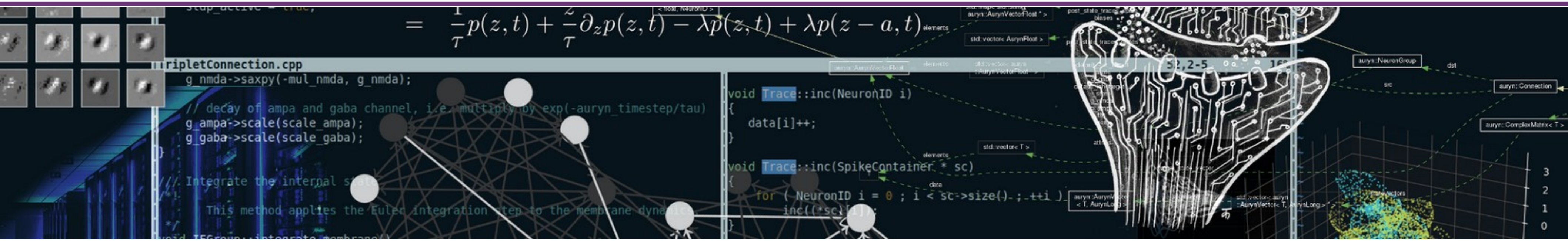
Friedrich Miescher Institute for Biomedical Research

University of Basel



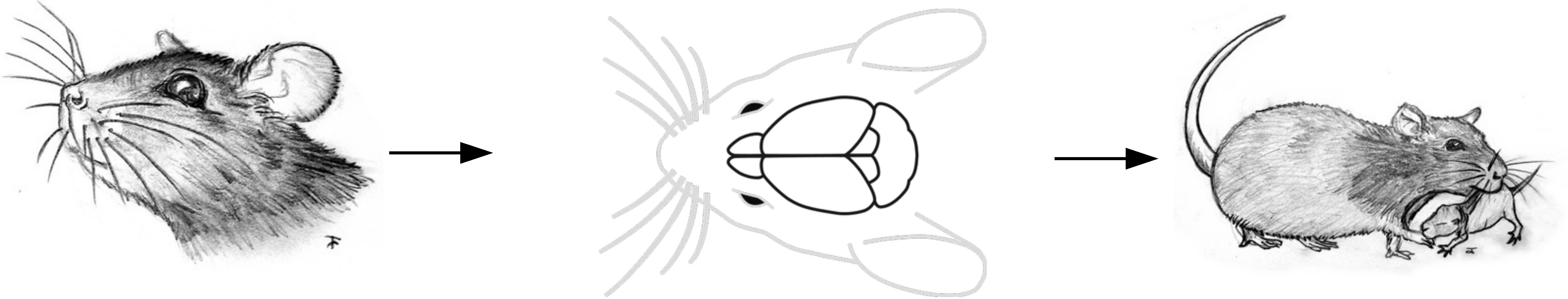
Principles of predictive representation learning in biological neural networks

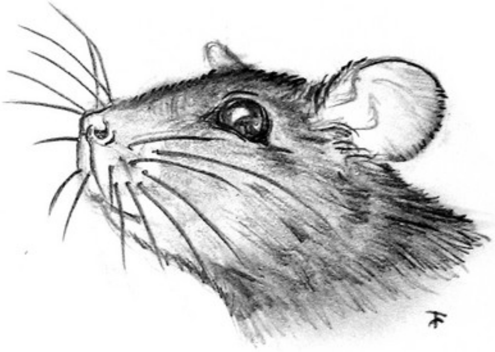
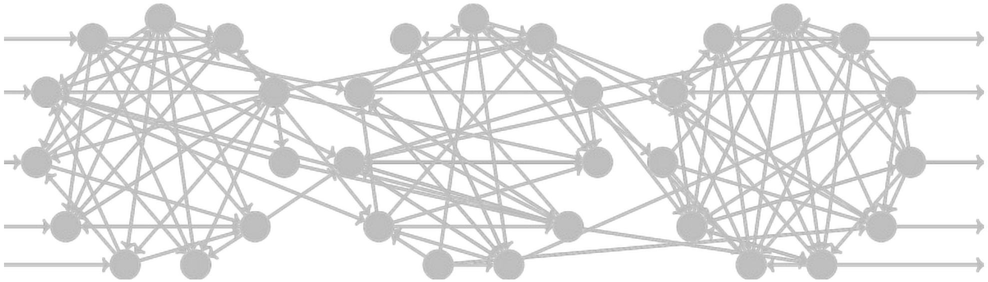
Preprint

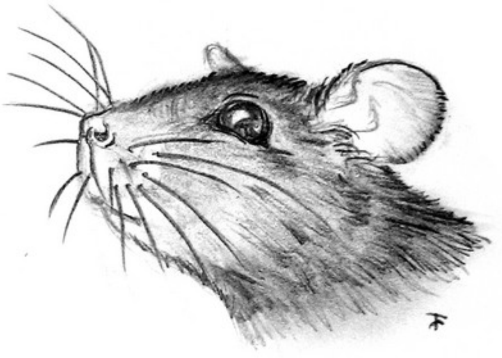
Friedemann Zenke

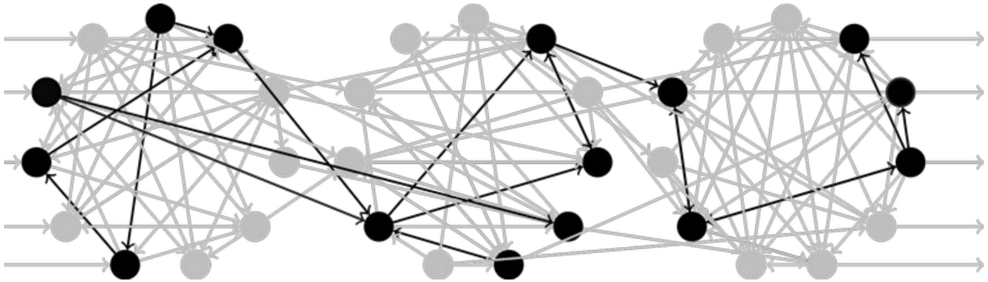
Computational Neuroscience @ FMI

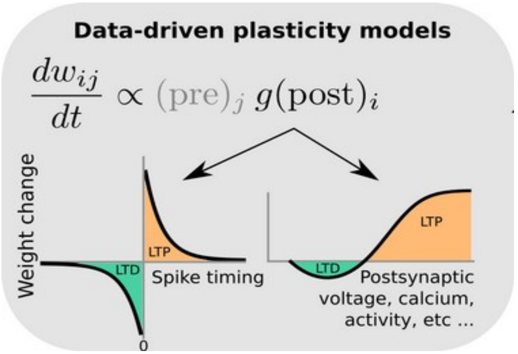
www.zenkelab.org





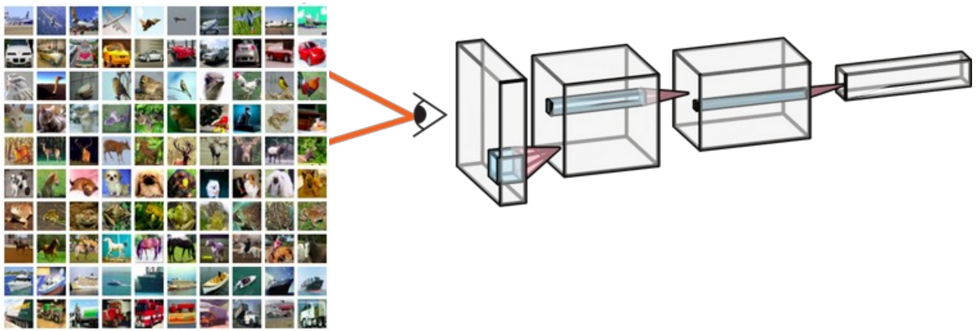


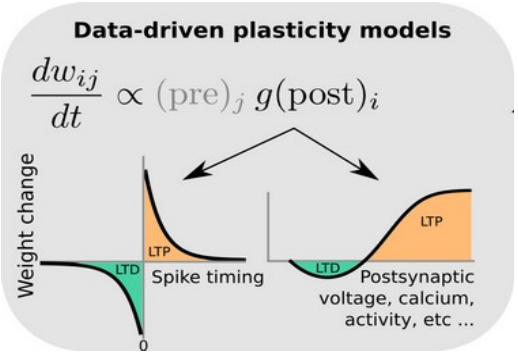
When an axon of cell A […] excites cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells so that A's efficiency as one of the cells firing B is increased.

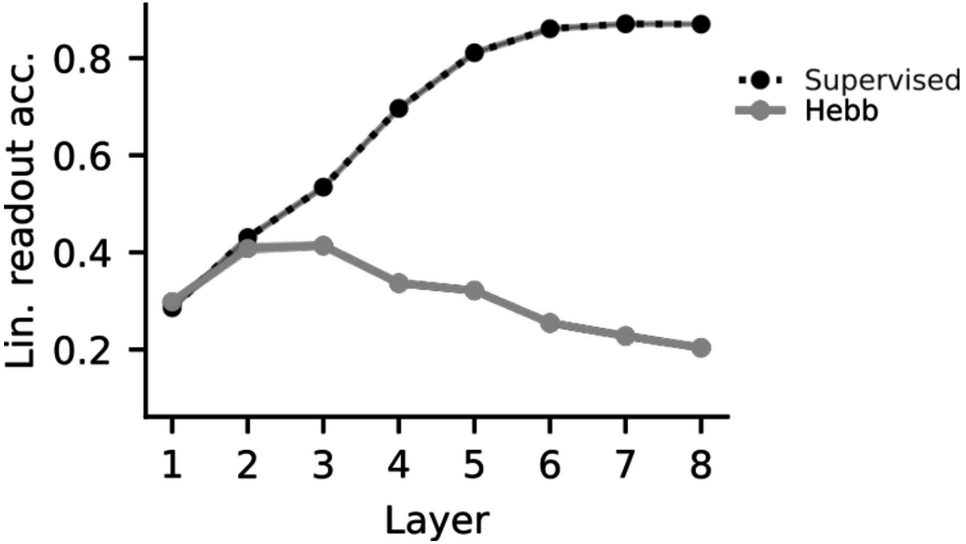
D.O. Hebb, The Organization of Behavior, 1949.

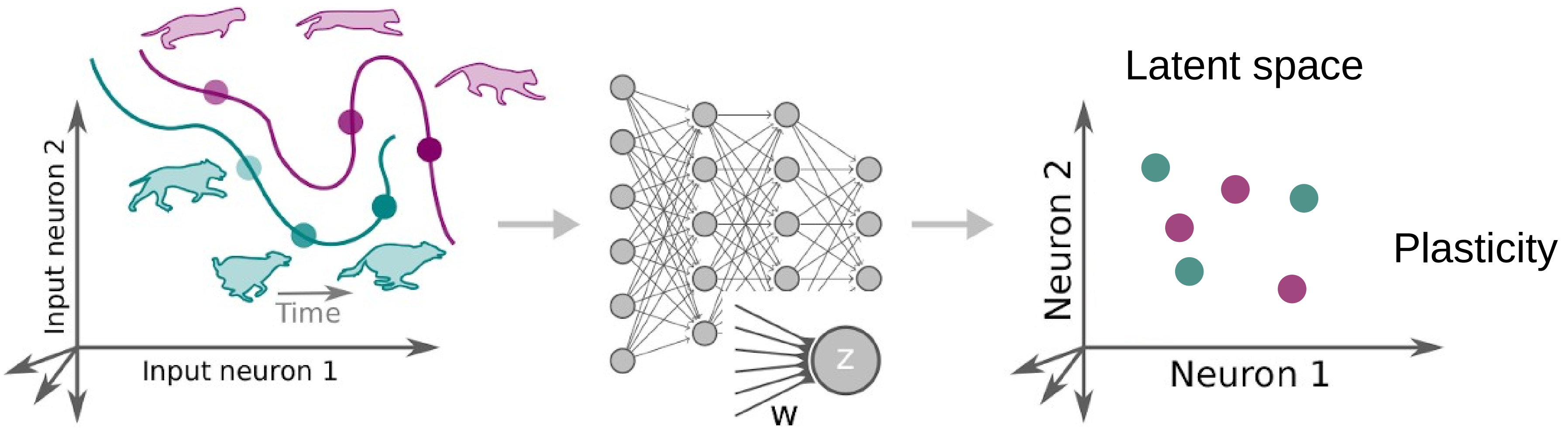
Problem

Hebbian plasticity does not learn *good* representations in deep neural networks

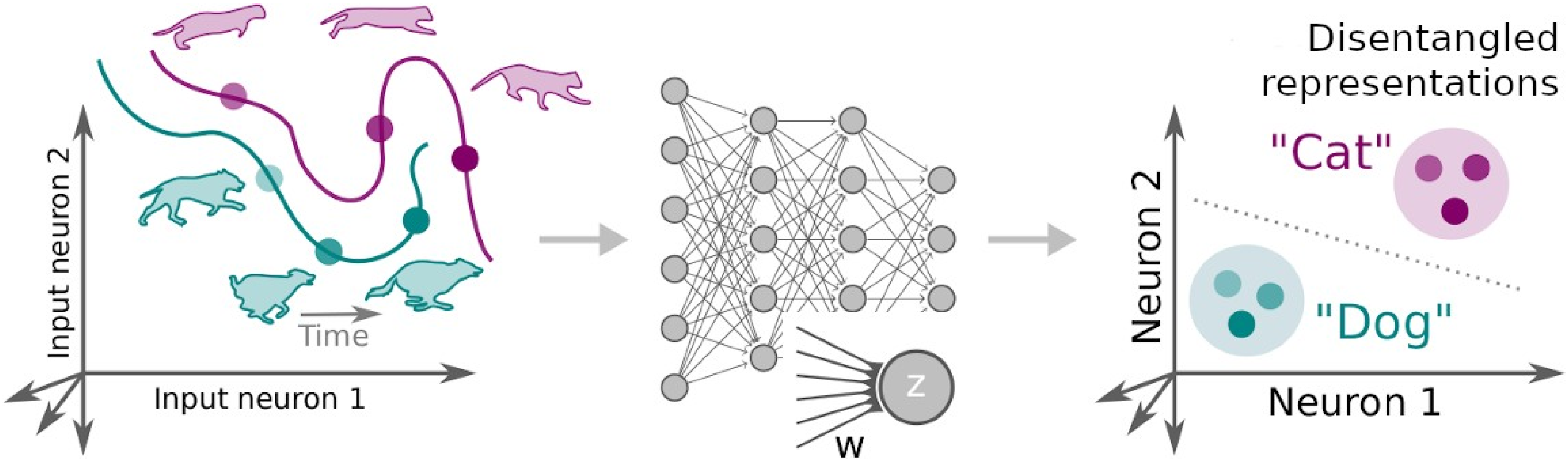


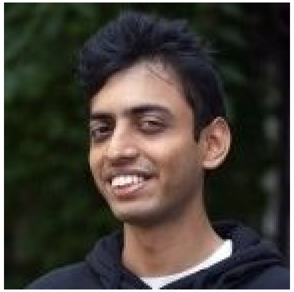





Input neuron 2
Input neuron 1
Time
z
w
Latent space
Neuron 2
Neuron 1
Plasticity

Halvagal and Zenke (2022)
https://www.biorxiv.org/content/10.1101/2022.03.17.484712v3



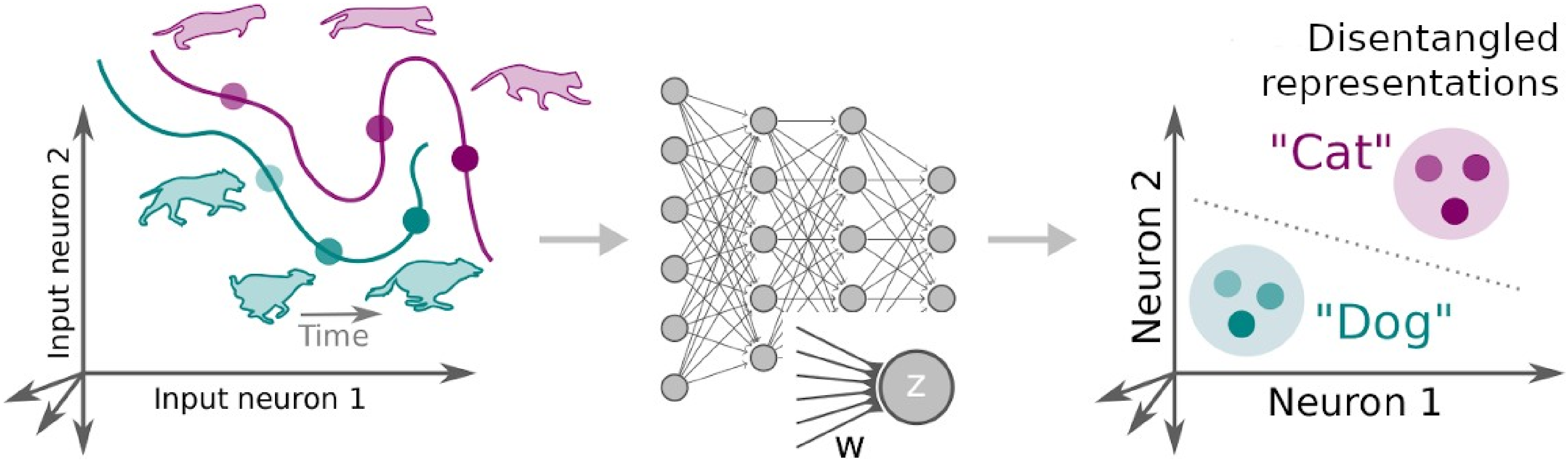


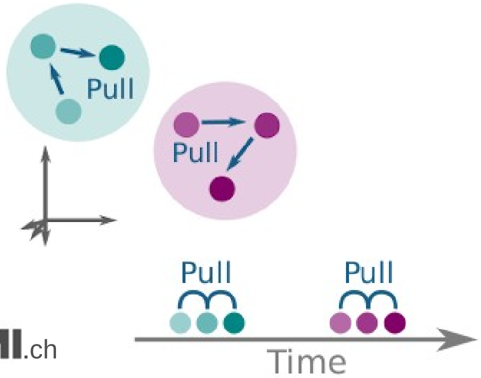
Manu Srinath Halvagal

Two fundamental questions

How does the brain do credit assignment?
Which objective do sensory networks optimize?

Idea: Use prediction in latent space as learning objective
(Also the core idea of self-supervised learning in ML)

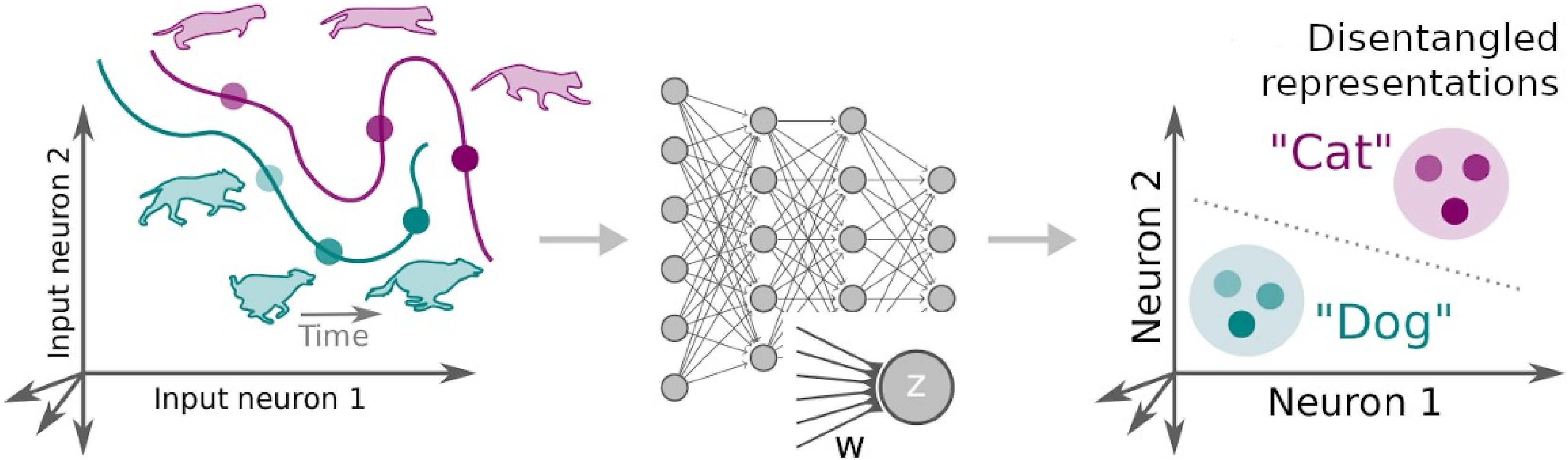




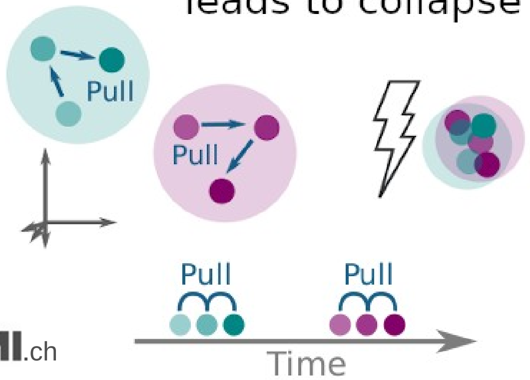
$$\mathcal{L} = (z_{\text{Cat1}} - z_{\text{Cat2}})^2 + (z_{\text{Dog1}} - z_{\text{Dog2}})^2$$

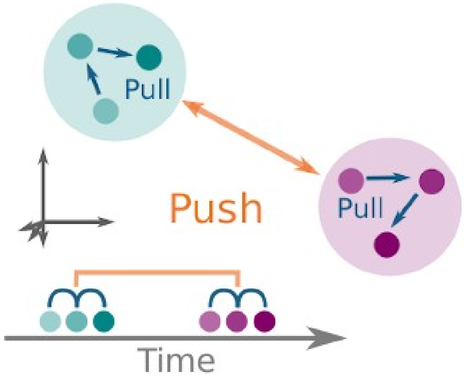
$$\mathcal{L} = (z(t_1) - z(t_1 + \Delta))^2 + (z(t_2) - z(t_2 + \Delta))^2$$

Idea: Use prediction in latent space as learning objective
(Also the core idea of self-supervised learning in ML)



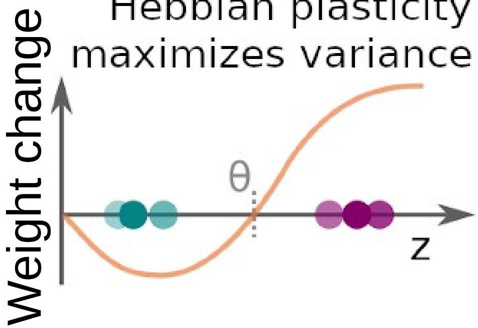
Purely predictive learning leads to collapse





Oja (1982):

Hebbian plasticity maximizes variance



Pull

Push

$$\mathcal{L} = \mathcal{L}_{\text{Pred.}} + \mathcal{L}_{\text{Hebb}}$$

A loss function combining prediction (pull) with variance maximization (push)

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \mathcal{L}_{\text{Hebb}}$$

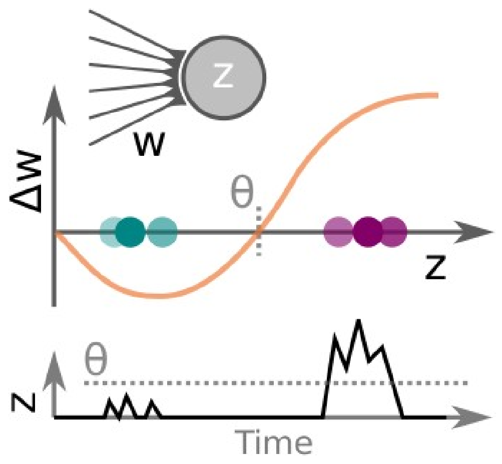
Same idea as in SFA: “Encourage slowness of representation”

$$\mathcal{L}_{\text{pred}} \propto \sum_t \left(\frac{dz}{dt}\right)^2 \approx \sum_t (z(t) - \boxed{\text{SG}}(z(t-\Delta t)))^2$$

important

$$\mathcal{L}_{\text{Hebb}} \propto -\log \sum_t (z(t) - \theta(t))^2 = -\boxed{\log} \sum_t \sigma^2(t)$$

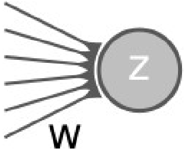
with $\theta(t) = \bar{z}(t)$



The **Latent Predictive Learning rule** (LPL)

A local learning rule that extends BCM theory

BCM: Bienenstock, Cooper, and Munroe (1982)



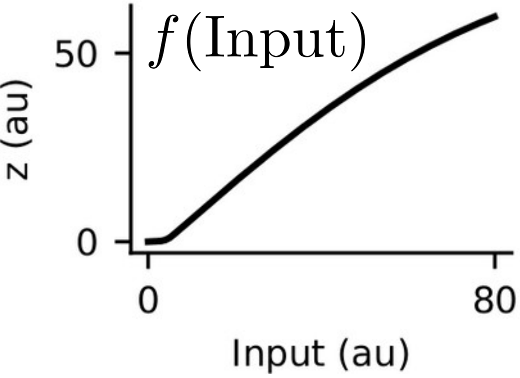
"pre"

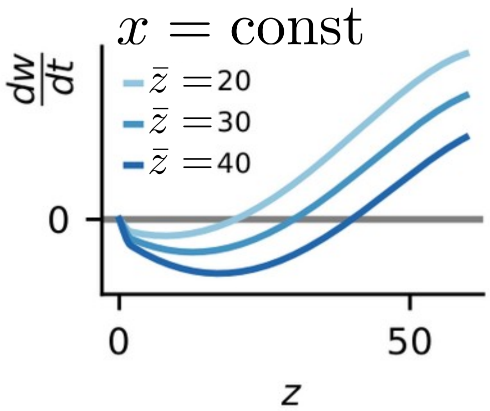
Beyond BCM

"post"

Moving threshold

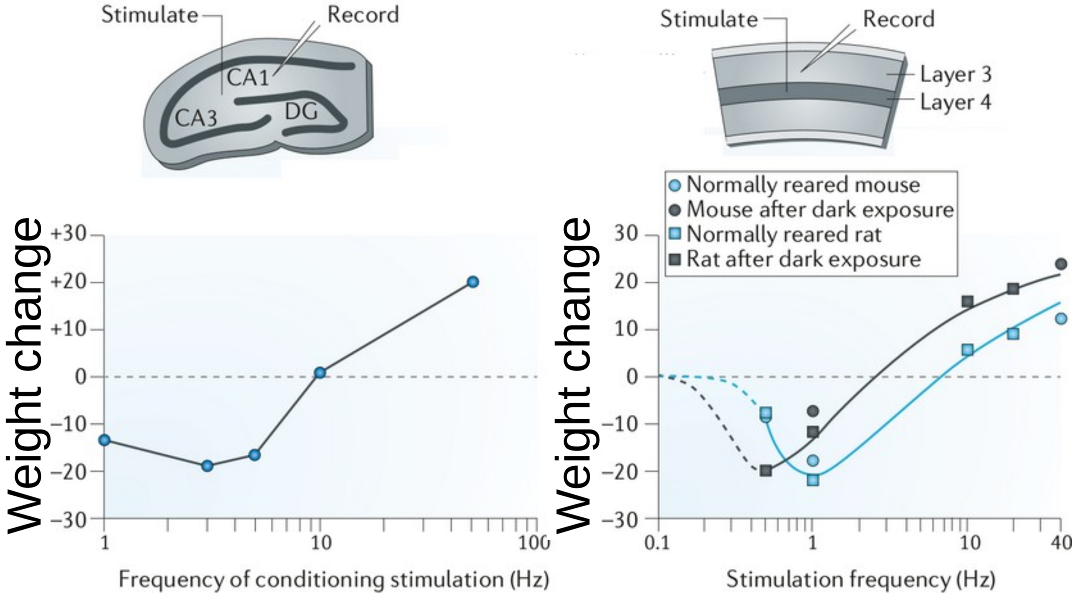
$$\frac{\mathrm{d}W_j}{\mathrm{d}t}(t) = \eta x_j(t) f'(a(t)) \left(\underbrace{\boxed{-\frac{\mathrm{d}z(t)}{\mathrm{d}t}}}_{\text{Pull}} + \underbrace{\frac{\lambda}{\boxed{\sigma_z(t)^2}}(z(t)-\bar{z}(t))}_{\text{Push}} \right)$$



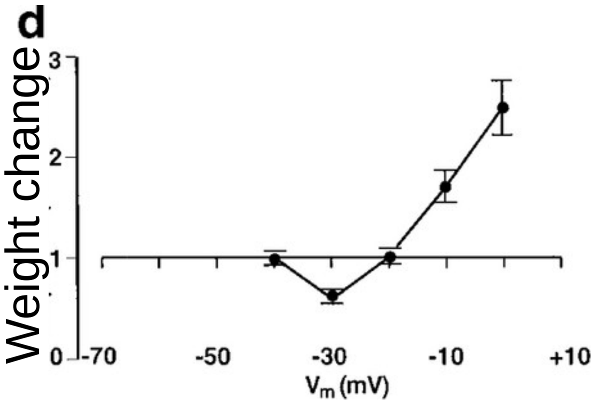


Postsynaptic activity

There is substantial experimental support at the synaptic and neuronal level for this type of Hebbian plasticity

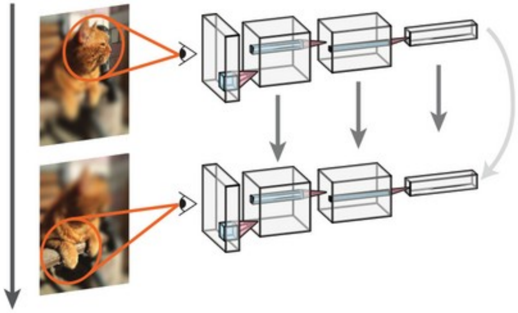


Cooper and Bear (2012). *Nat Rev Neurosci* 13, 798–810.

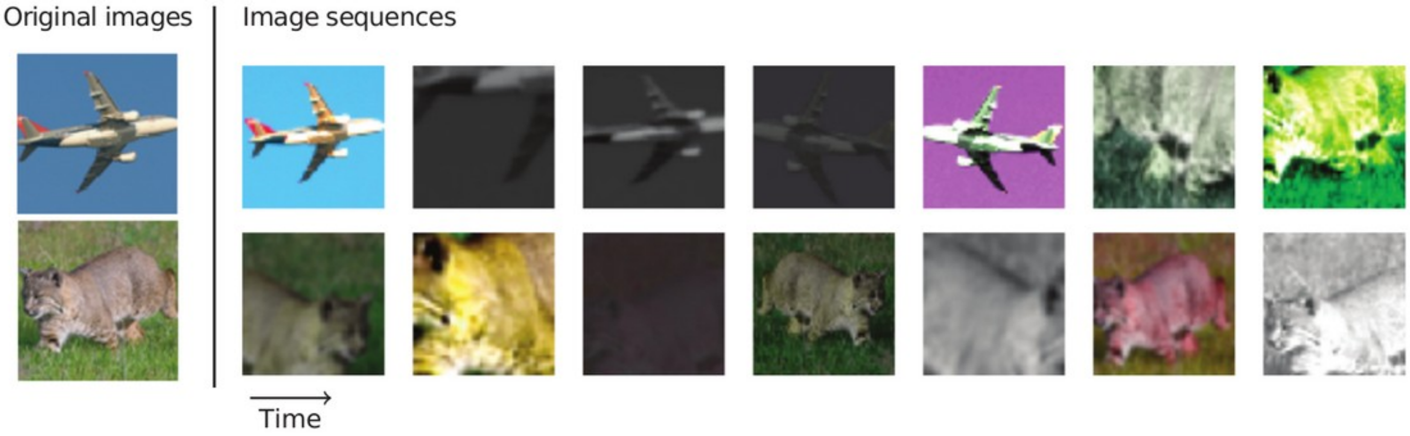


Ngezahayo, Schachner, and Artola (2000).
J. Neuroscience 20, 2451–2458

Training deep networks with LPL and “positive samples”

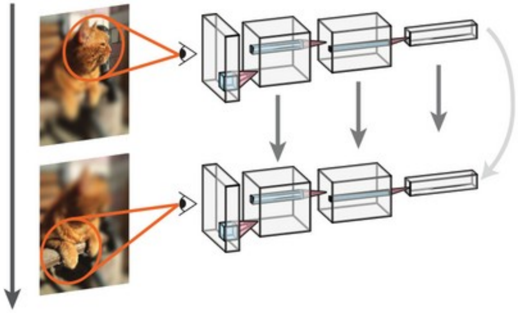


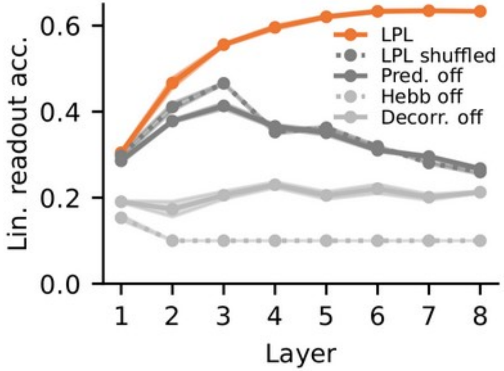
- VGG-11 model
- Neurons in all layers learn with **LPL**
- Layer-local: no backprop
- Prevent neuronal co-adaptation (e.g., lateral inhibition)

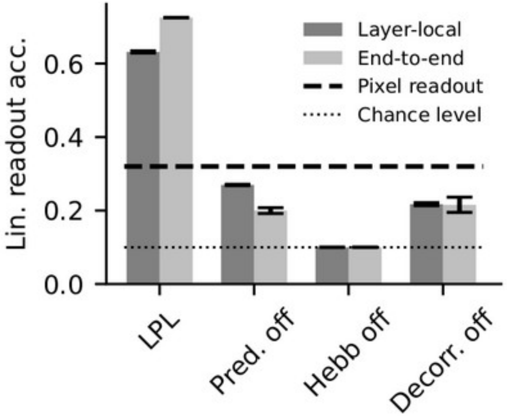


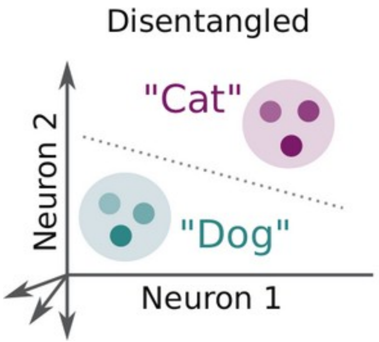
LPL progressively learns invariant representations

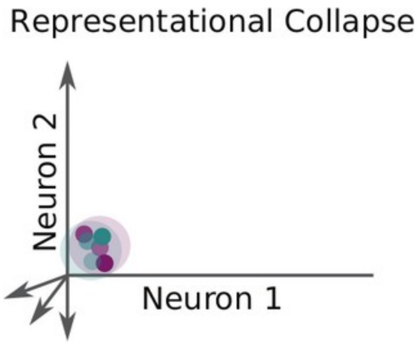
After “watching” millions of image sequences …

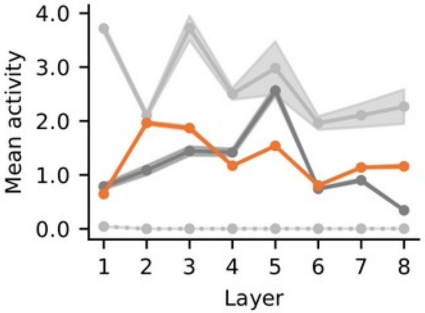


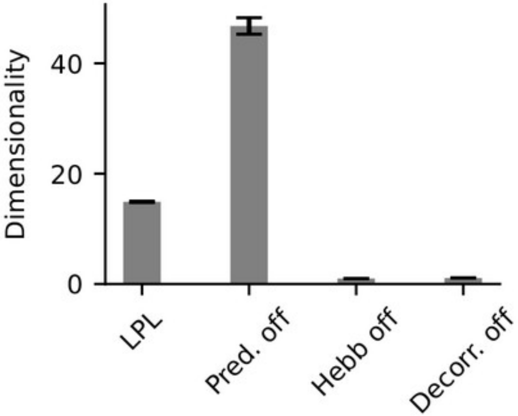












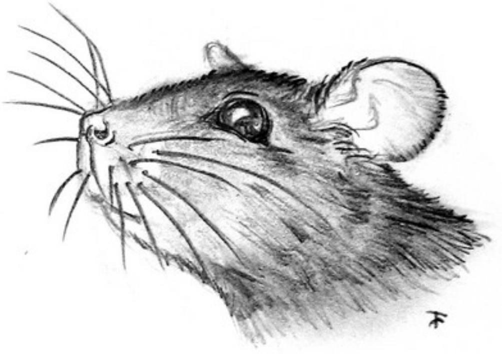
Interim summary

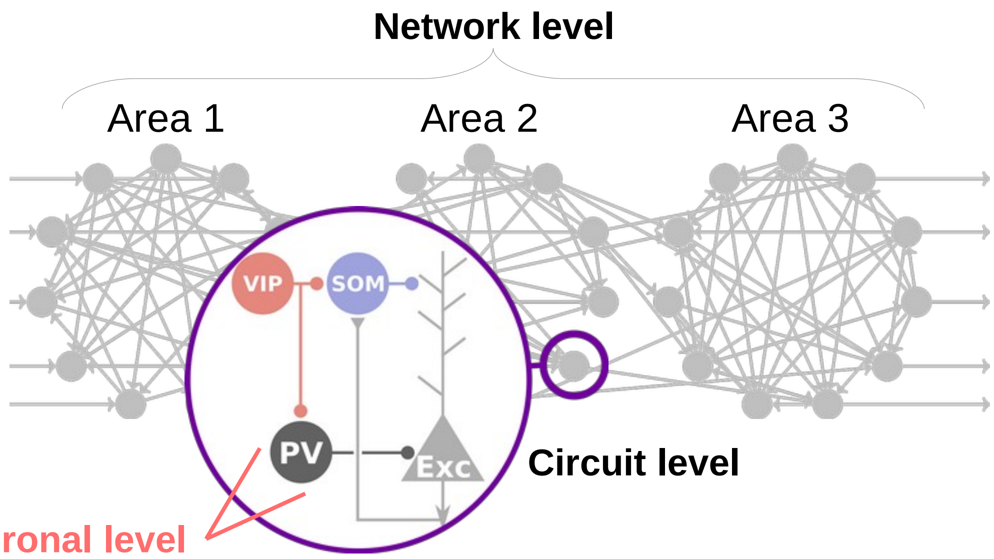
- Yes, LPL disentangles represention in deep nets
- LPL combines Hebbian and predictive plasticity
- LPL is a local learning rule that combined with decorrelation does quite well
- Looks similar to and extends BCM learning rule

Big question

Is LPL consistent with experiments?

Is LPL consistent with experiments?

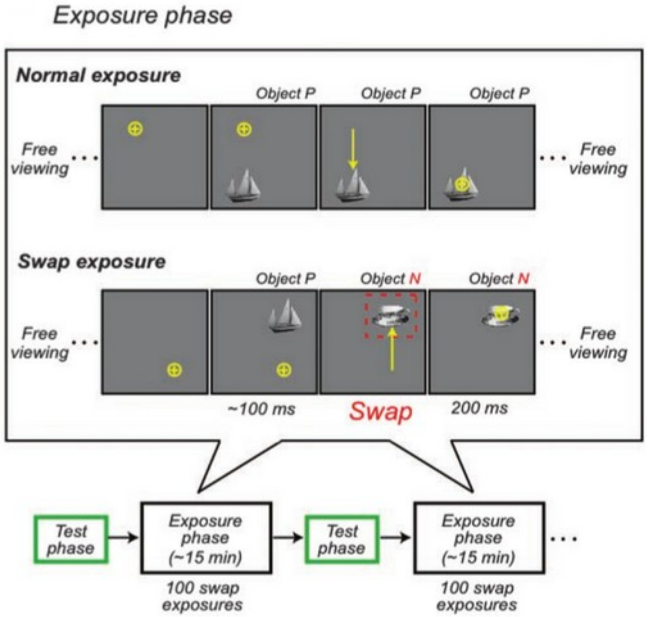




1) Network-level
2) Neuronal/synaptic level

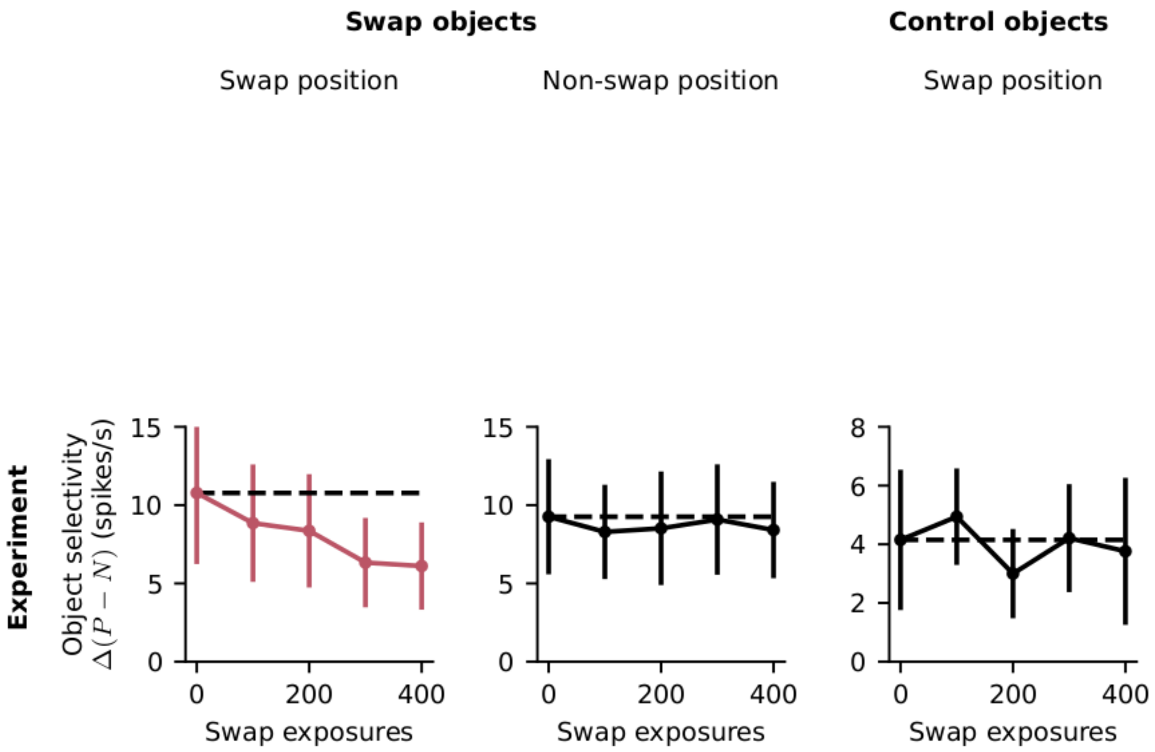
Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex

Nuo Li and James J. DiCarlo*

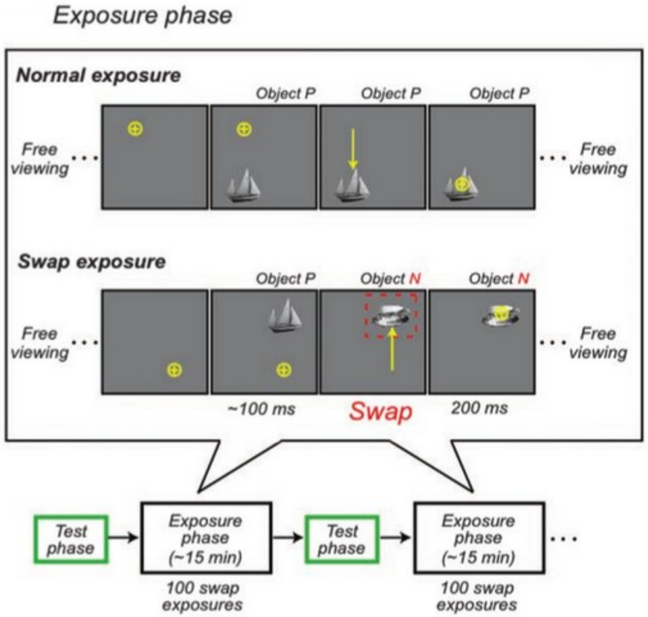


 Li, N., and DiCarlo, J.J. (2008). *Science* 321, 1502–1507.

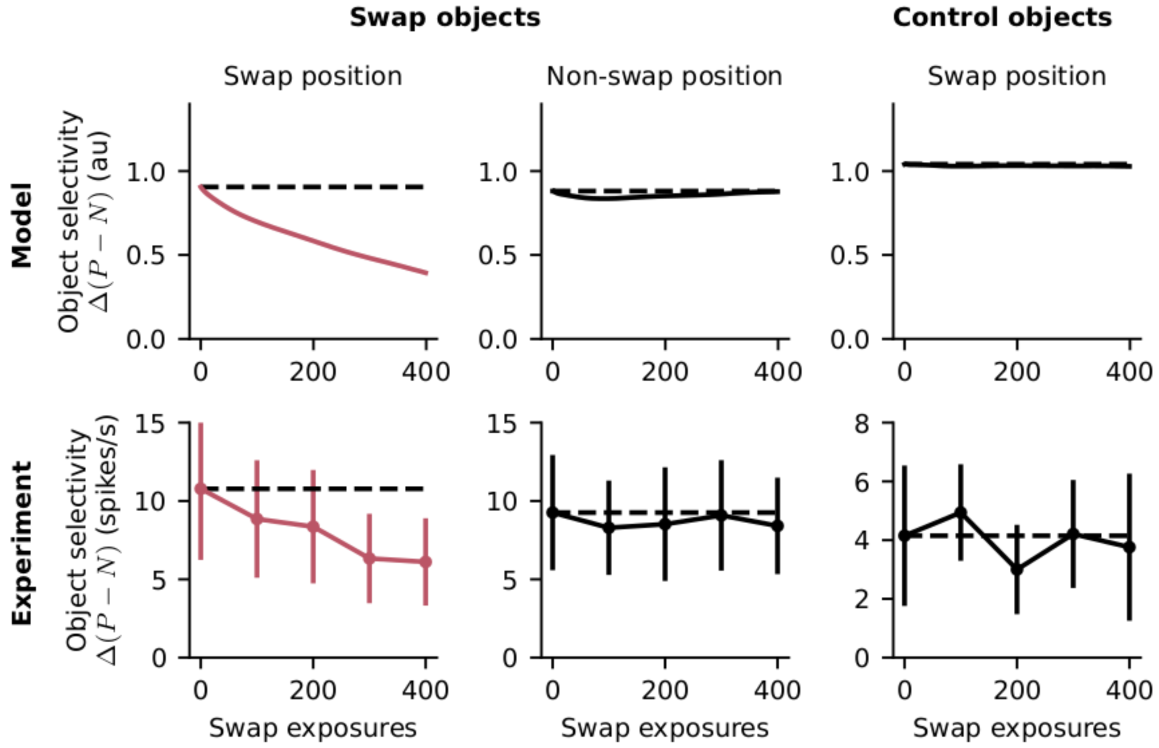
Unsupervised Natural Experience Rapidly Alters Invariant Object Representation in Visual Cortex



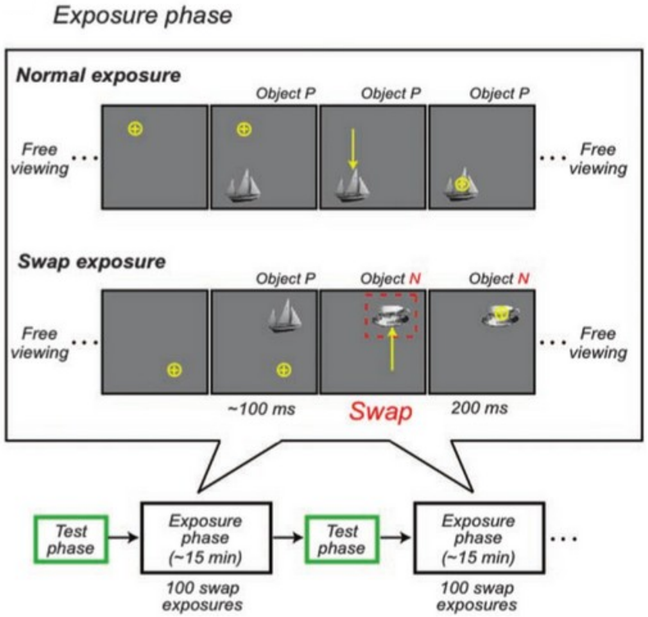
(Li and DiCarlo, 2008)



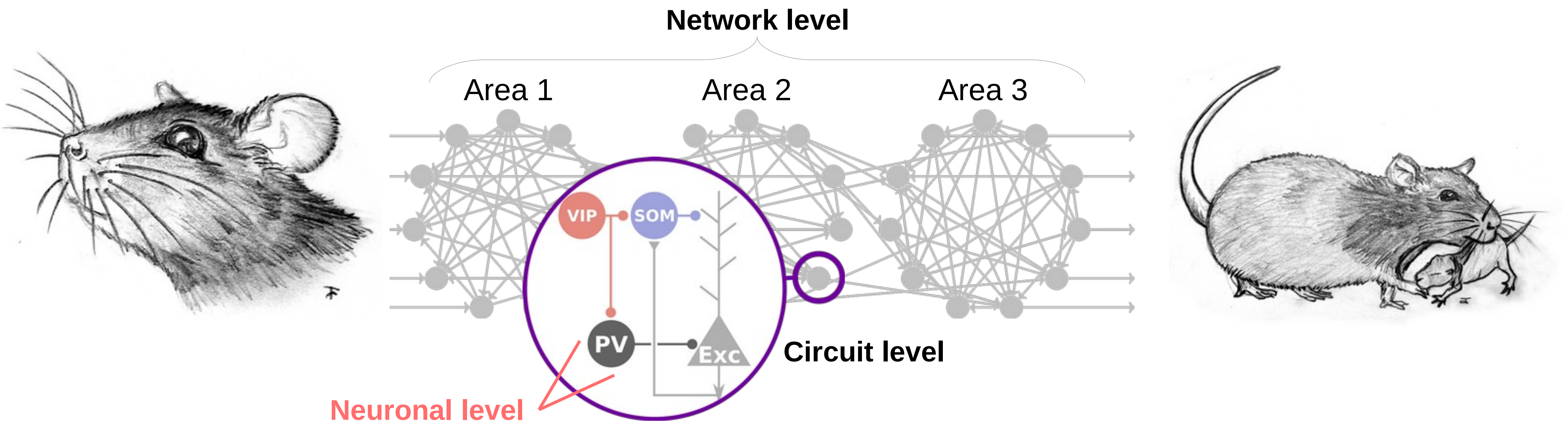
LPL accounts for experiments



(Li and DiCarlo, 2008)



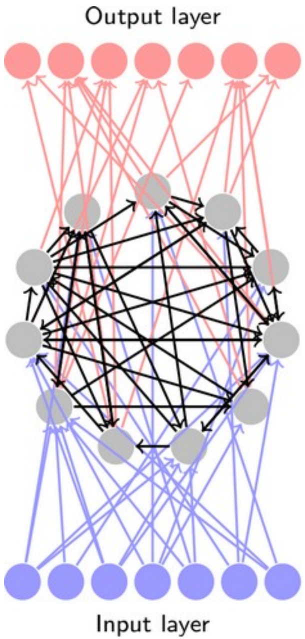
LPL is consistent with representation changes at the network level

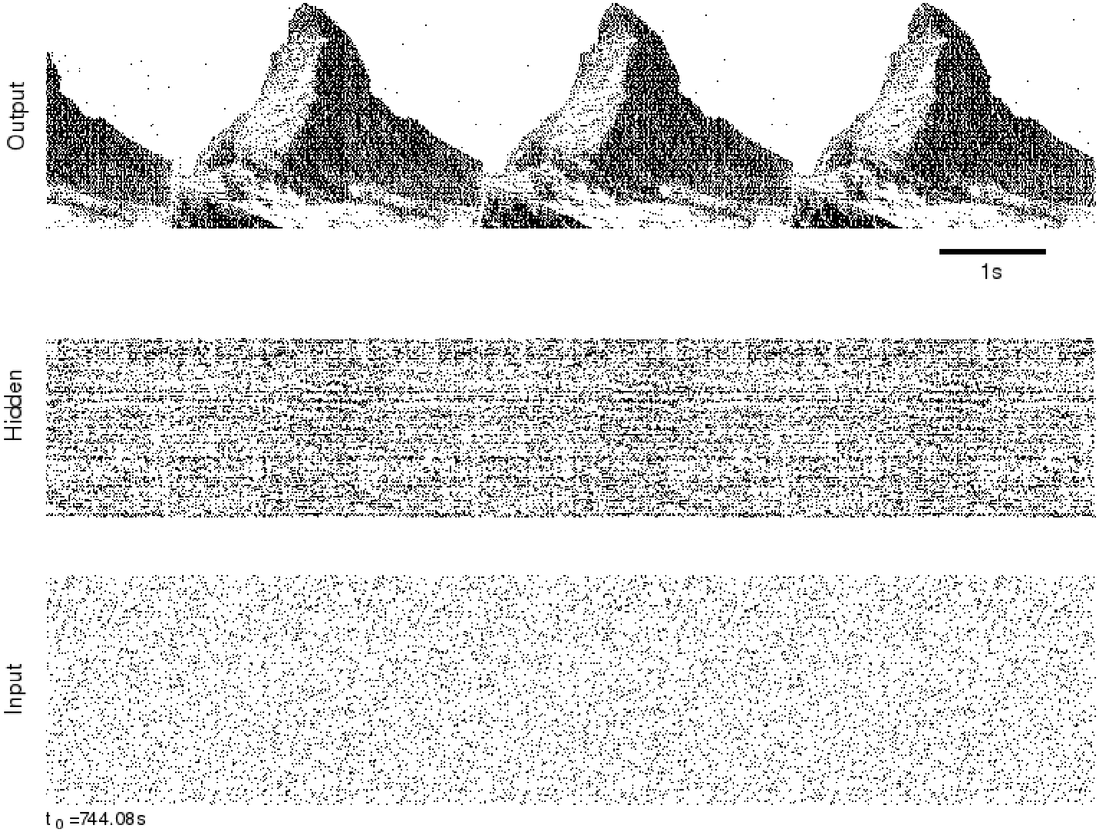


1) ~~Network level~~
2) Neuronal/synaptic level

What about spiking nets and STDP?

Training spiking networks online

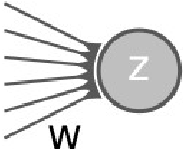




Zenke and Ganguli (2018). Neural Computation
Zenke and Neftci (2021). Proceedings of the IEEE

The spiking LPL rule

Based on SuperSpike: Zenke & Ganguli (2018)



$$S_i(t)$$

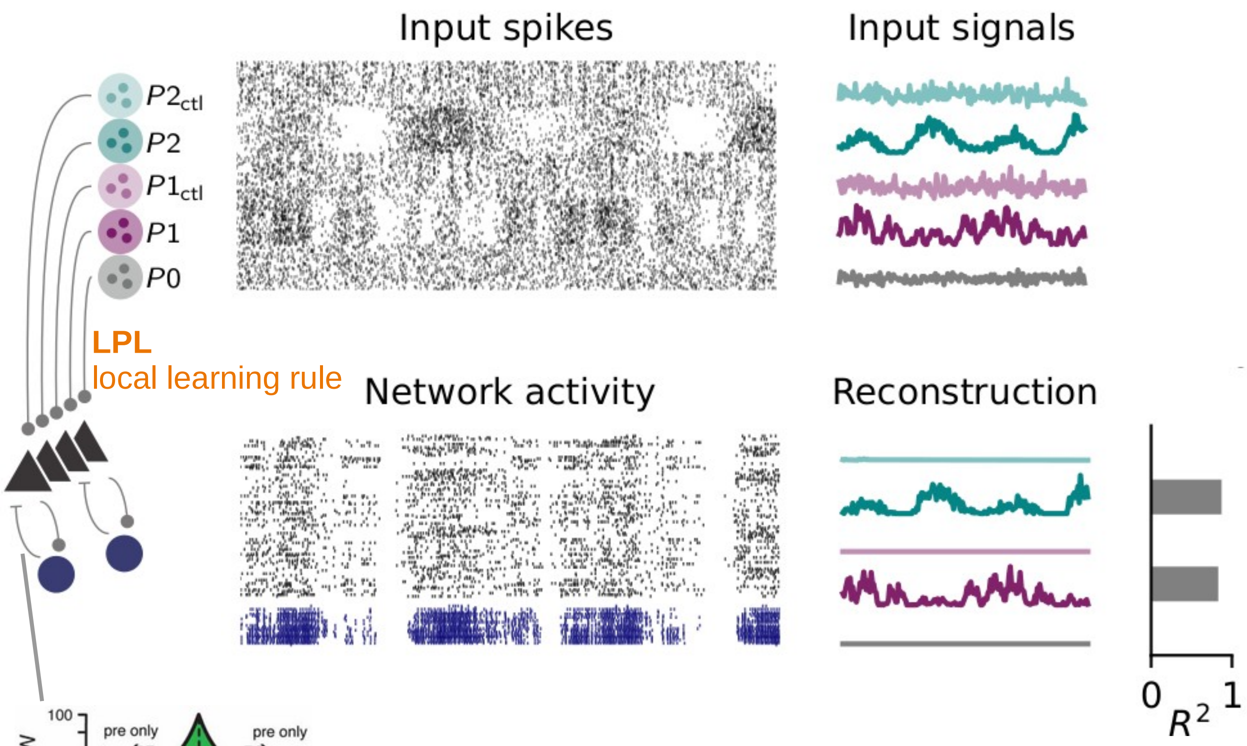
$$\mathcal{L} = \mathcal{L}_{\mathrm{pred}} + \mathcal{L}_{\mathrm{Hebb}}$$

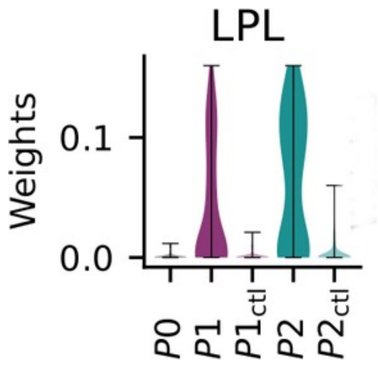
+ inhibitory neurons & plasticity

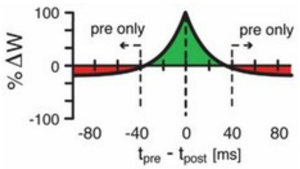
$$\frac{\mathrm{d}w_{ij}}{\mathrm{d}t} = \eta\, \alpha * \left(\underbrace{\epsilon * S_j(t)}_{\text{pre}} \underbrace{f'(U_i(t))}_{\text{post}} \right) \left[\alpha * \left(\underbrace{-\left(S_i(t) - S_i(t - \Delta t)\right)}_{\text{predictive}} + \underbrace{\frac{\lambda}{\sigma_i^2 + \xi} \left(S_i(t) - \bar{S}_i(t)\right)}_{\text{Hebb}} \right) \right]$$

$$+\, \eta \qquad \underbrace{\delta S_j(t)}_{\text{transmitter-triggered}}$$

LPL in spiking neural networks (SuperSpike + LPL Loss)



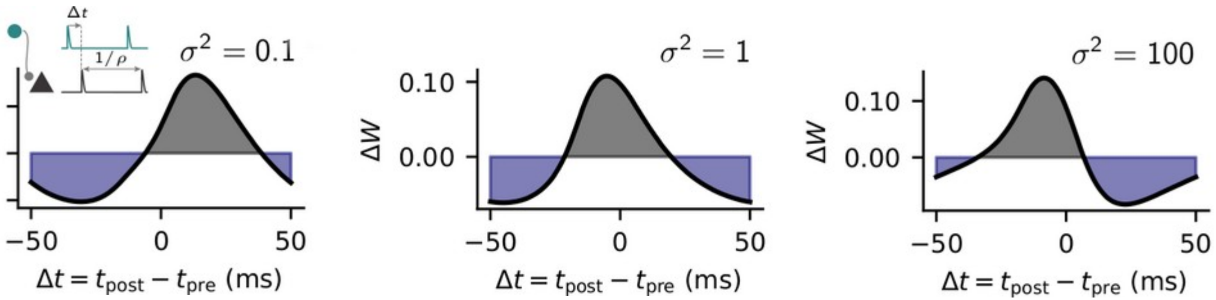


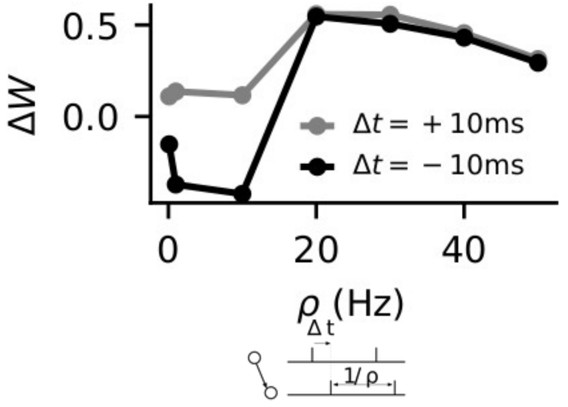


Inhibitory Plasticity
Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011)

LPL accounts for STDP and rate-based plasticity effects and predicts variance-dependent metaplasticity

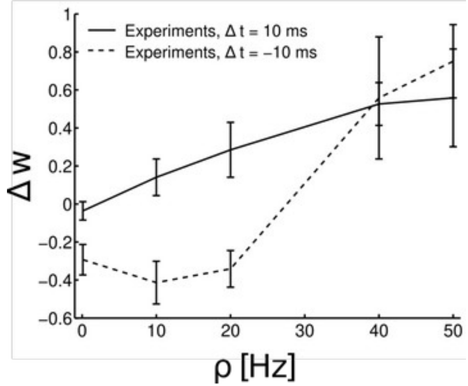
LPL





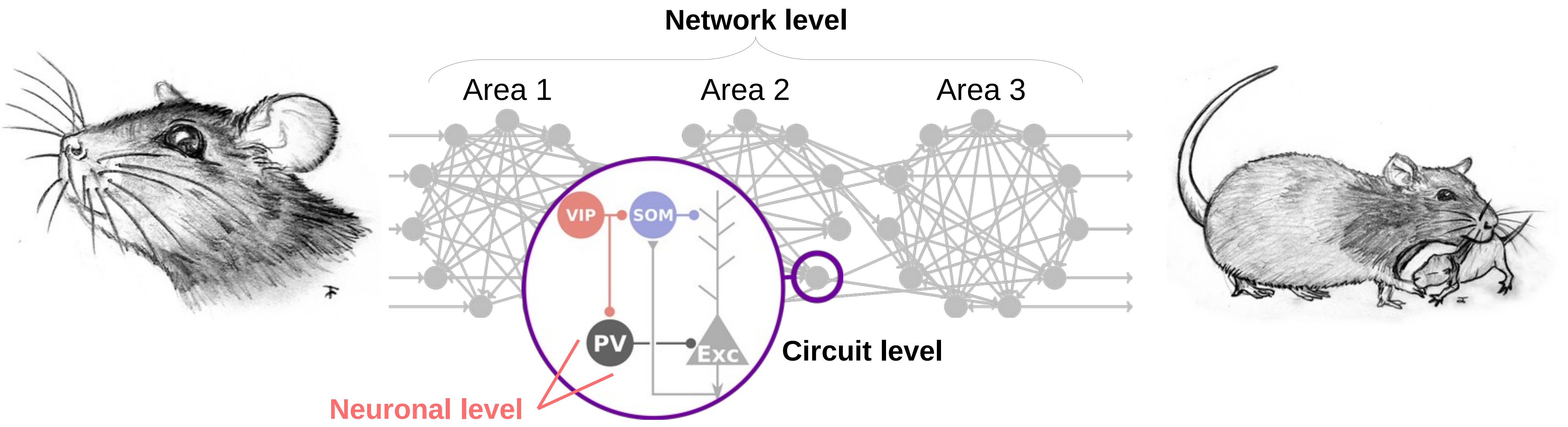
Predicts change of STDP window
with postsynaptic variance

$$\frac{\mathrm{d}W_j}{\mathrm{d}t}(t) = \eta x_j(t) f'(a(t)) \left(\underbrace{-\frac{\mathrm{d}z(t)}{\mathrm{d}t}}_{\text{predictive}} + \underbrace{\frac{\lambda}{\boxed{\sigma_z(t)^2}}(z(t) - \bar{z}(t))}_{\text{Hebbian}} \right)$$



Plot from: Pfister & Gerstner (2006)
Data: Sjöström et al. (2001)

LPL is consistent with observed synaptic plasticity



1) ~~Network level~~
2) ~~Neuronal/synaptic level~~

Summary

- LPL: A plausible local online rule that learns *good* representations in deep networks through prediction
- Consistent with experiments
  - Synaptic plasticity level (BCM, STDP)
  - Network representation level
- Predicts new form of metaplasticity which should manifest in modulation of STDP window
- Ongoing work: More realistic video input and check whether LPL accounts for other experiments (e.g., Matteucci and Zoccolan, 2020)

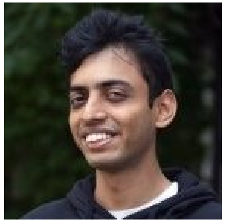
Thanks

We are hiring!

Preprint:

The team https://zenkelab.org/team/

Axel Laborieux



Manu Srinath Halvagal

Julian Rossbroich

Julia Gygax

Peter Buttaroni

Jeremias Seitz

Atena Mohammadi

Collaborators

Andreas Luthi (FMI), Georg Keller (FMI), Rainer Friedrich (FMI), Emre Neftci (FZ Julich), Johannes Schemmel (Uni Heidelberg), Richard Naud (uOttawa), Blake Richards (Mila)

Funding

Artwork: kyadava.net