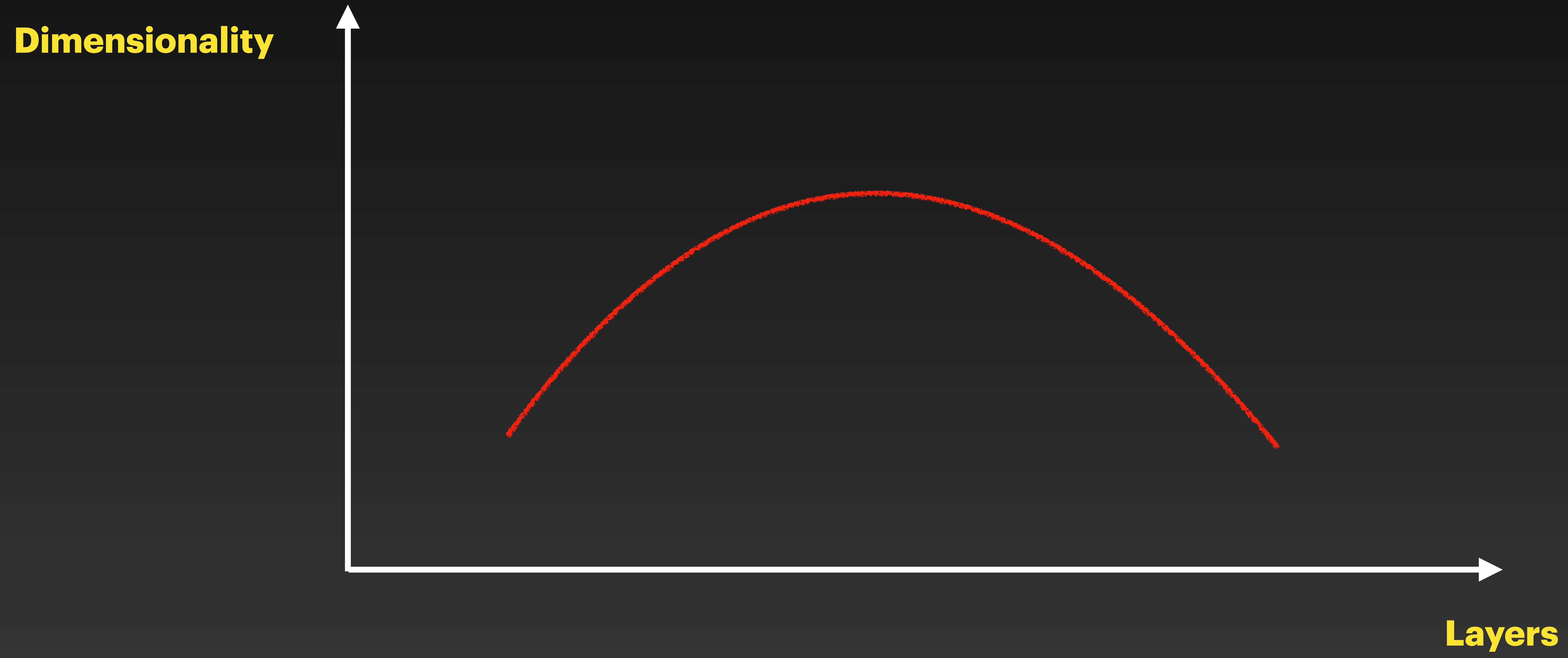
Measuring the dimensionality and disentanglement of energy-based networks with prospective configuration as a learning rule

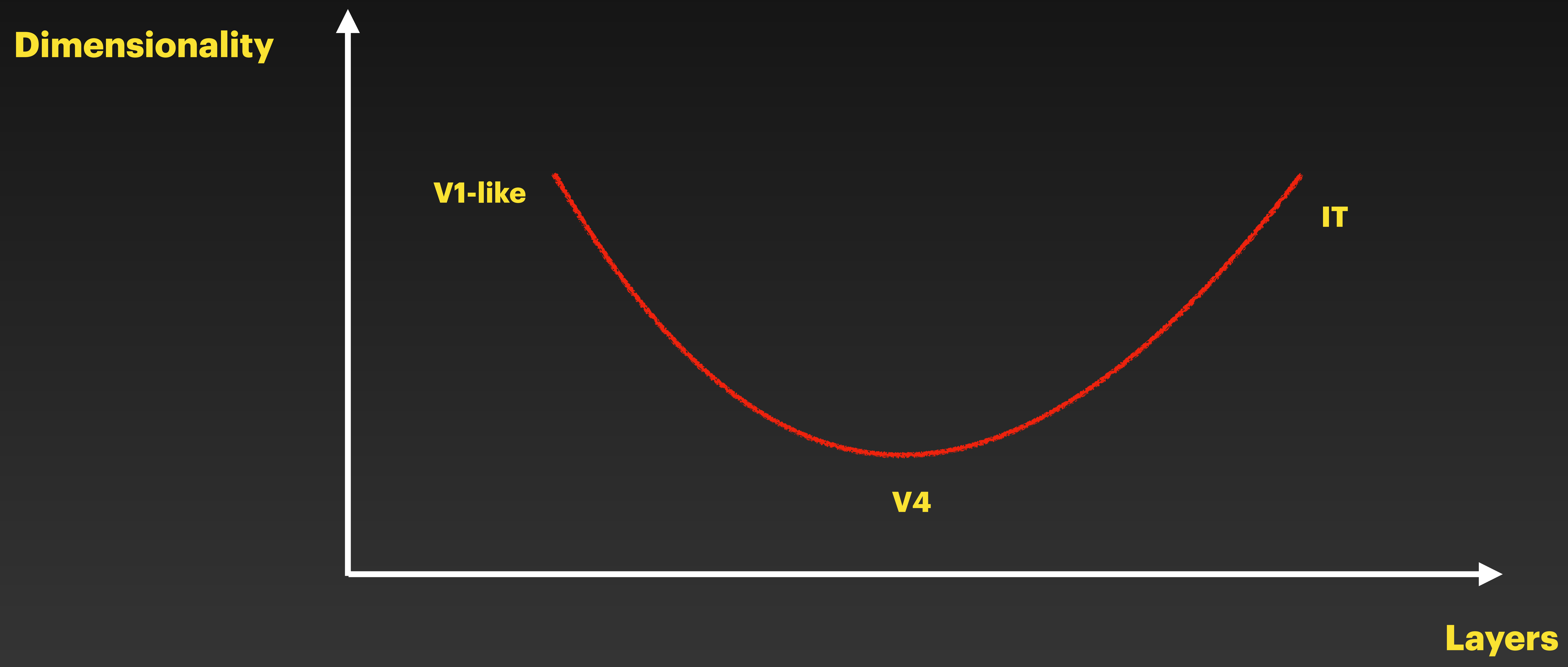
Aslan Satary Dizaji - NeuroAILab - 30.08.2024 - ISRC-CN3-2024

Dimensionality versus deep hierarchy in DNN



Sorscher et al. 2022

Dimensionality versus deep hierarchy in biological brain



Sorscher et al. 2022

Adding biological constraints to DNN to reconcile this discrepancy...

- Adding biological constraints to the loss function of deep neural networks:
  - Non-negativity of activity: $\mathbf{L}_{\mathbf{nonneg}} = \beta_{nonneg} \sum_i \max(-a_i, 0)$
  - Minimising energy efficiency in both activity and weight: $\mathbf{L}_{\mathbf{activity}} = \beta_{activity} \sum_l \|a_l\|^2, \mathbf{L}_{\mathbf{weight}} = \beta_{weight} \sum_l \|W_l\|^2$
  - Total loss function: $\mathbf{L} = \mathbf{L}_{\mathbf{nonneg}} + \mathbf{L}_{\mathbf{activity}} + \mathbf{L}_{\mathbf{weight}} + \mathbf{L}_{\mathbf{prediction}}$

Whittington et al. 2022

NeuroAILab

Measuring dimensionality

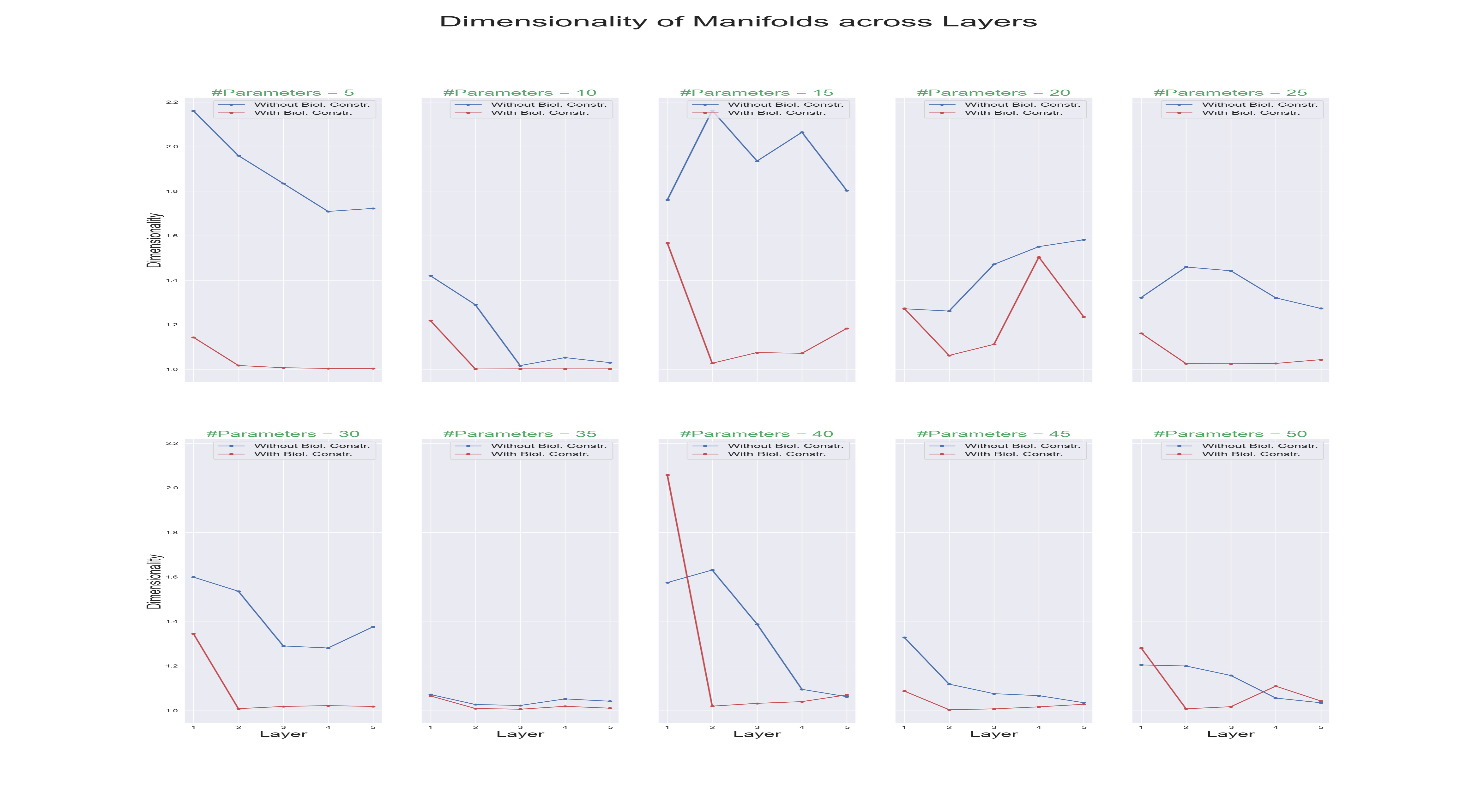
- Method 1: (Effective Dimensionality)
  - $\mathbf{x} \in R^N, X \in R^{N \times P}, \mathbf{x_0} \in R^N, C = \frac{1}{P} X X^T - \mathbf{x_0} \mathbf{x_0}^T$
  - $C = \frac{1}{P} \sum_{i=1}^{N} R_i^2 \mathbf{u_i} \mathbf{u_i}^T, R^2 = \frac{1}{N} \sum_{i=1}^{N} R_i^2, D = \frac{\sum_{i=1}^{N} (R_i^2)^2}{\sum_{i=1}^{N} R_i^4}$
- Method 2: (Intrinsic Dimensionality)
  - $f(\mu) = \frac{d}{\mu^{1+d}}, F(\mu) = 1 - \mu^{-d}, d \sim - \frac{\log(1 - F(\mu_i))}{\log(\mu_i)}$

Sorscher et al. 2022

Facco et al. 2017

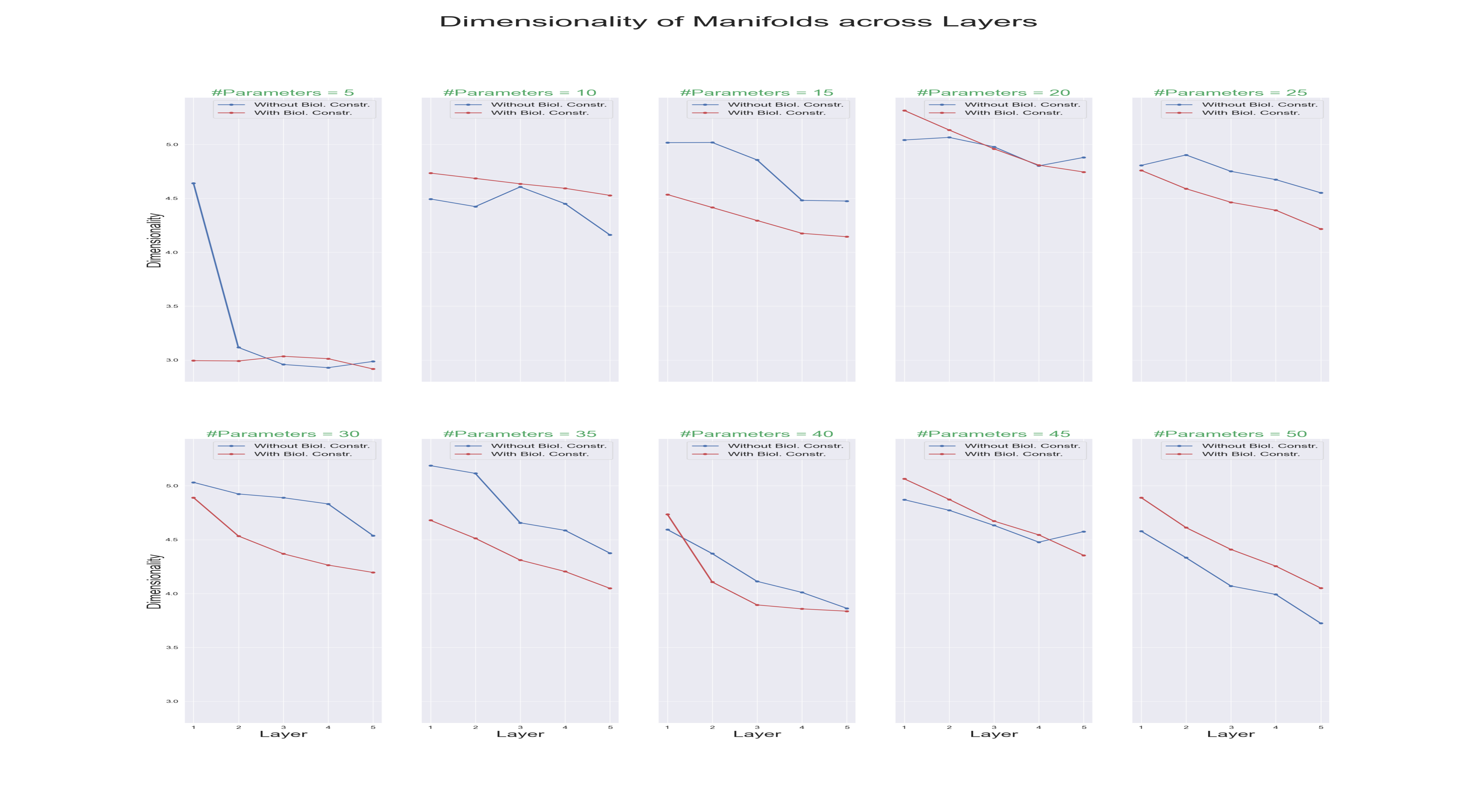
NeuroAILab

Results from my previous work (Dimension = 6) (Method 1)

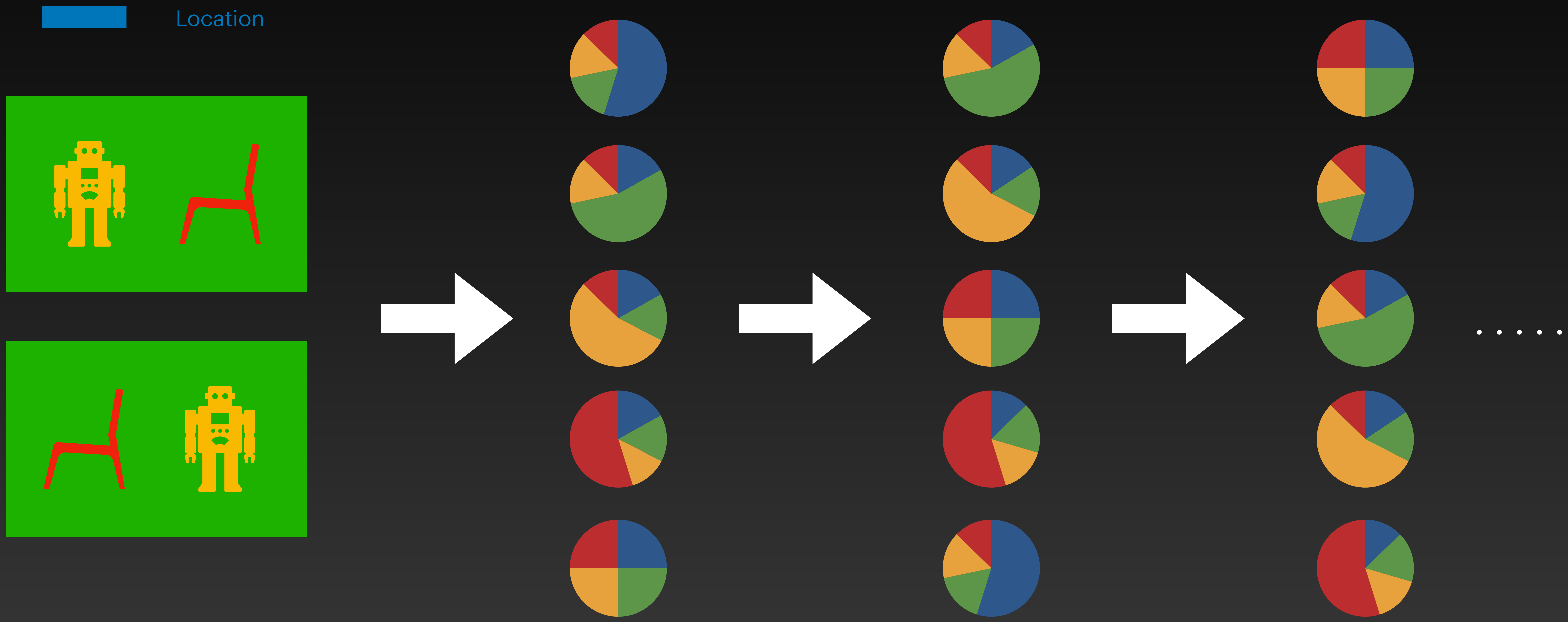


NeuroAILab

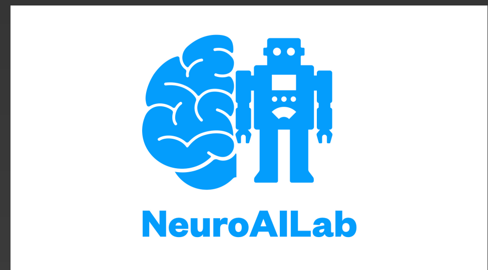
Results from my previous work (Dimension = 6) (Method 2)



Disentanglement representations in DNN



Kim & Minh (2018)
Whittington et al. (2022)



Adding biological constraints to force disentanglement representations in DNN

$$L = L_{nonneg} + L_{activity} + L_{weight} + L_{prediction}$$

$$L = \beta_{nonneg} \sum_{i} max(-a_i, 0) + \beta_{activity} \sum_{l} \|\mathrm{a}_l\|^2 + \beta_{weight} \sum_{l} \|\mathrm{W}_l\|^2 + L_{prediction}$$

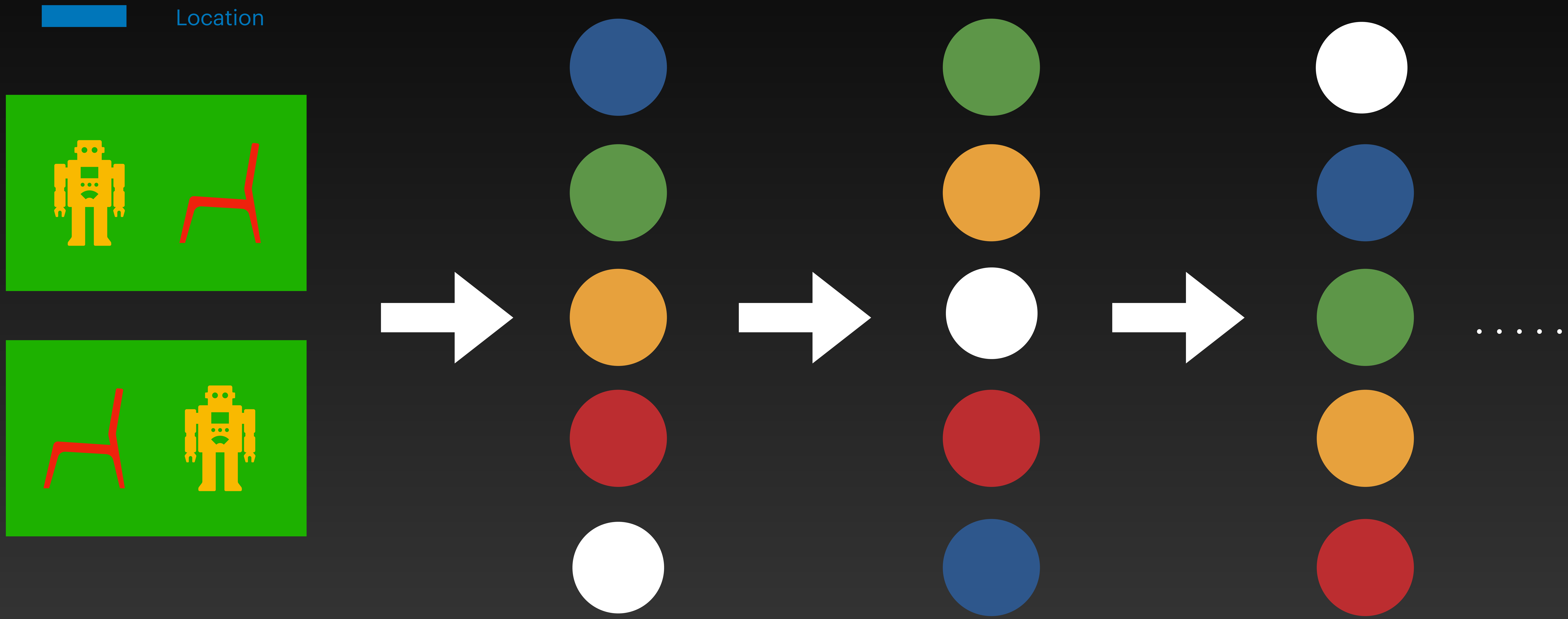
$$r_n = \frac{max_f(I_{n,f})}{\sum_f (I_{n,f})}$$

$$MIR = \frac{\frac{\sum_n r_n}{n_n} - \frac{1}{n_f}}{1 - \frac{1}{n_f}}$$

Whittington et al. (2022)

NeuroAILab

Disentanglement representations in DNN



Kim & Minh (2018)
Whittington et al. (2022)

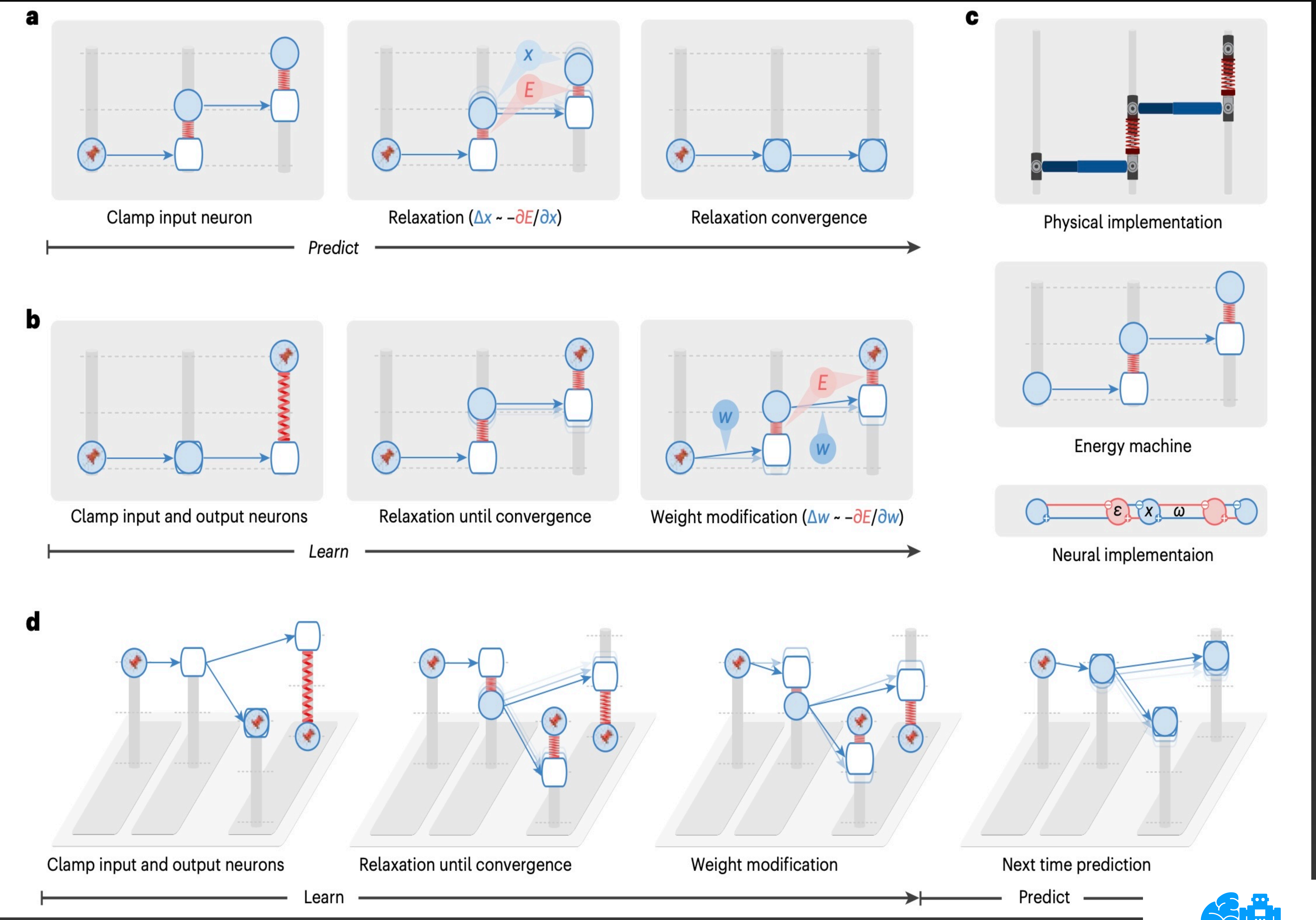
Energy-based networks and prospective configuration

- Backpropagation is not biologically plausible due to requiring non-local information and assuming symmetries in backward and forward weights. Furthermore, it requires more exposure to training data. Also, it suffers from catastrophic interference.
- Energy-based networks have been widely used in describing biological neural systems. In these models, a neural circuit is described by a dynamical system driven by reducing an abstract 'energy', for example, reflecting errors made by neurons. Neural activity and weights change to reduce this energy; hence, they can be considered 'movable parts' of the dynamical system. Energy-based networks can be shown that they are mathematically equivalent to a physical machine (we call it 'energy machine'), where the energy function has an intuitive interpretation, and its dynamics are straightforward; the energy machine simply adjusts its movable parts to reduce energy.
- In prospective configuration, before synaptic weights are modified, neural activity changes across the network so that output neurons better predict the target output; only then are the synaptic weights (hereafter termed 'weights') modified to consolidate this change in neural activity. By contrast, in backpropagation, the order is reversed; weight modification takes the lead, and the change in neural activity is the result that follows.

Song et al. (2024)

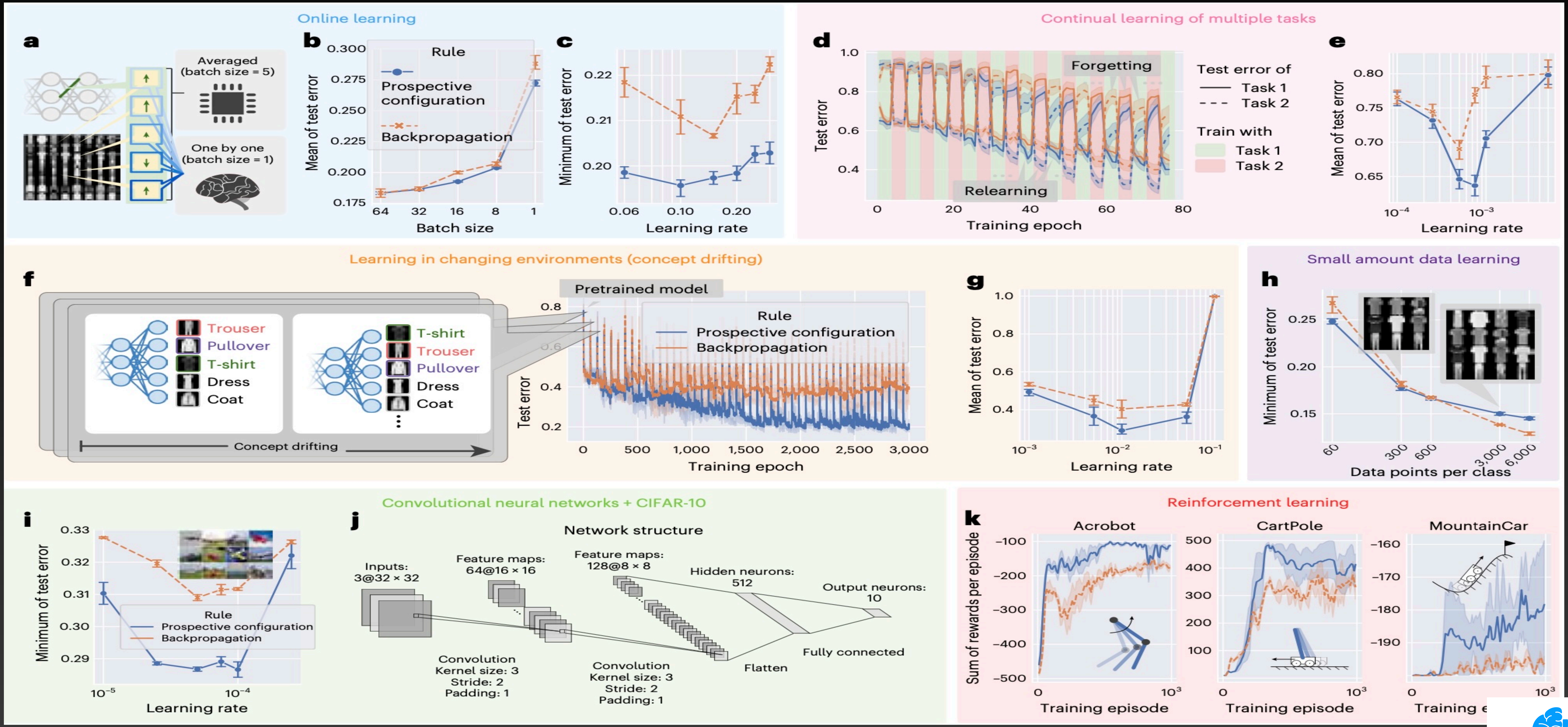
Energy-based networks and prospective configuration

In the energy machine, the activity of a neuron corresponds to the height of a node (represented by a solid circle) sliding on a post. The input to the neuron is represented by a hollow node on the same post. A synaptic connection corresponds to a rod pointing from a solid node to a hollow node. The weight determines how the input to a postsynaptic neuron depends on the activity of a presynaptic neuron; hence, it influences the angle of the rod. In energy-based networks, relaxation (that is, neural dynamics) and weight modification (that is, weight dynamics) are both driven by minimizing the energy, which corresponds to relaxation of the energy machine by moving the nodes and tuning the rods (that is, weight dynamics) are both driven by minimizing the energy, which corresponds to relaxation of the energy machine by moving the nodes and tuning the rods, respectively.



Song et al. (2024)

Energy-based networks and prospective configuration



Song et al. (2024)

What I am planning to do with energy-based networks and prospective configuration as a learning rule?

- I would like to measure the dimensionality and disentanglement of intermediate representations in energy-based networks trained by prospective configuration on simple toy datasets and compare them to the case of ordinary DNNs with or without biological constraints. Also, I would like to know which one is more similar to the biological neural network.

- Moreover, energy-based networks and prospective configuration need less data to be trained and if they are implemented on analog hardware they use less energy. These make them suitable for next generation of Large Language Models.

- Furthermore, after converting any ordinary DNN to an energy-based network, prospective configuration can be used as a learning rule for training. As an ambitious project, it would be interesting to convert a GPT-like neural network to an energy-based network and train it from scratch via prospective configuration.

- Finally, energy-based networks and prospective configuration are close to what is called cybernetic mode of computation, bottom-up physical computation, or unconventional computing. On the other hand, it is close to the formulation of free energy principle, and the debate around reward hypothesis in reinforcement learning literature or origin of life and agency in artificial life literature. I have this feeling that unconventional computing is the ultimate way of solving the problem of agency, free will, and consciousness in reward free artificial intelligence.

Song et al. (2024)
Jaeger et al. (2023)
Bowling et al. (2023)
Abel et al. (2024)
Watson (2023)
Seifert et al. (2023)
Fields & Levin (2023)
Roleau & Levin (2023)

NeuroAILab

NeuroAILab

Thanks for your attention!

I was introduced to this field when I started my work at the European Neuroscience Institute - Göttingen, so I would like to thank all my colleagues there specially:

Caspar Schwiedrzik, Giorgio Manenti