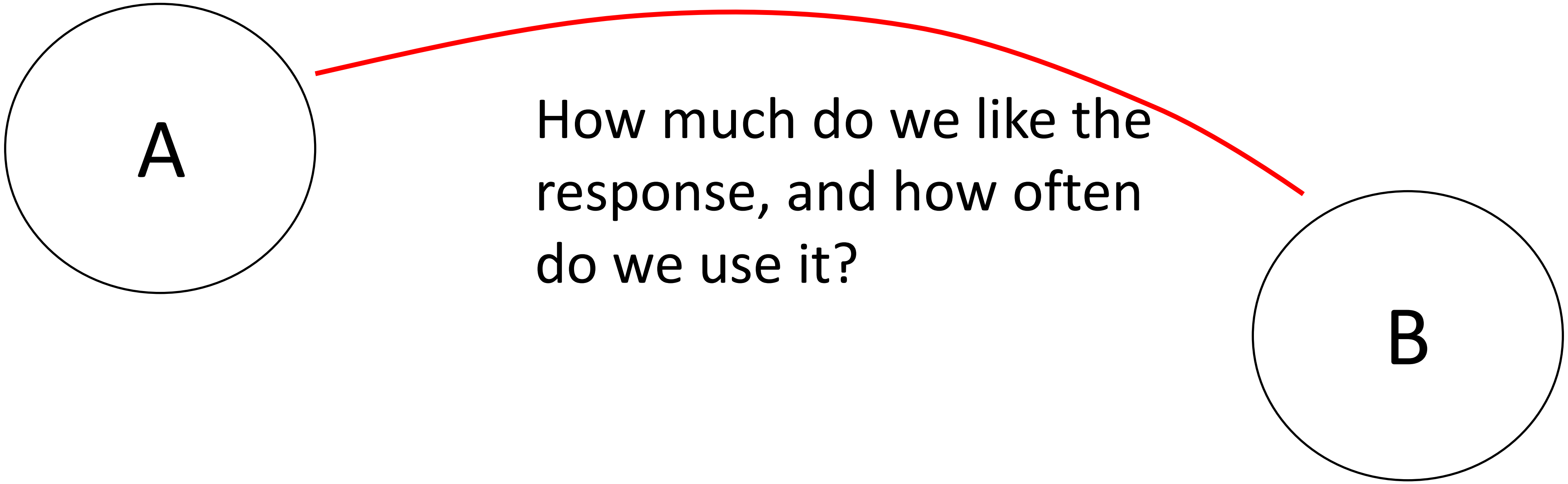


Figure 7. Directed connections between neurones x and y, sets indexed by i and j, respectively

Neural synapses: Hebbian learning

- How do we decide to increase weight?



How much do we like the response, and how often do we use it?

Figure 8. Neural activation pathway representing a learned response to a stimulus.

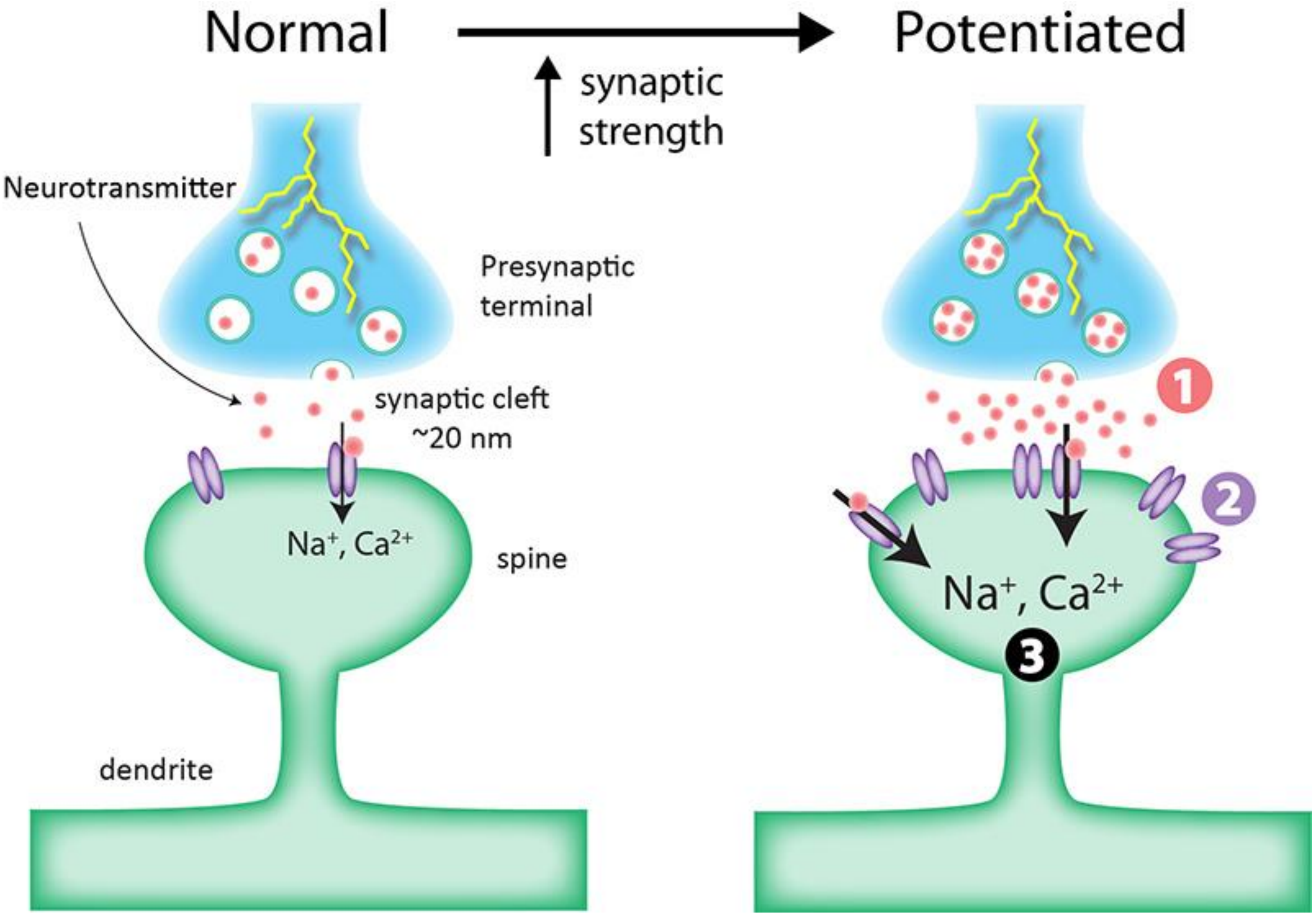


Figure 9. Long - term potentiation. Credit: https://qbi.uq.edu.au/brain-basics/brain/brain-physiology/long-term-synaptic-plasticity

Neural networks

- What is an artificial neural network?

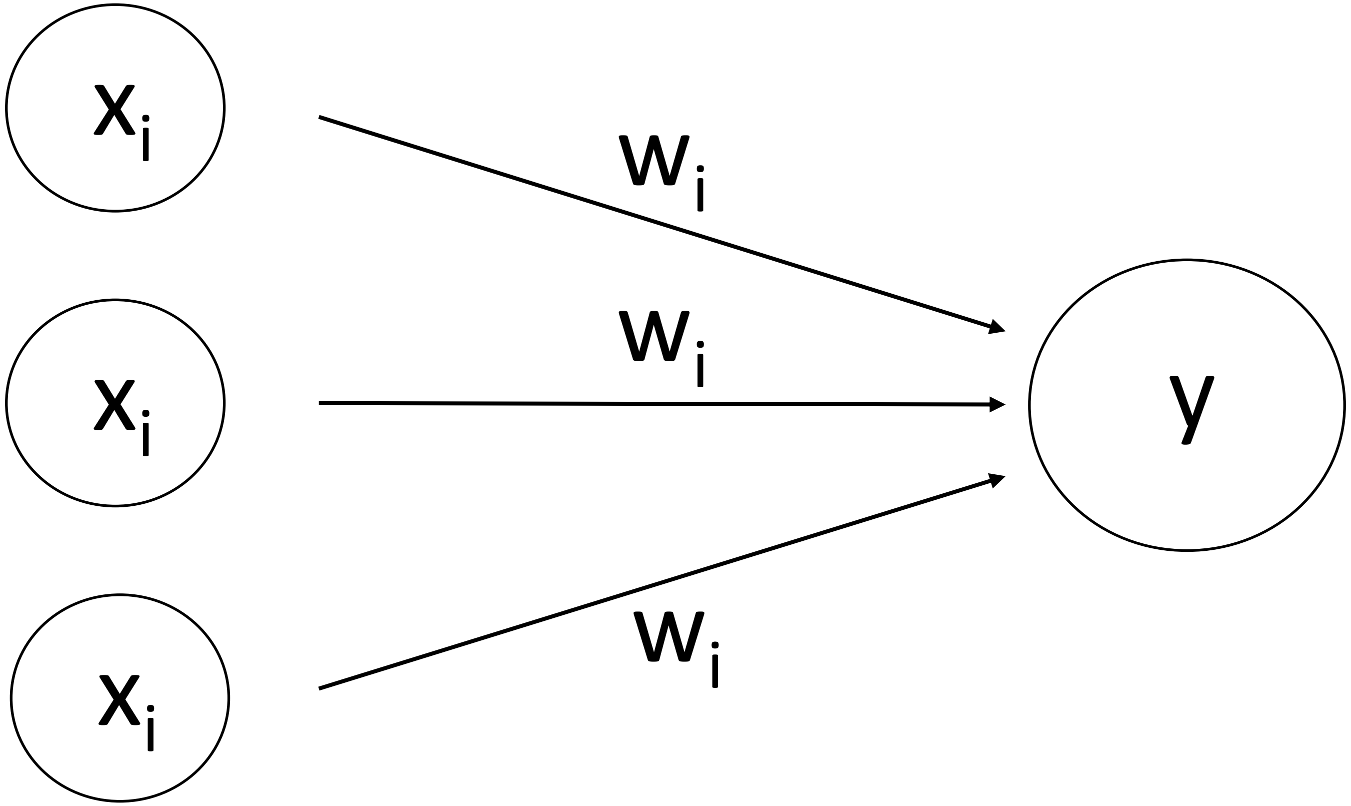


Figure 10. Schematic diagramme of an artificial neural network, perceptron model.

Neural networks

$$y(\mathbf{x}) = \sum_{i=1}^{l} x_i \cdot w_i$$

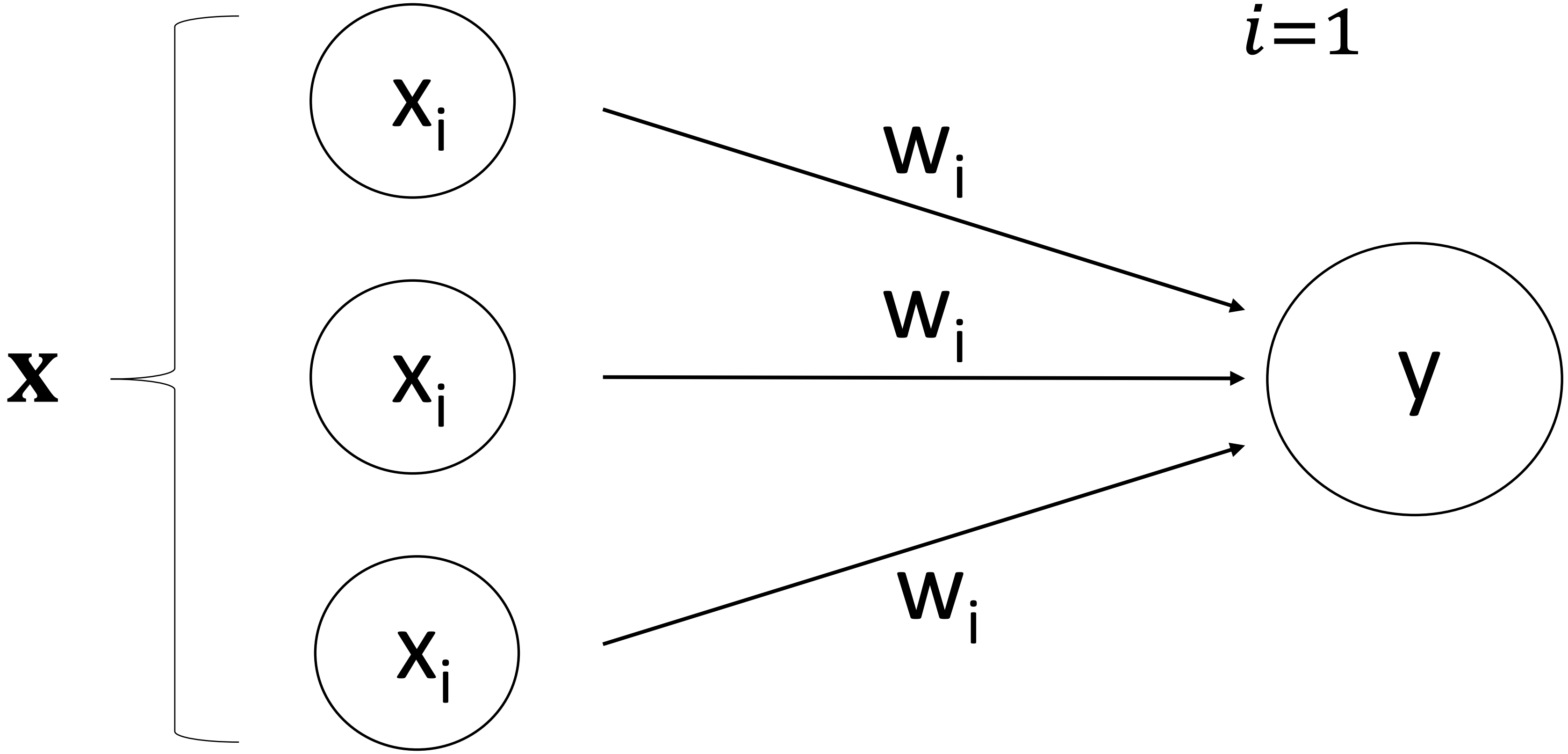
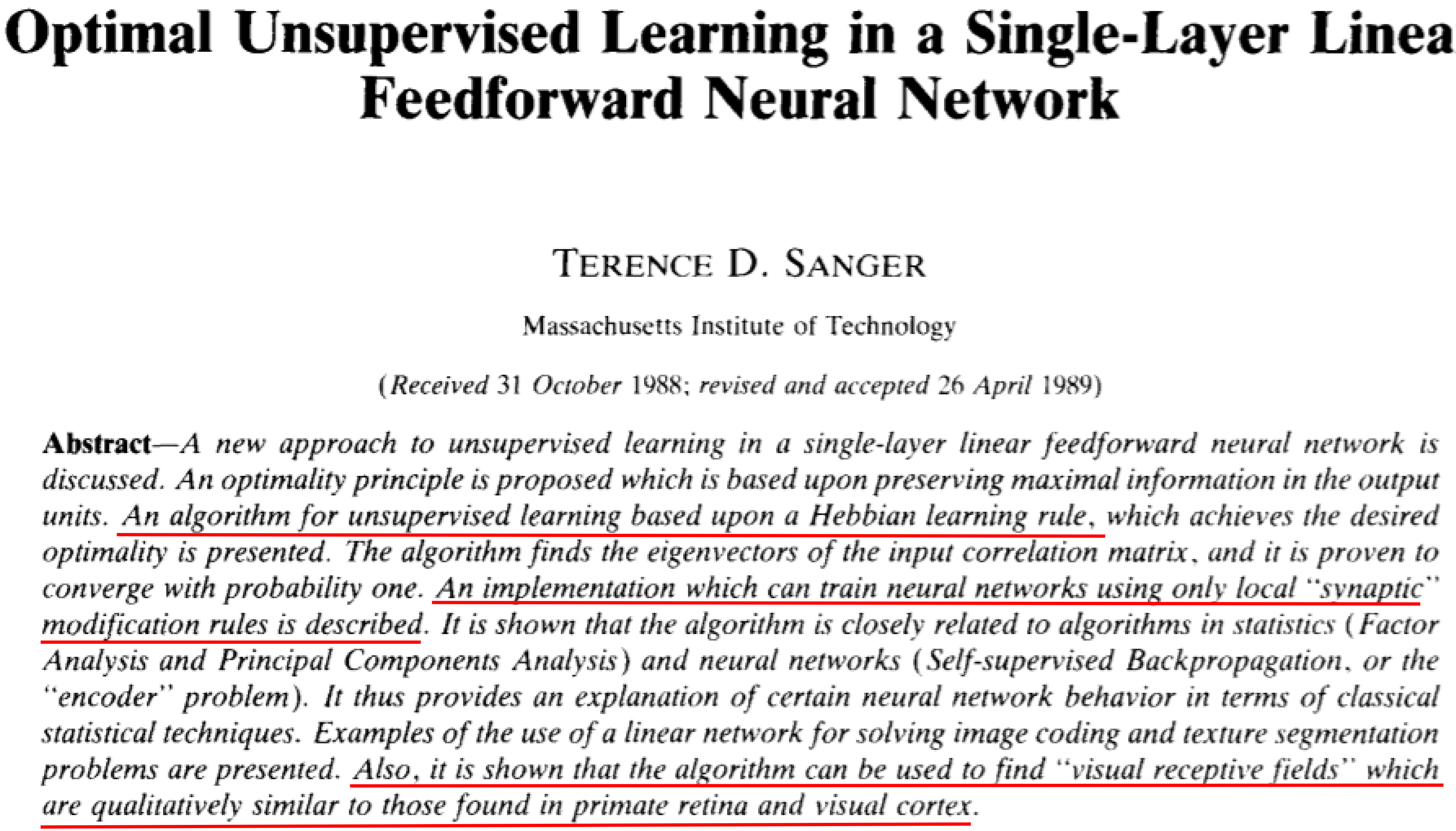


Figure 10. Schematic diagramme of an artificial neural network, perceptron model.

Neural networks: Hebbian learning

- Neural networks also adjust their weights to learn something
- Plays an essential role in training an algorithm
- Iteratively find a weight vector $\mathbf{w}$ for which learning is most accurate



Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network

TERENCE D. SANGER

Massachusetts Institute of Technology

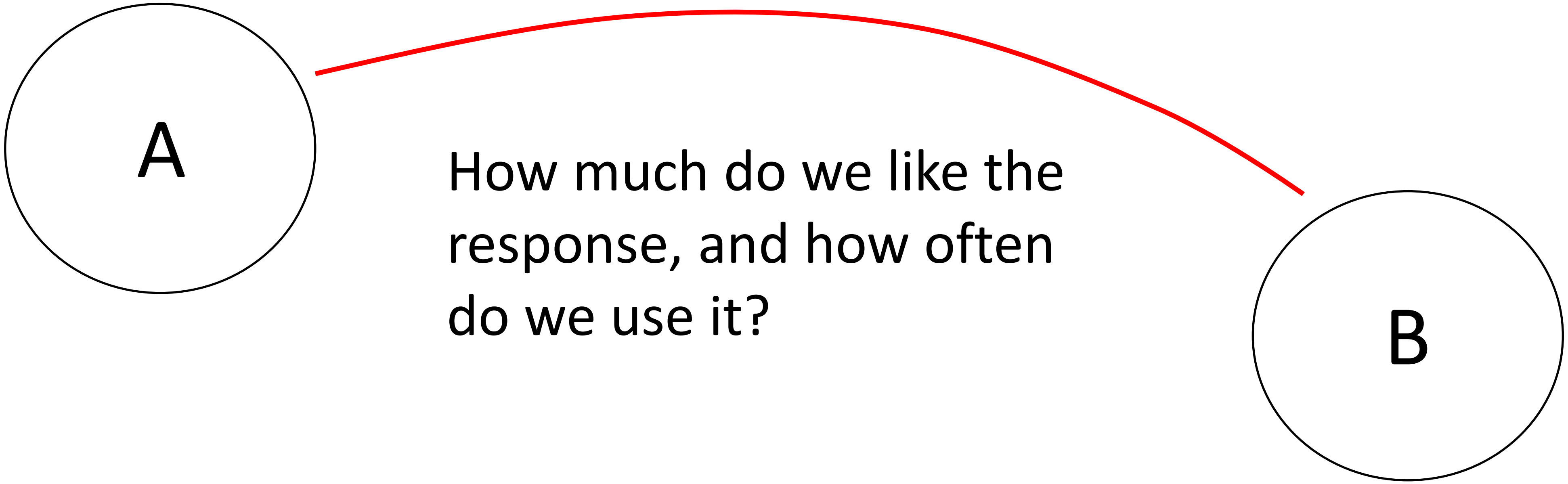
(*Received 31 October 1988; revised and accepted 26 April 1989*)

Abstract—*A new approach to unsupervised learning in a single-layer linear feedforward neural network is discussed. An optimality principle is proposed which is based upon preserving maximal information in the output units. An algorithm for unsupervised learning based upon a Hebbian learning rule, which achieves the desired optimality is presented. The algorithm finds the eigenvectors of the input correlation matrix, and it is proven to converge with probability one. An implementation which can train neural networks using only local "synaptic" modification rules is described. It is shown that the algorithm is closely related to algorithms in statistics (Factor Analysis and Principal Components Analysis) and neural networks (Self-supervised Backpropagation, or the "encoder" problem). It thus provides an explanation of certain neural network behavior in terms of classical statistical techniques. Examples of the use of a linear network for solving image coding and texture segmentation problems are presented. Also, it is shown that the algorithm can be used to find "visual receptive fields" which are qualitatively similar to those found in primate retina and visual cortex.*

Sanger's rule for Hebbian adjustment –
the Generalised Hebbian Algorithm

Neural synapses: Hebbian learning

- How do we decide on a weight?



How much do we like the response, and how often do we use it?

Figure 8. Neural activation pathway representing a learned response to a stimulus.

Energetics

- The study of energy at multiple scales and in different states

- Includes thermodynamics (statistical mechanics) and biochemistry

- Foundation of the physical world

Thermodynamics, statistics, and chemistry: Helmholtz free energy

- Helmholtz drew an equivalence between entropy and a concept called surprise, or free energy

- Distinct from free energy in the brain but shares a number of similarities

- If the brain system is at equilibrium then internal states minimise Helmholtz free energy

- Helmholtz also studied perception, but never united ideas due to differences in application

Thermodynamics, statistics, and the brain: Friston free energy

Karl Friston unites the study of energy with information perception in 2006, applying physics and statistics to psychology

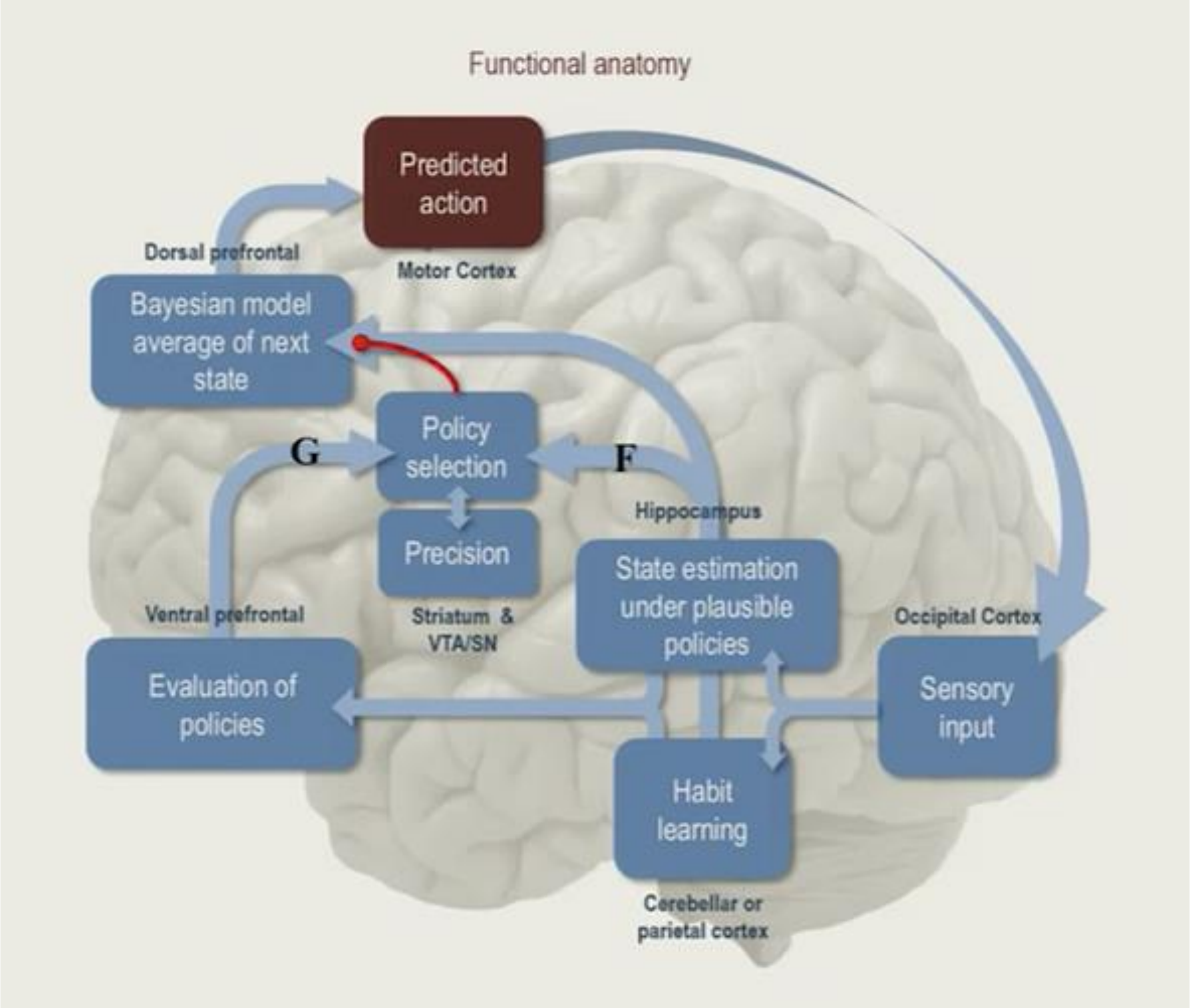
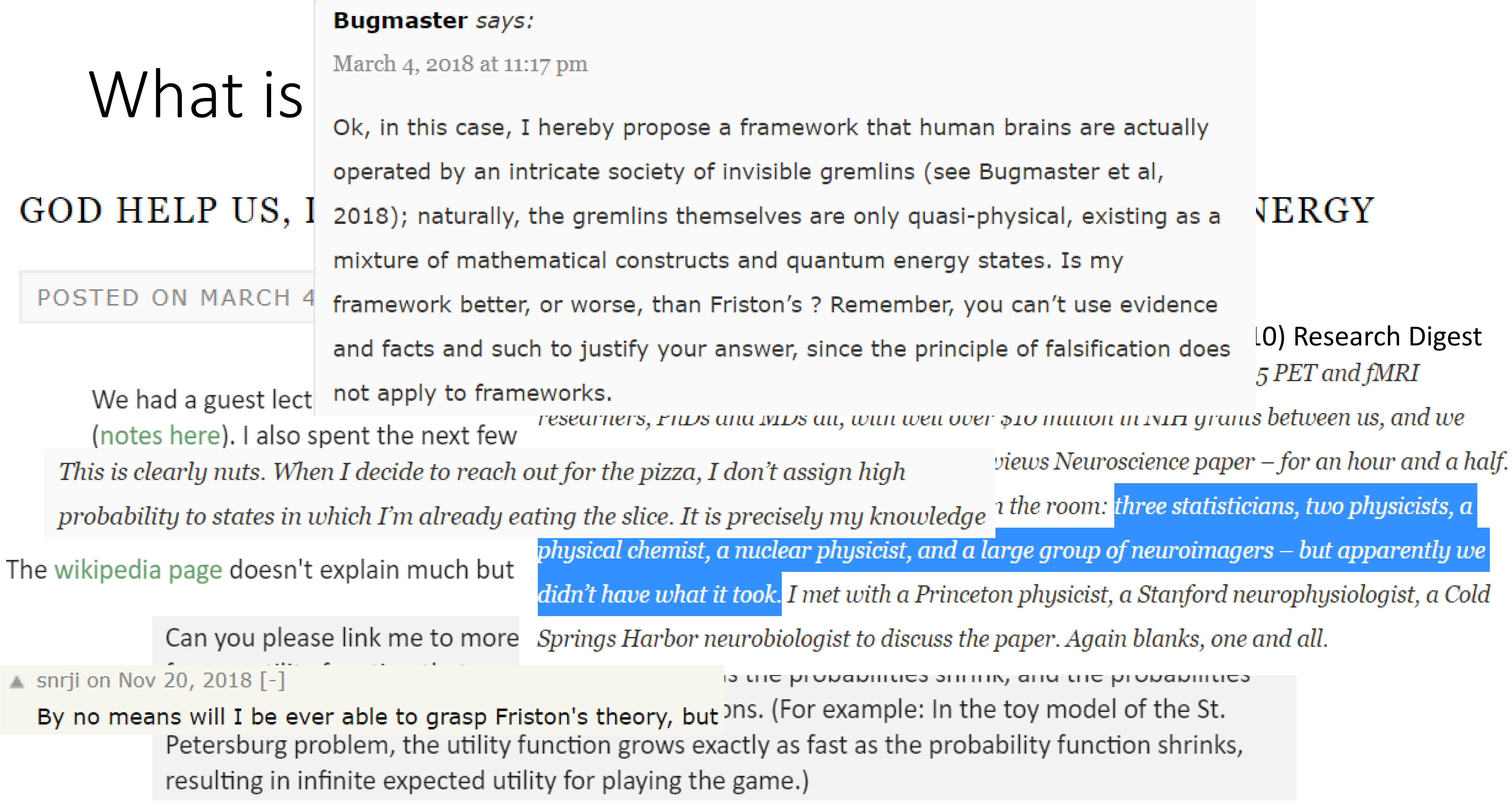


Figure 11. Free energy schematic demonstrating the calculation of free energy in the brain. Credit: Karl Friston - 2016 CCN Workshop: Predictive Coding

What is



Bugmaster *says:*

March 4, 2018 at 11:17 pm

Ok, in this case, I hereby propose a framework that human brains are actually operated by an intricate society of invisible gremlins (see Bugmaster et al, 2018); naturally, the gremlins themselves are only quasi-physical, existing as a mixture of mathematical constructs and quantum energy states. Is my framework better, or worse, than Friston's ? Remember, you can't use evidence and facts and such to justify your answer, since the principle of falsification does not apply to frameworks.

GOD HELP US, I

NERGY

POSTED ON MARCH 4

We had a guest lect

(notes here). I also spent the next few

10) Research Digest

5 PET and fMRI researchers, PhDs and MDs all, with well over $10 million in NIH grants between us, and we

views Neuroscience paper – for an hour and a half.

n the room: three statisticians, two physicists, a physical chemist, a nuclear physicist, and a large group of neuroimagers – but apparently we didn't have what it took. I met with a Princeton physicist, a Stanford neurophysiologist, a Cold Springs Harbor neurobiologist to discuss the paper. Again blanks, one and all.

This is clearly nuts. When I decide to reach out for the pizza, I don't assign high probability to states in which I'm already eating the slice. It is precisely my knowledge

The wikipedia page doesn't explain much but

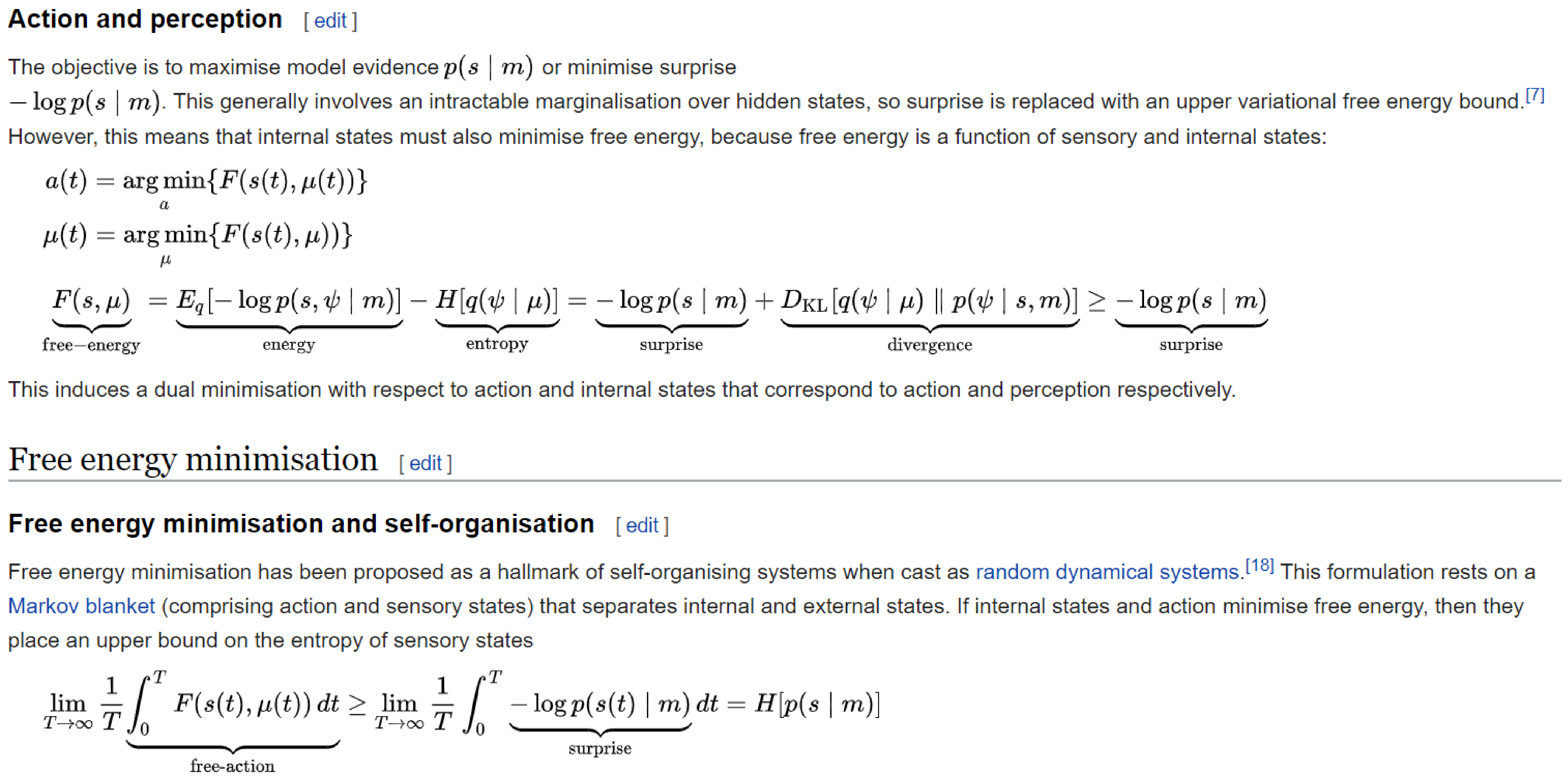
Can you please link me to more

s the probabilities shrink, and the probabilities

ons. (For example: In the toy model of the St. Petersburg problem, the utility function grows exactly as fast as the probability function shrinks, resulting in infinite expected utility for playing the game.)

▲ snrji on Nov 20, 2018 [-]

By no means will I be ever able to grasp Friston's theory, but



Action and perception [edit]

The objective is to maximise model evidence $p(s \mid m)$ or minimise surprise $-\log p(s \mid m)$. This generally involves an intractable marginalisation over hidden states, so surprise is replaced with an upper variational free energy bound.[7] However, this means that internal states must also minimise free energy, because free energy is a function of sensory and internal states:

$$a(t) = \underset{a}{\arg\min}\{F(s(t), \mu(t))\}$$

$$\mu(t) = \underset{\mu}{\arg\min}\{F(s(t), \mu))\}$$

$$\underbrace{F(s,\mu)}_{\text{free}-\text{energy}} = \underbrace{E_q[-\log p(s,\psi \mid m)]}_{\text{energy}} - \underbrace{H[q(\psi \mid \mu)]}_{\text{entropy}} = \underbrace{-\log p(s \mid m)}_{\text{surprise}} + \underbrace{D_{\text{KL}}[q(\psi \mid \mu) \parallel p(\psi \mid s,m)]}_{\text{divergence}} \geq \underbrace{-\log p(s \mid m)}_{\text{surprise}}$$

This induces a dual minimisation with respect to action and internal states that correspond to action and perception respectively.

Free energy minimisation [edit]

Free energy minimisation and self-organisation [edit]

Free energy minimisation has been proposed as a hallmark of self-organising systems when cast as random dynamical systems.[18] This formulation rests on a Markov blanket (comprising action and sensory states) that separates internal and external states. If internal states and action minimise free energy, then they place an upper bound on the entropy of sensory states

$$\lim_{T\to\infty} \frac{1}{T} \underbrace{\int_0^T F(s(t), \mu(t))\, dt}_{\text{free-action}} \geq \lim_{T\to\infty} \frac{1}{T} \int_0^T \underbrace{-\log p(s(t) \mid m)}_{\text{surprise}}\, dt = H[p(s \mid m)]$$

Thermodynamics, statistics, and chemistry: Friston free energy

- Why minimise free energy as a measure of model correctness?

- More free energy means more entropy

- Entropy is not good, because the brain tries to stay ordered, and have as little false information as possible

- “It turns out that the problems of inferring the causes of sensory input (perceptual inference) and learning the relationship between input and cause (perceptual learning) can be resolved using exactly the same principle. Specifically, both inference and learning rest on minimizing the brain’s free energy, as defined in statistical physics.” Friston, 2005

Bayesian statistics for model building (applied to the brain)

Percent chance that our internal model is still correct in light of new observation

$$P(M|O) = \frac{\boxed{P(O|M)} \cdot \boxed{P(M)}}{\boxed{P(O)}} \qquad \text{Eq. 4}$$

Internal model

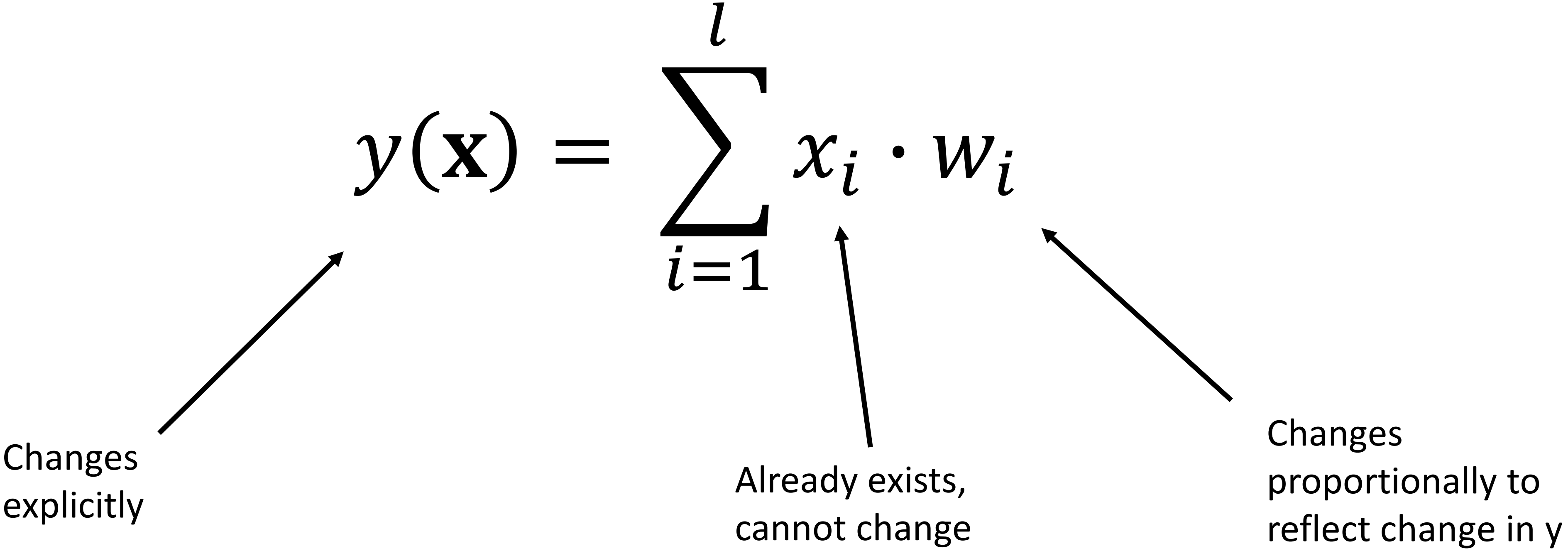
New observation

So what is free energy, really?

- “Predictive coding” under free energy is a model for perception and learning
- Physical, neuronal representations of reality must change to minimise error
- Hebbian synapses are thus subject to free energy

Error minimisation

- Error minimisation at the synapse level, unlike the more global nature of ‘psychological’ learning



$$y(\mathbf{x}) = \sum_{i=1}^{l} x_i \cdot w_i$$

Changes explicitly

Already exists, cannot change

Changes proportionally to reflect change in y

Hebbian free energy: a simplified derivation

- Create a model of the world around you, based on relevant prior experiences
- Model is characterised by actual data v, observation u, expected data v_p, function $g(v)$ for mapping, sensory noise Σ_u, and expected noise Σ_p
- Based on your observation u and your past experience you will try to estimate the true data, v_p and Σ_p

Hebbian free energy: a simplified derivation

Phi, characterises (new) prior model

Error terms

$$\varepsilon_p = \frac{\phi - v_p}{\Sigma_p} \qquad \text{Eq. 5}$$

$$\varepsilon_u = \frac{u - g(\phi)}{\Sigma_u} \qquad \text{Eq. 6}$$

Hebbian free energy: a simplified derivation

$$\varepsilon_p = \frac{\phi - v_p}{\Sigma_p}$$

$$\varepsilon_u = \frac{u - g(\phi)}{\Sigma_u}$$

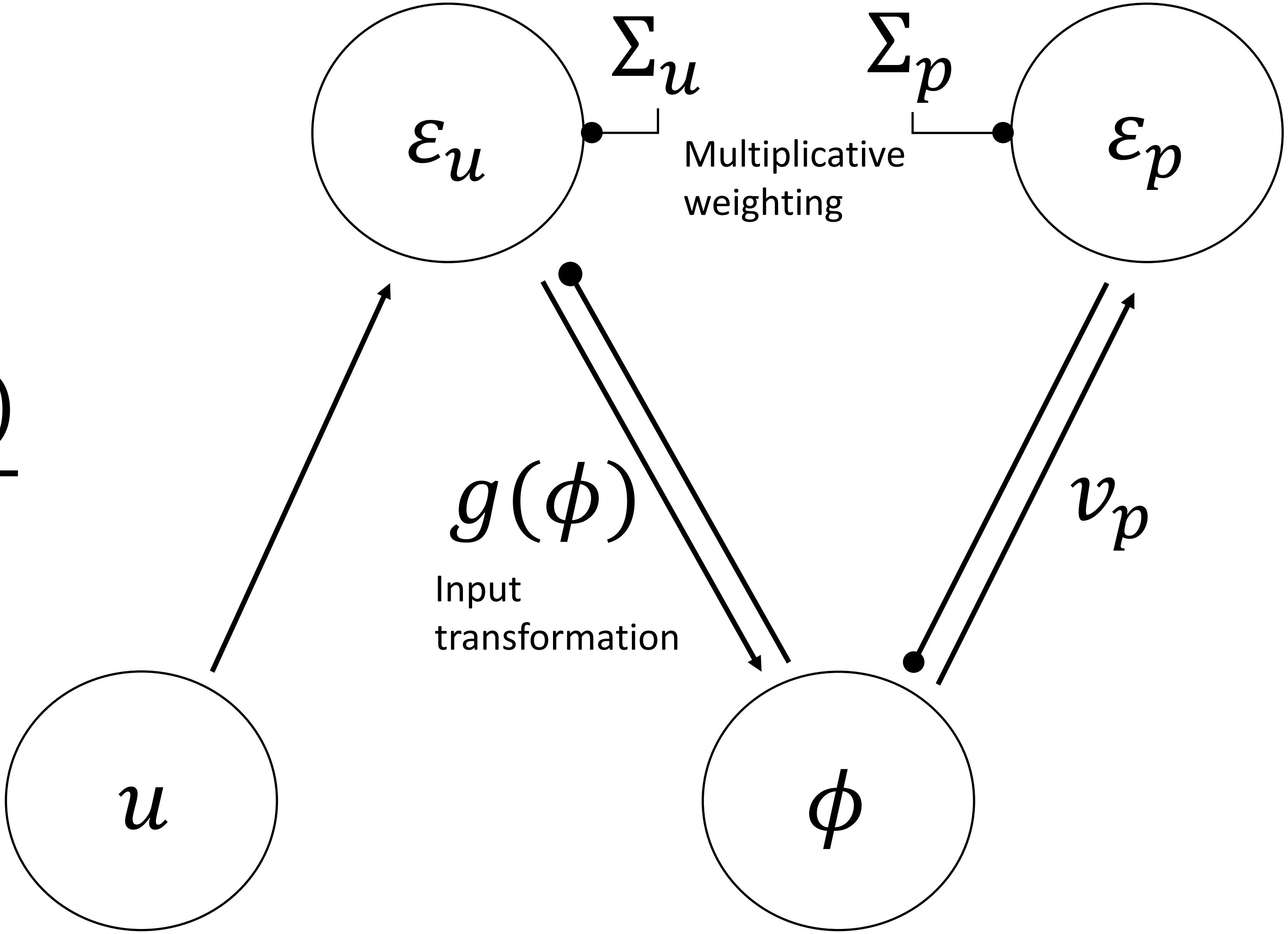


Figure 12. Schematic diagramme of a Hebbian free energy architecture. Includes neural states and weights, or activity and connections.

Hebbian free energy: a simplified derivation

$$\Delta v_p = \varepsilon_p$$ Eq. 7

$$\Delta \Sigma_p = \frac{1}{2}({\varepsilon_p}^2 - {\Sigma_p}^{-1})$$ Eq. 8

$$\Delta \Sigma_u = \frac{1}{2}({\varepsilon_u}^2 - {\Sigma_u}^{-1})$$ Eq. 9

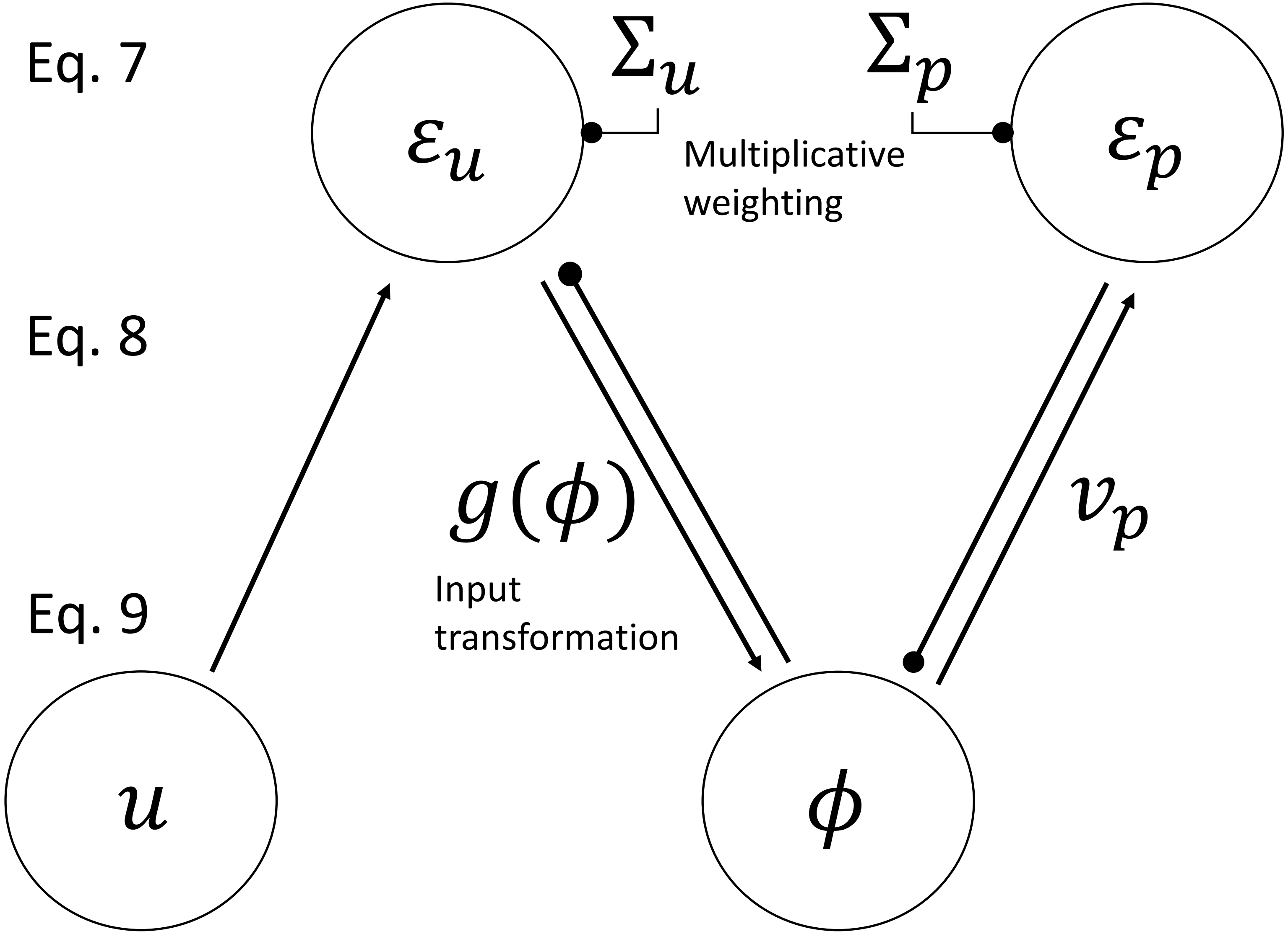
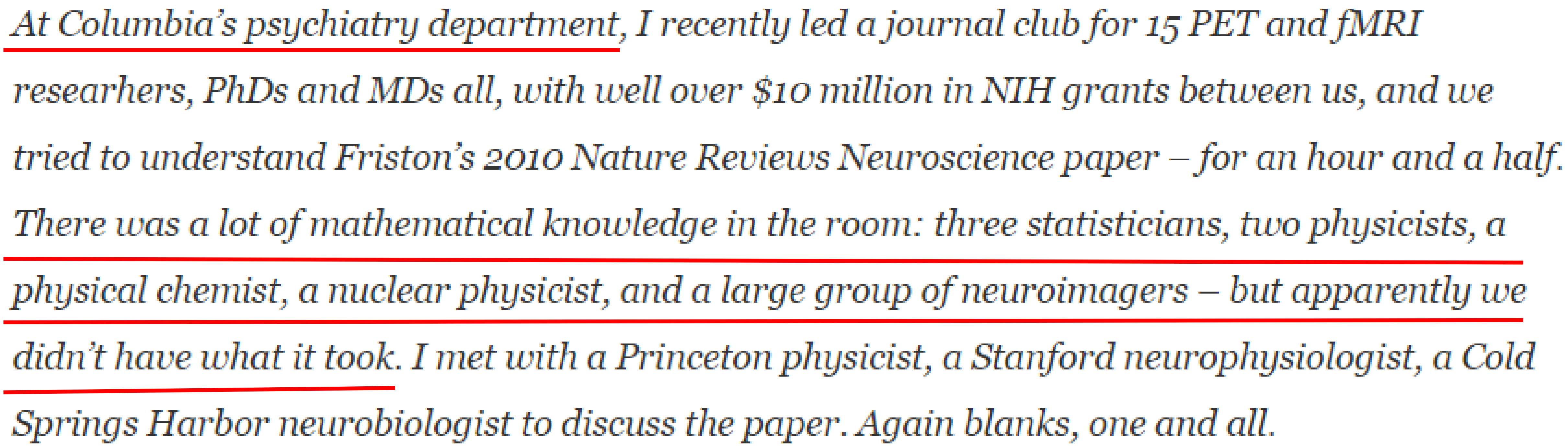


Figure 12. Schematic diagramme of a Hebbian free energy architecture. Includes neural states and weights, or activity and connections.

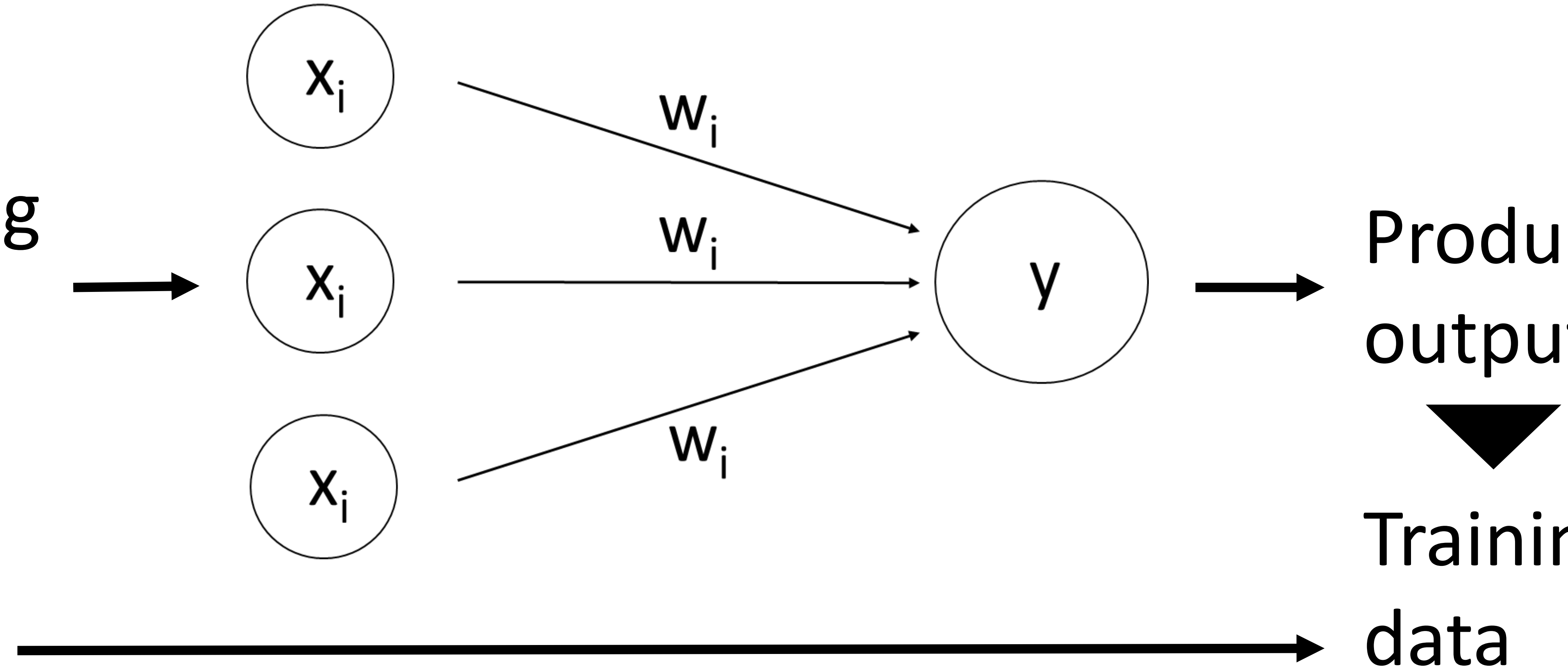
Hebbian free energy: we’re finally done



At Columbia’s psychiatry department, I recently led a journal club for 15 PET and fMRI researhers, PhDs and MDs all, with well over $10 million in NIH grants between us, and we tried to understand Friston’s 2010 Nature Reviews Neuroscience paper – for an hour and a half. There was a lot of mathematical knowledge in the room: three statisticians, two physicists, a physical chemist, a nuclear physicist, and a large group of neuroimagers – but apparently we didn’t have what it took. I met with a Princeton physicist, a Stanford neurophysiologist, a Cold Springs Harbor neurobiologist to discuss the paper. Again blanks, one and all.

Peter Freed (2010) Research Digest

Hebb and efficient algorithm training



Training data input

Produced output

Training data output

Figure 13. Training a neural network through an error minimisation, or comparative, method.

Training and Bayesian algorithm learning

$$\underbrace{f(X) \rightarrow Y}_{\text{f(x) must map given inputs in all previous training data to given outputs}} : \underbrace{P(\, f(x) \mid (x, y)\,) \approx 1}_{\text{f(x) must also represent any new data set, or give an output for all possible inputs from similar distributions}} \qquad \text{Eq. 10}$$

Integrating Hebb, Friston, and Samuel

- Besides explicitly coding Friston’s architectures, some rules already approximate this

Hebb and brain - like error minimisation

$$w[n+1] = w[n] + \Delta w$$

Eq. 3

Delta rule

$$\Delta w_{ij} = \; x_i \cdot \eta \cdot (y_j - f(x_i)_j)$$

Eq. 11

$$f(x) \dashv y \therefore P(\, f(x) \mid (x, y) \,) \ll 1,$$

$$\Delta f(x) : y - f(x) \approx 0$$

Zooming out

- The link between learning in the brain and learning in ML is statistical inference

- Cells and nodes are uniquely capable of computing on data by using inputs and outputs to calibrate themselves

- Synaptic operations and error minimisation give the human brain learning ability, and form the same substrates of learning and inference in machine learning algorithms. ■

For the future

- What other characteristics of the brain can we use to improve ML performance?

- How can inquiry into ML continue to be driven by neuroscience?

- Can neuroscience ever lead to more complex systems?

1. Hebb, 1949. **The organization of behavior: a neuropsychological theory.** Slide 9, quote, Hebbian concepts.

2. Oja, 1982. **Simplified neuron model as a principal component analyser.** Slide 11, eq. 3.

3. Shors and Matzel, 1997. **Long-term potentiation: what's learning got to do with it?** Slide 13, LTP.

4. Freed, 2010. **Research digest.** Slide 22 and 32, quote on free energy.

5. Friston, 2005. **A theory of cortical responses.** Slide 24, quote on learning, free energy in the brain.

6. Bogacz, 2017. **A tutorial on the free-energy framework for modelling perception and learning.** Slide 29, eqs. 5 and 6.

Get in touch!

- dalton.sakthi@gmail.com
- dalton.a.sakthivadivel@stonybrook.edu
- @DaltonSakthi
    - DMs open