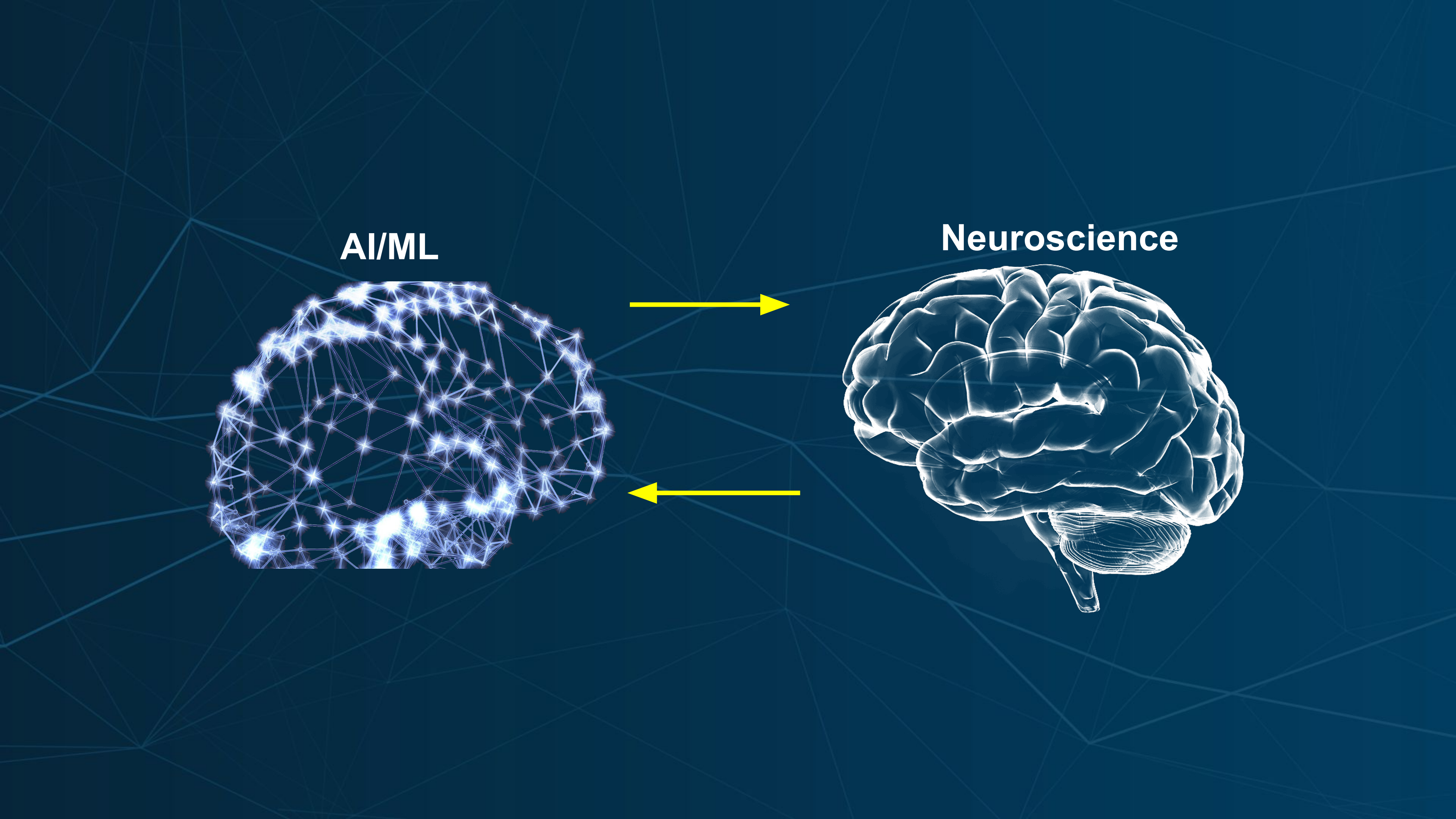
Recent advancements at the intersection of neuroscience and AI

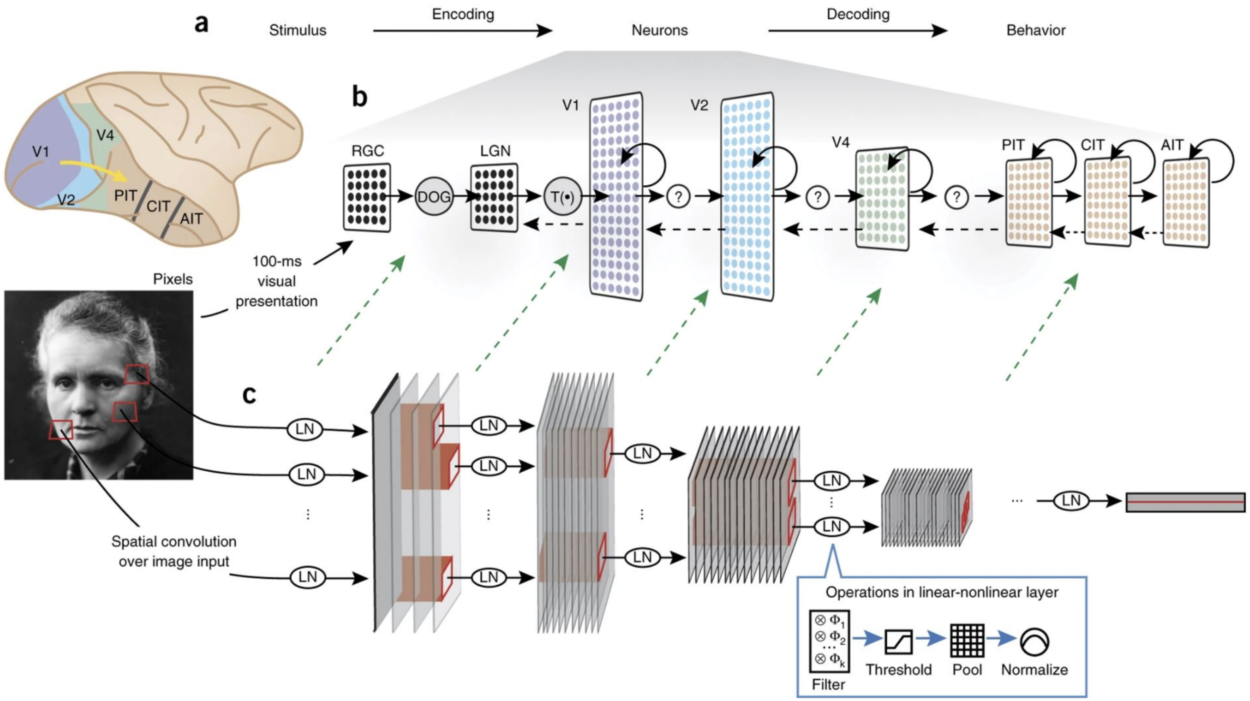
Jane Wang

Submit questions here:

https://app.sli.do/event/92gy6nuo

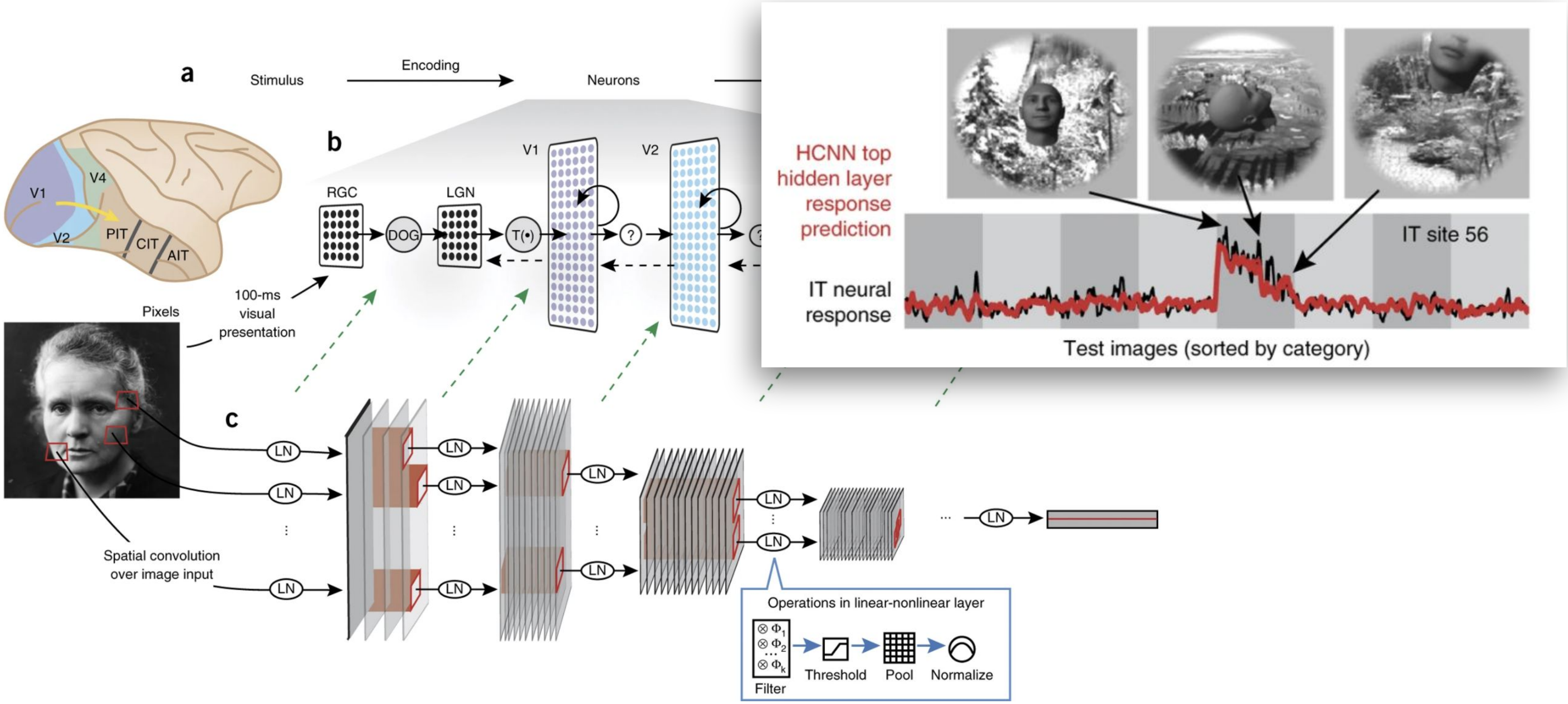

AI/ML
Neuroscience

DNNs as models for neuroscience – perception



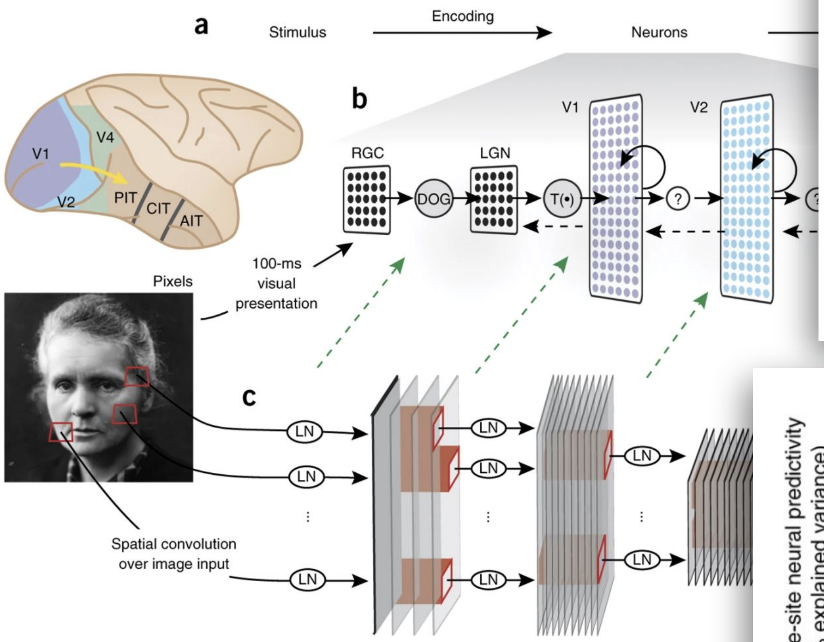
Using goal-driven deep learning models to understand sensory cortex. Yamins & DiCarlo, 2016 Nature Neuroscience
Performance-optimized hierarchical models predict neural responses in higher visual cortex. Yamins et al, 2014 PNAS

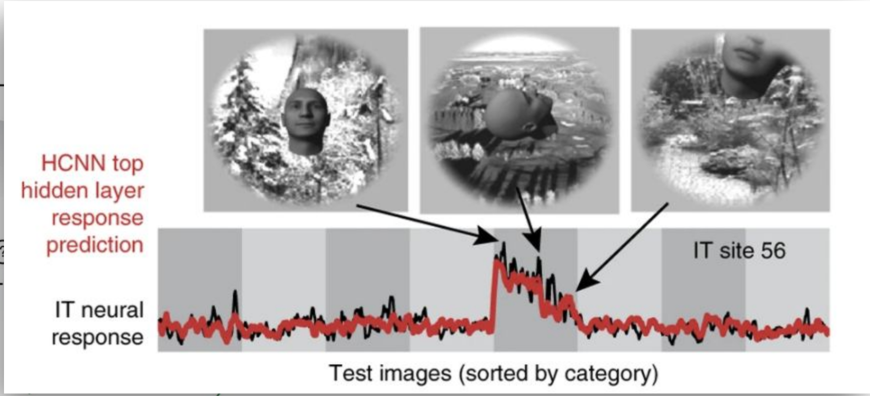
DNNs as models for neuroscience – perception

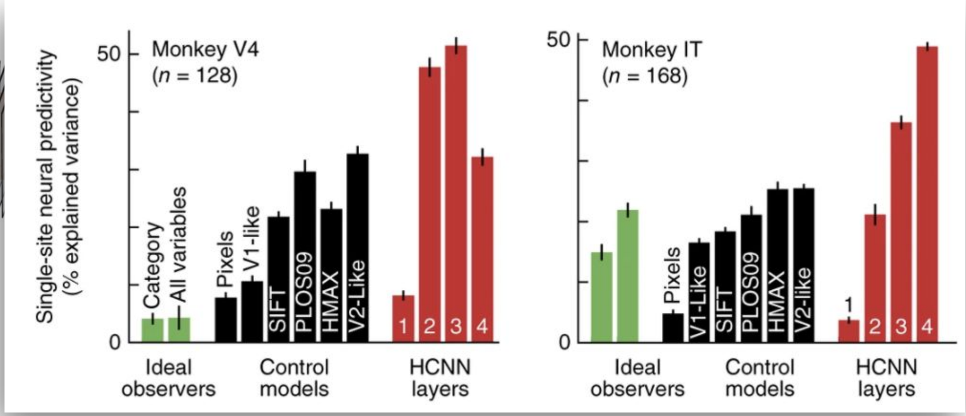


Using goal-driven deep learning models to understand sensory cortex. Yamins & DiCarlo, 2016 Nature Neuroscience
Performance-optimized hierarchical models predict neural responses in higher visual cortex. Yamins et al, 2014 PNAS

DNNs as models for neuroscience – perception



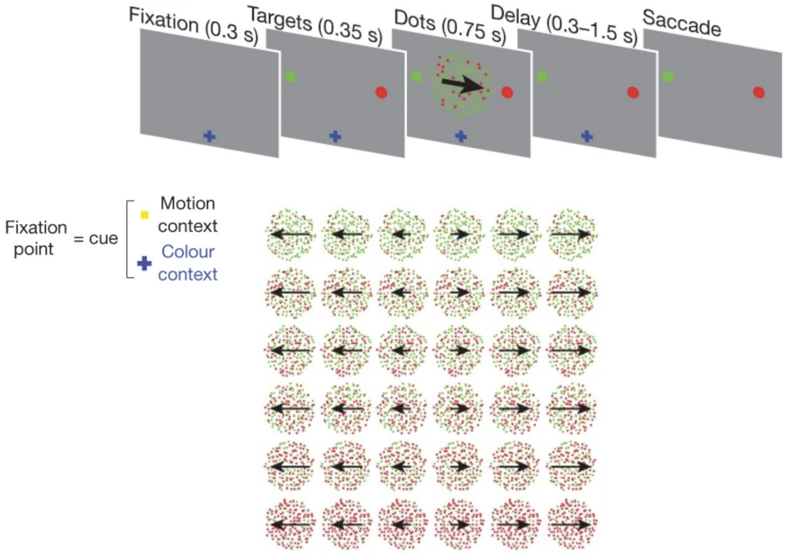


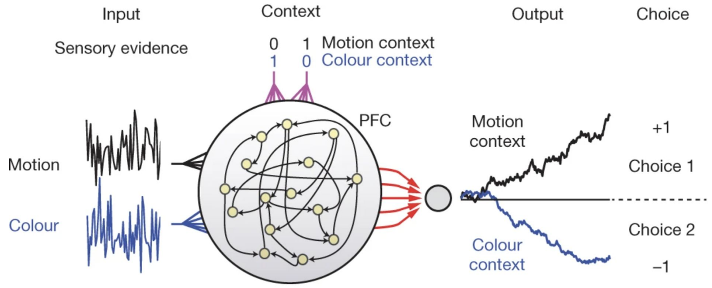


Yamins & DiCarlo, 2016 Nature Neuroscience
Yamins et al, 2014 PNAS

DNNs as models for neuroscience - decision-making

- Perceptual-based decision-making

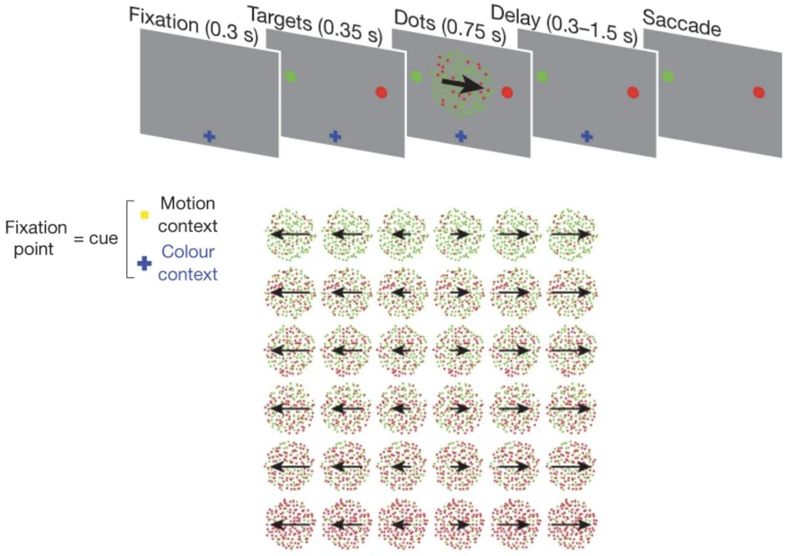


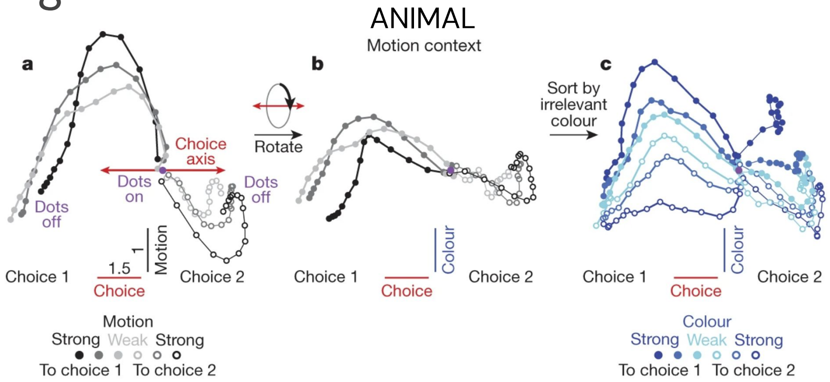


Context-dependent computation by recurrent dynamics in prefrontal cortex. Mante et al. 2013 Nature

DNNs as models for neuroscience – decision-making

- Perceptual-based decision-making

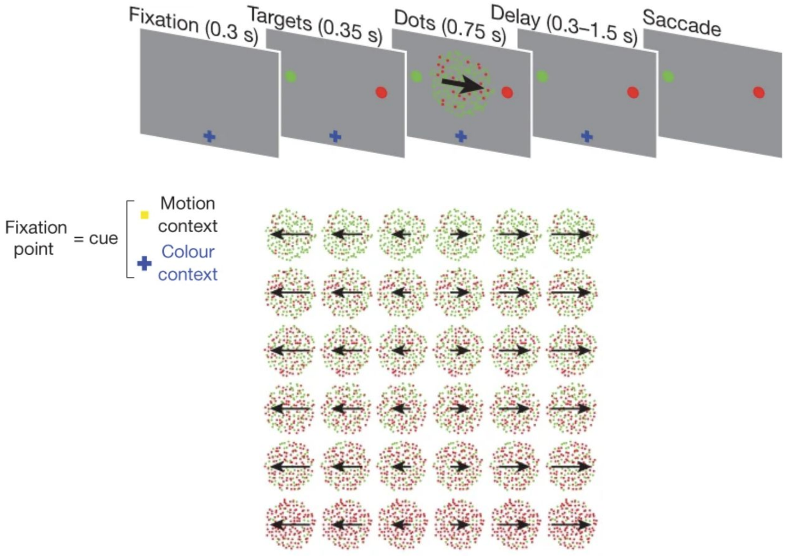


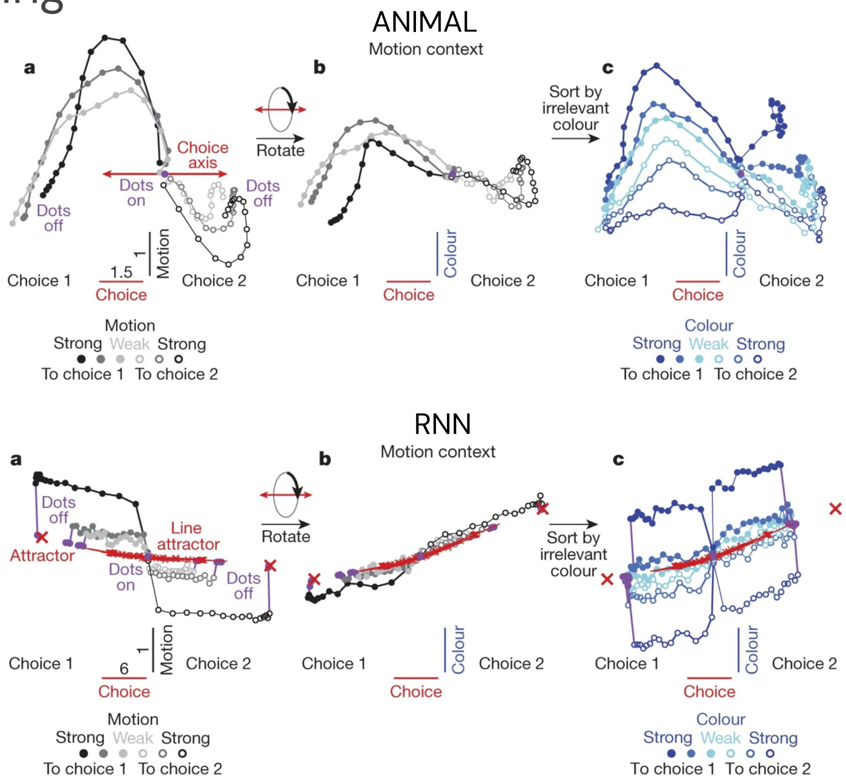


Context-dependent computation by recurrent dynamics in prefrontal cortex. Mante et al. 2013 Nature

DNNs as models for neuroscience - decision-making

- Perceptual-based decision-making



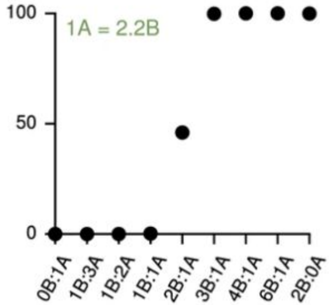


Context-dependent computation by recurrent dynamics in prefrontal cortex. Mante et al. 2013 Nature

DNNs as models for neuroscience – decision-making

- Value-based decision-making

OR

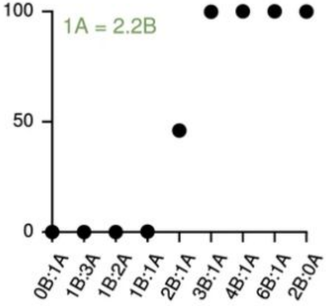


Reward-based training of recurrent neural networks for cognitive and value-based tasks. Song et al. 2017 eLife

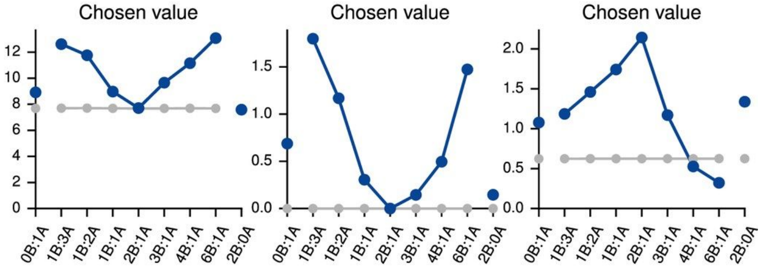
DNNs as models for neuroscience – decision-making

- Value-based decision-making

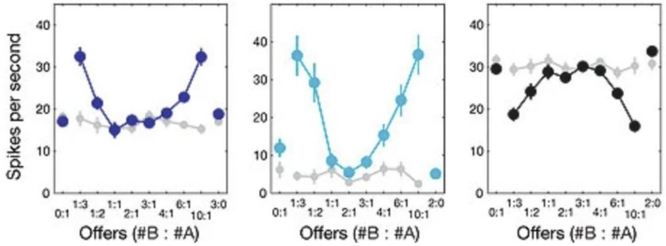
OR



RNN



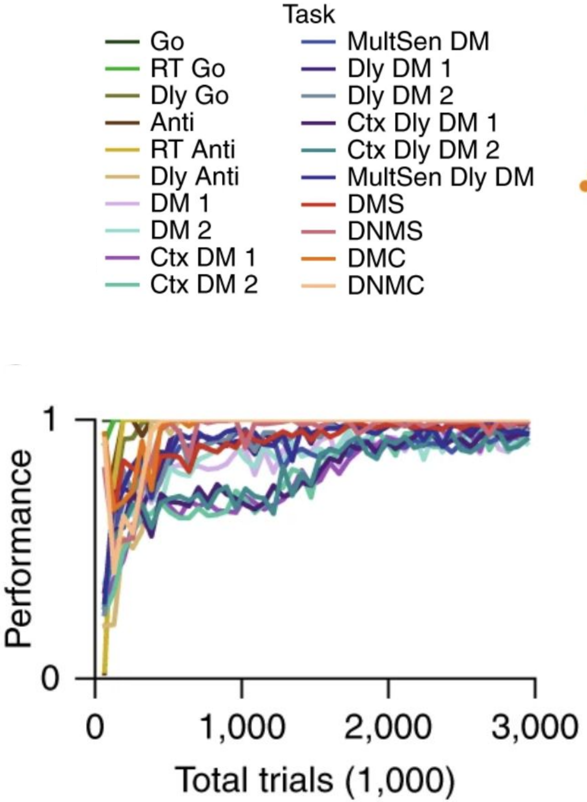
ANIMAL

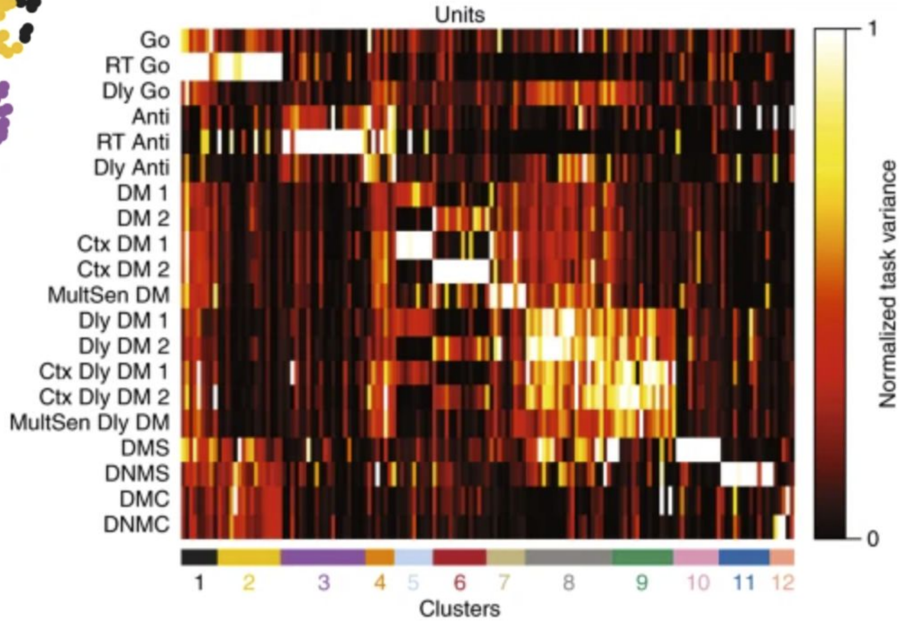


Reward-based training of recurrent neural networks for cognitive and value-based tasks. Song et al. 2017 eLife

Neurons in the orbitofrontal cortex encode economic value. Padoa-Schioppa & Assad. 2006 Nature

DNNs as models for neuroscience – many tasks

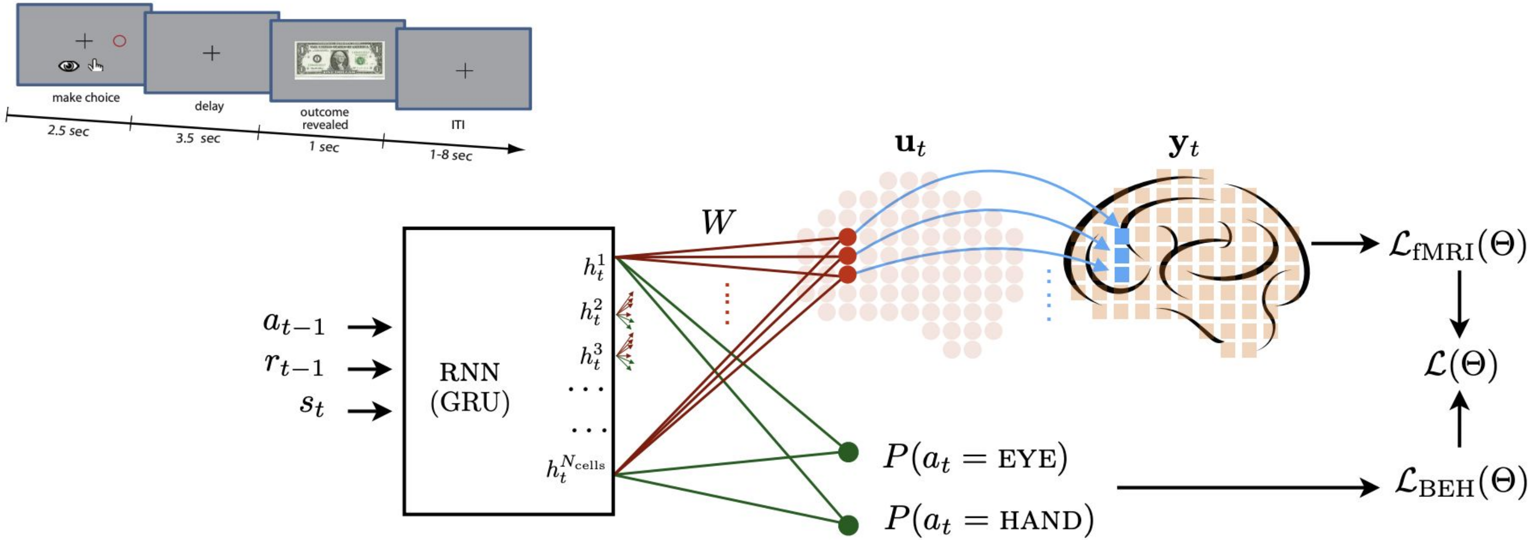




Task representations in neural networks trained to perform many cognitive tasks. Yang et al. 2019 Nat Neuro

DNNs as models for neuroscience – human behavior

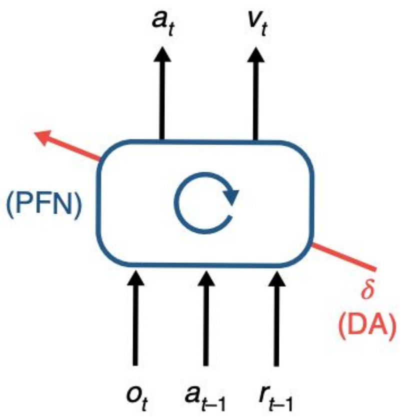
- Train RNN to simultaneously predict behavior and neural response data (fMRI)
- Use this model to assess the impact of reward on future actions, pinpointing specific brain regions involved in decision-making in this task

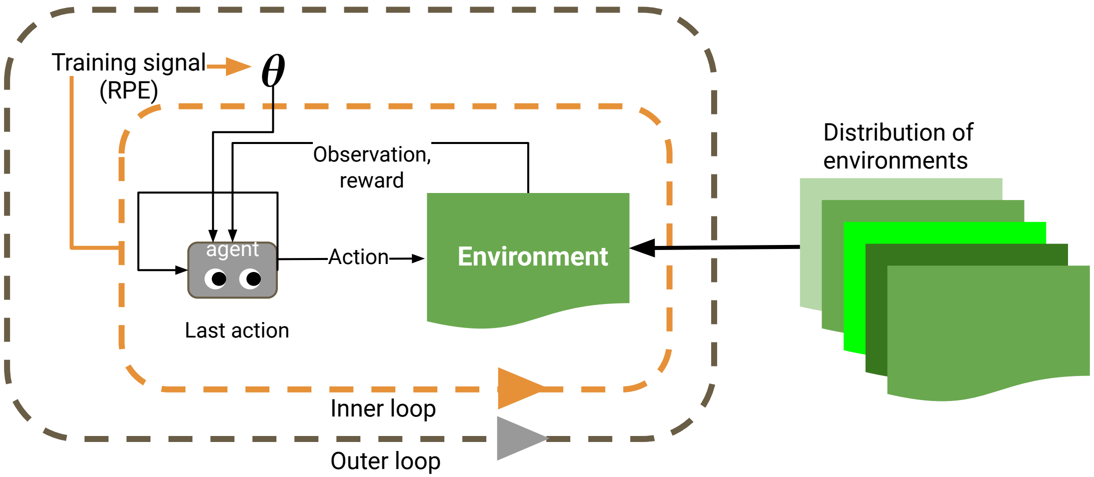


Integrated accounts of behavioral and neuroimaging data using flexible recurrent neural network models.
Dezfouli et al. 2018 NeurIPS

Deep meta-reinforcement learning

- Use an LSTM to meta-learn a reinforcement learning algorithm
- Train on a distribution of related tasks
- Learns to quickly adapts to new tasks

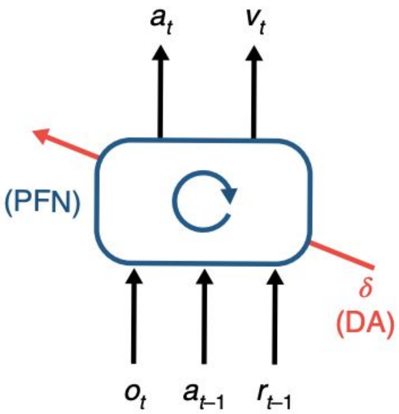


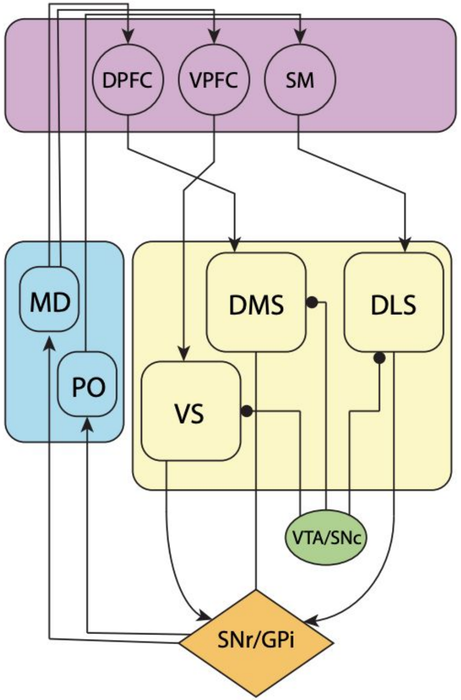


Prefrontal cortex as a meta-reinforcement learning system. Wang et al, 2018, Nat Neurosci

Deep meta-reinforcement learning

- Use an LSTM to meta-learn a reinforcement learning algorithm
- Train on a distribution of related tasks
- Learns to quickly adapts to new tasks





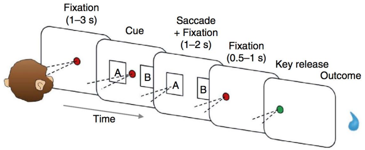
Prefrontal cortex as a meta-reinforcement learning system. Wang et al, 2018, Nat Neurosci

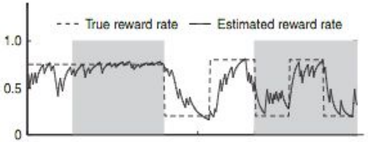

Articles
https://doi.org/10.1038/s41593-018-0147-8
nature neuroscience

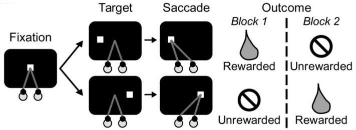
Prefrontal cortex as a meta-reinforcement learning system

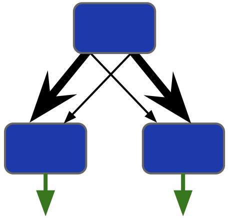
Jane X. Wang[1,5], Zeb Kurth-Nelson[1,2,5], Dharshan Kumaran[1,3], Dhruva Tirumala[1], Hubert Soyer[1], Joel Z. Leibo[1], Demis Hassabis[1,4] and Matthew Botvinick[1,4*]

Over the past 20 years, neuroscience research on reward-based learning has converged on a canonical model, under which the neurotransmitter dopamine 'stamps in' associations between situations, actions and rewards by modulating the strength of synaptic connections between neurons. However, a growing number of recent findings have placed this standard model under



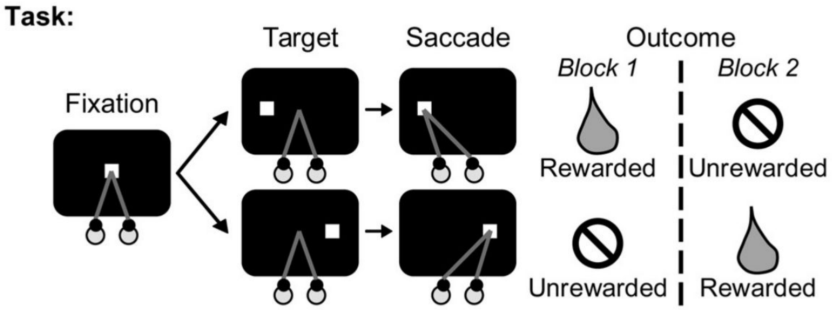






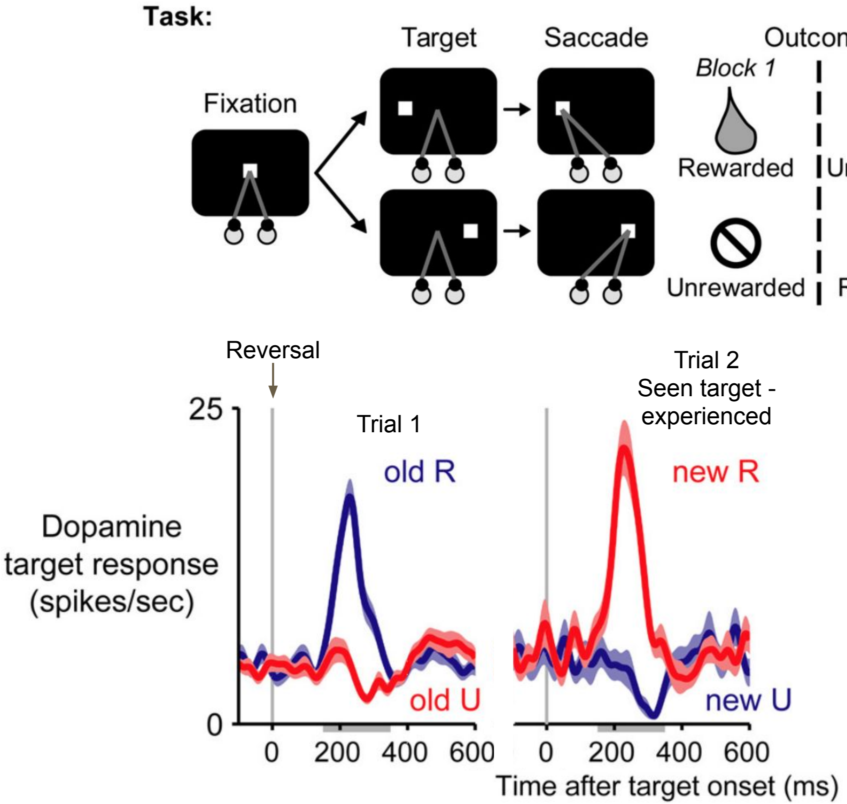
Prefrontal cortex as a meta-reinforcement learning system. Wang et al, 2018, Nat Neurosci

Dopamine reward prediction errors (RPEs) reflect indirect, inferred value



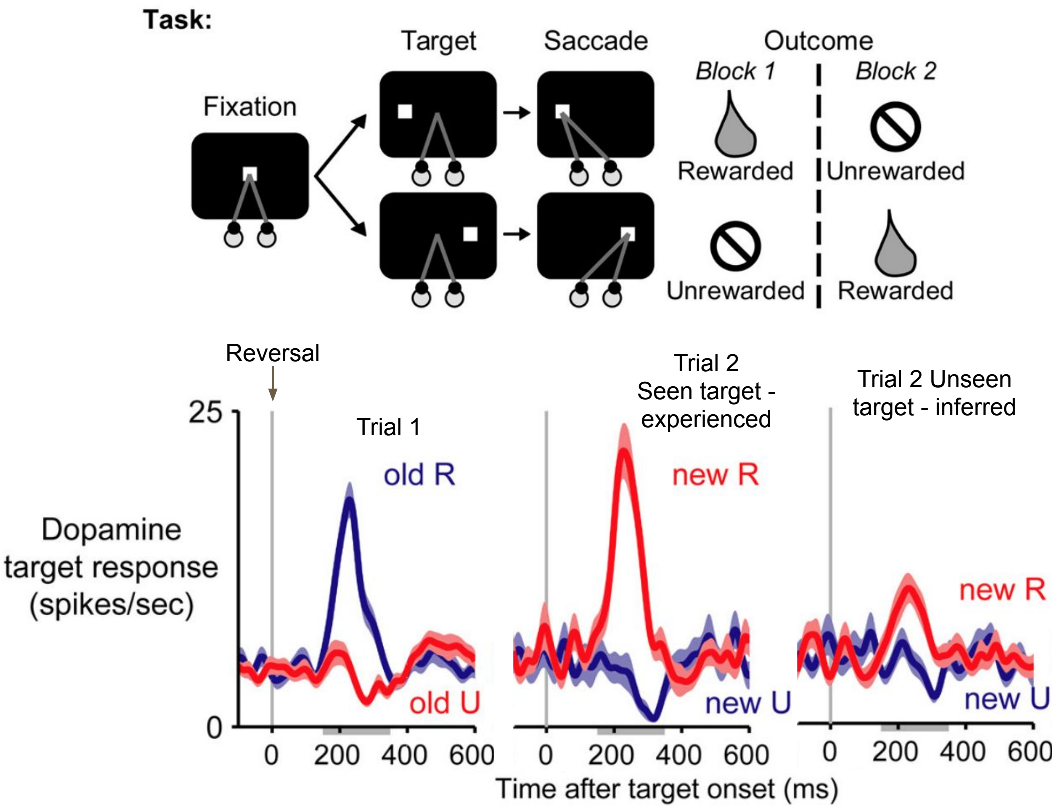
A pallidus-habenula-dopamine pathway signals inferred stimulus values. Bromberg-Martin et al., J Neurophys, 2010

Dopamine reward prediction errors (RPEs) reflect indirect, inferred value



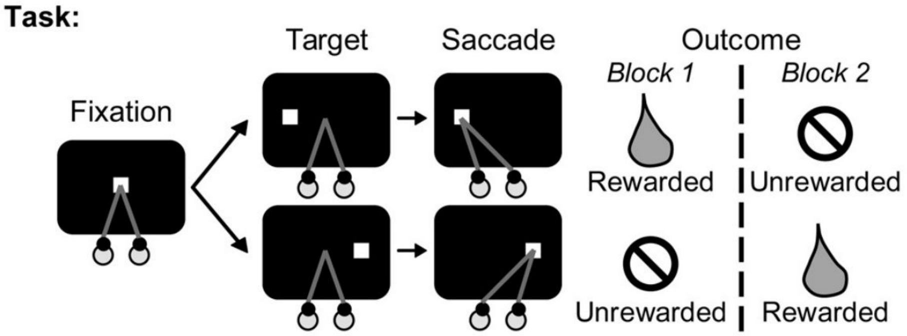
Bromberg-Martin et al., J Neurophys, 2010

Dopamine reward prediction errors (RPEs) reflect indirect, inferred value

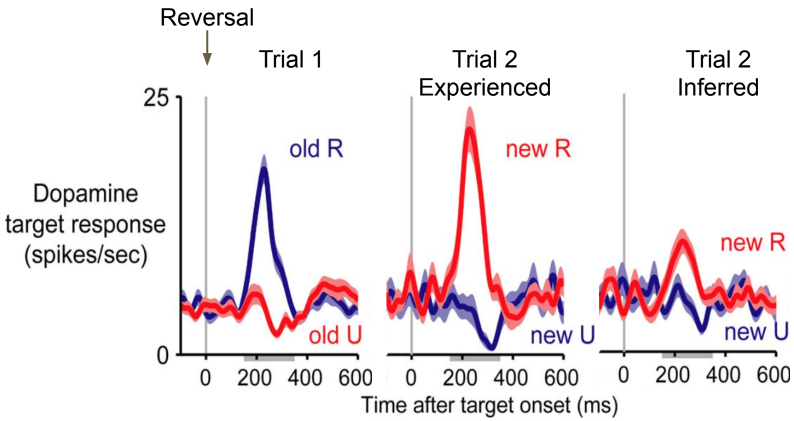


Bromberg-Martin et al., J Neurophys, 2010

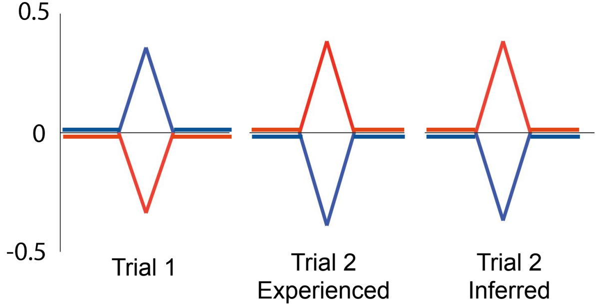
Reward prediction error signal reflects model-based inference



Bromberg-Martin et al, J Neurophys, 2010

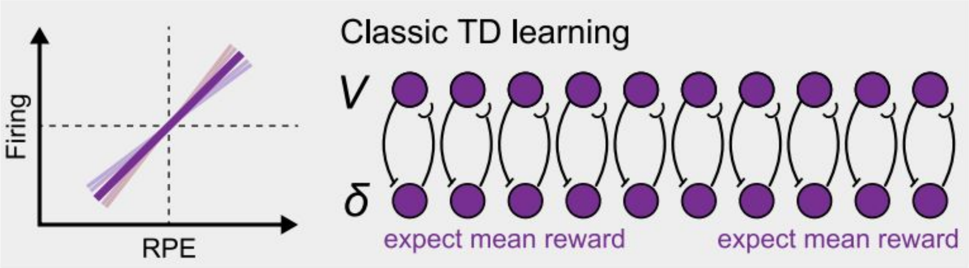


Meta-RL

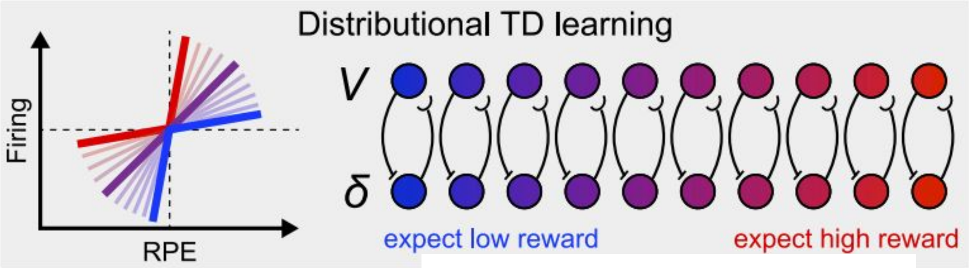


Wang et al, 2018, Nat Neurosci

Distributional reinforcement learning



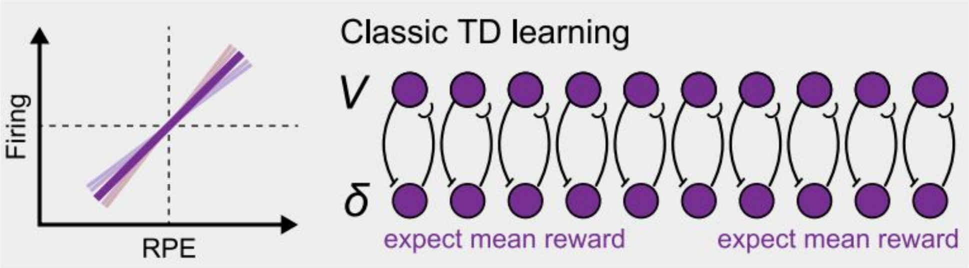
$$V_i(t) \leftarrow V_i(t) + \alpha_i \delta_i(t)$$



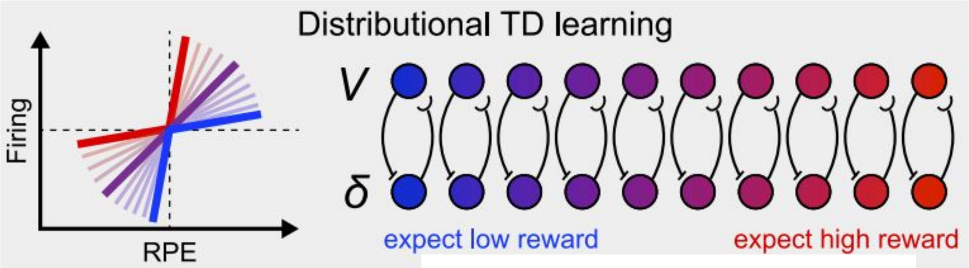
$$V_i(t) \leftarrow V_i(t) + \alpha_i^+ \delta_i(t), \;\; \delta_i(t) > 0$$
$$V_i(t) \leftarrow V_i(t) + \alpha_i^- \delta_i(t), \;\; \delta_i(t) \leq 0$$

A distributional code for value in dopamine-based reinforcement learning. Dabney et al. 2020 Nature

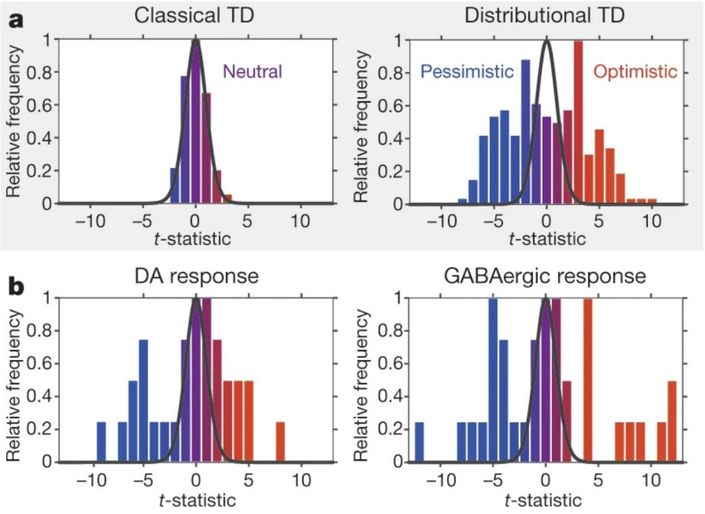
Distributional reinforcement learning



$$V_i(t) \leftarrow V_i(t) + \alpha_i \delta_i(t)$$



$$V_i(t) \leftarrow V_i(t) + \alpha_i^{+} \delta_i(t), \;\; \delta_i(t) > 0$$
$$V_i(t) \leftarrow V_i(t) + \alpha_i^{-} \delta_i(t), \;\; \delta_i(t) \leq 0$$



A distributional code for value in dopamine-based reinforcement learning. Dabney et al. 2020 Nature

Moving towards biological realism

Deep neural networks

★ Discrete time

Biological neural networks

★ Continuous time

Moving towards biological realism

Deep neural networks

★ Discrete time

★ Continuous activations

Biological neural networks

★ Continuous time

★ Spiking, stochastic

Moving towards biological realism

Deep neural networks

- ★ Discrete time
- ★ Continuous activations
- ★ "Supervised" / global loss signal

Biological neural networks

- ★ Continuous time
- ★ Spiking, stochastic
- ★ Associative (Hebbian) / local learning

Moving towards biological realism

Deep neural networks

★ Discrete time

★ Continuous activations

★ "Supervised" / global loss signal

★ Backpropagation for optimization

Biological neural networks

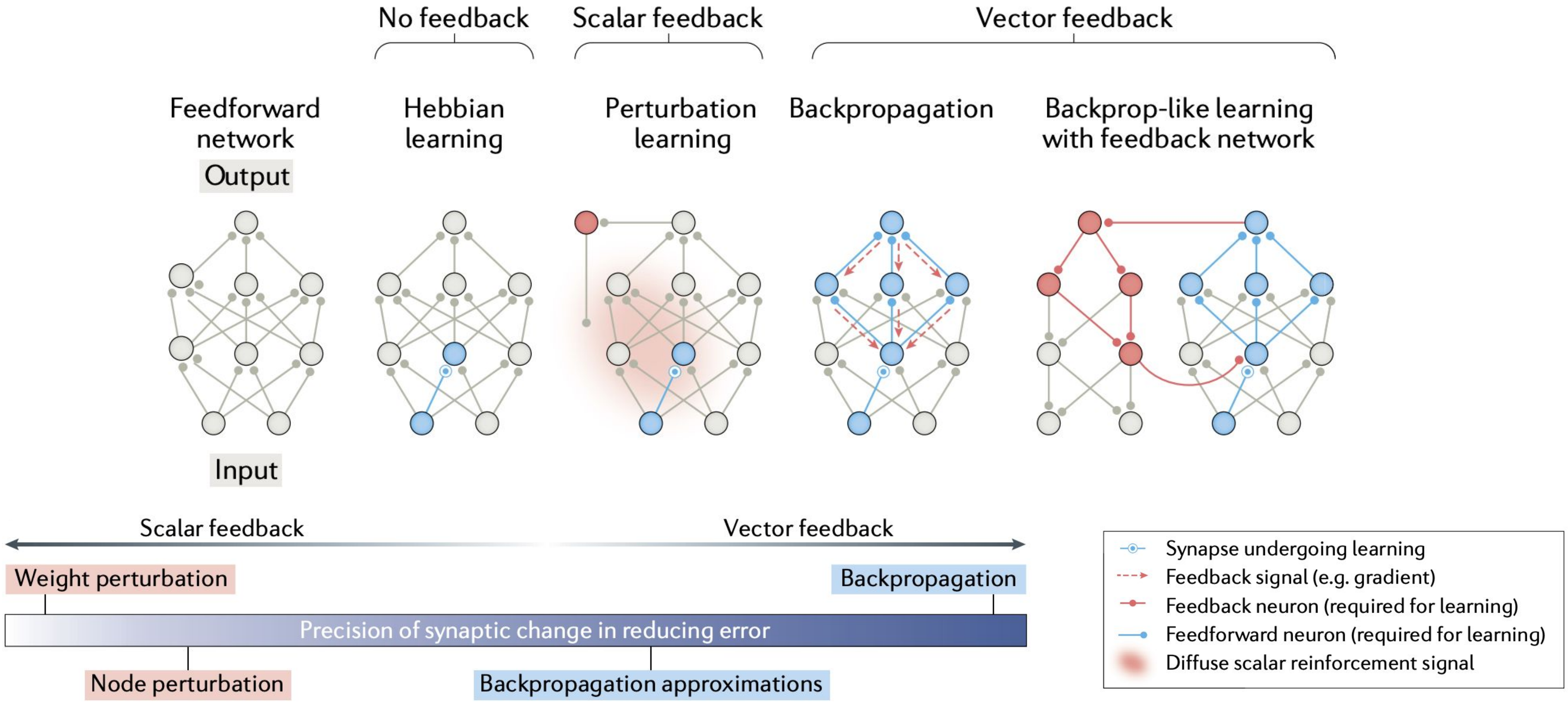
★ Continuous time

★ Spiking, stochastic

★ Associative (Hebbian) / local learning

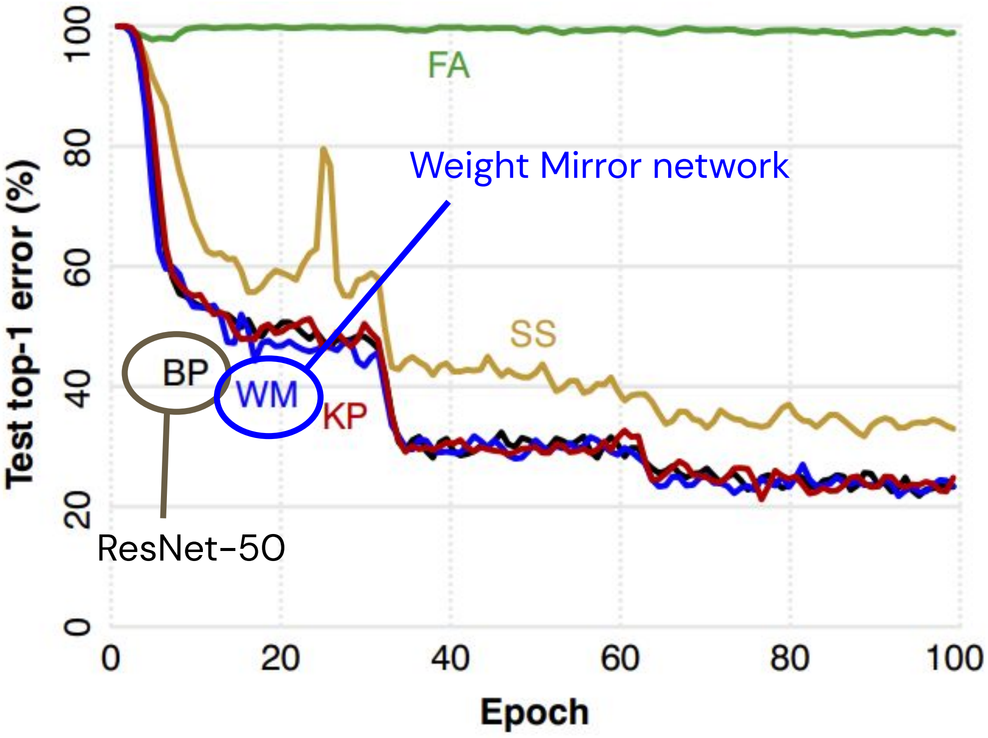
★ No backpropagation!

Moving towards biological realism



Backpropagation and the brain. Lillicrap et al. 2020 Nature Rev Neurosci

Moving towards biological realism



Deep learning without weight transport. Akrout et al. 2019 NeurIPS

Moving towards biological realism

Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks

Thomas Miconi
The Neurosciences Institute, United States

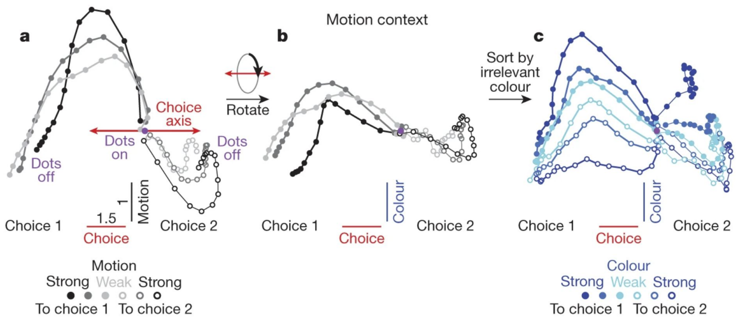
- Biologically plausible implementation of (continuous time) RNN trained to perform multiple cognitive tasks
  - Reward-modulated Hebbian variant of node perturbation
  - No backprop required

Biologically realistic implementation of RNNs. Miconi. 2017, eLife

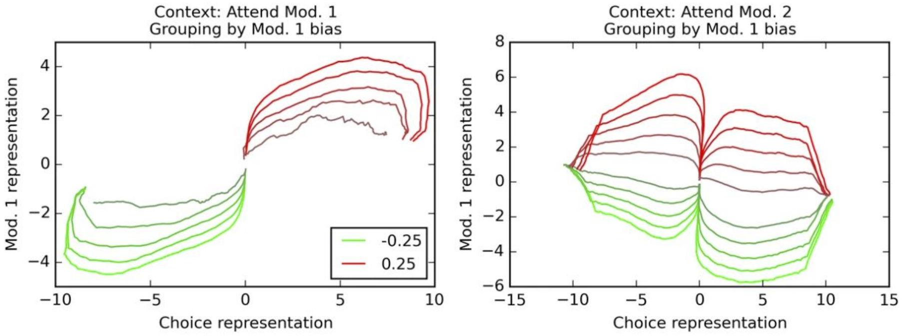
Moving towards biological realism

Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks

Thomas Miconi
The Neurosciences Institute, United States



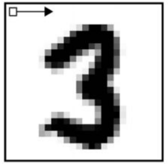
Mante et al. 2013 Nature

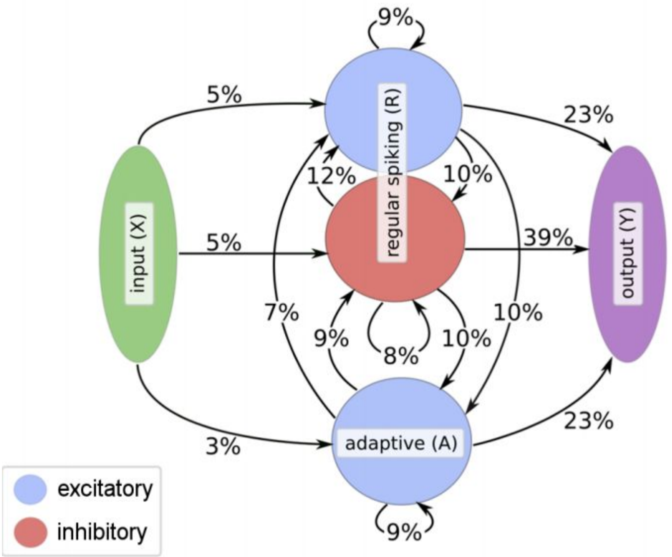


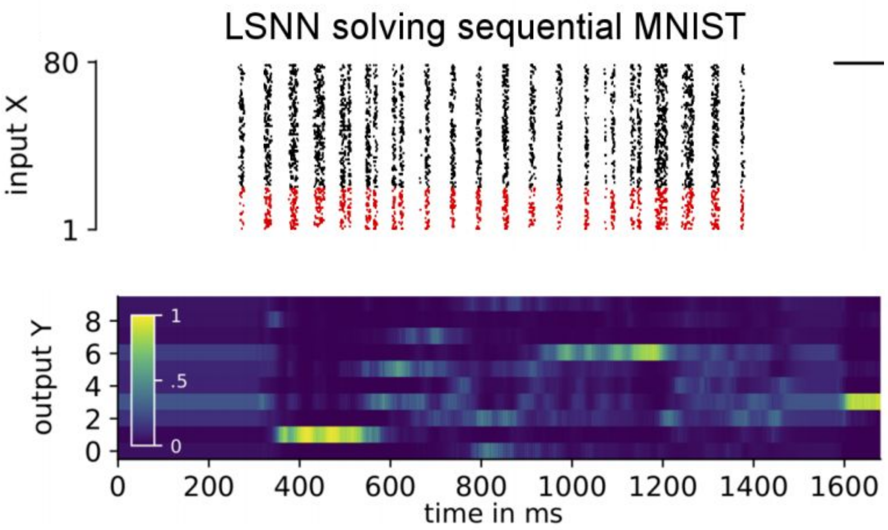
Miconi. 2017, eLife

Moving towards biological realism

- Learning with spiking neural networks
- Sequential MNIST task



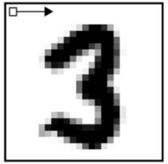


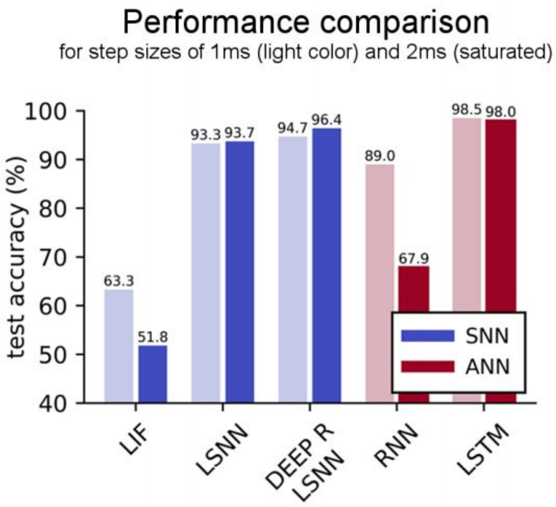


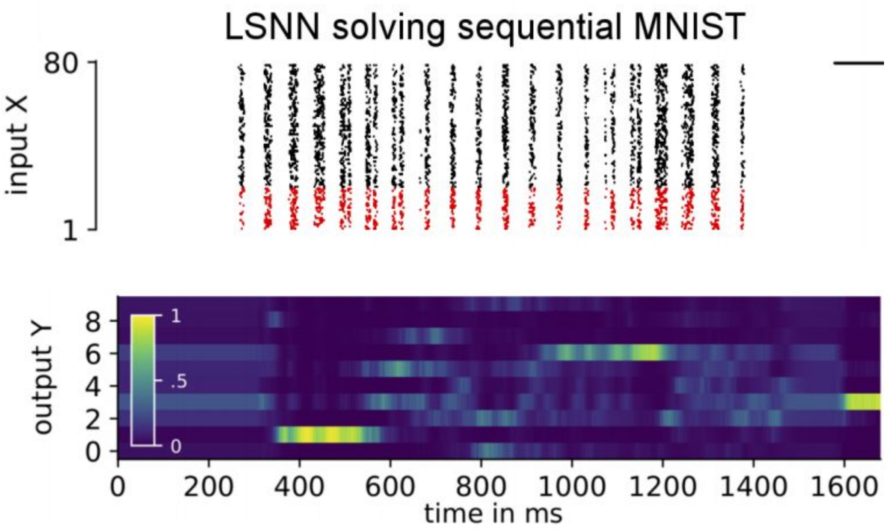
Long short-term memory and learning-to-learn in networks of spiking neurons. Bellec et al. 2018 NeurIPS

Moving towards biological realism

- Learning with spiking neural networks
- Sequential MNIST task



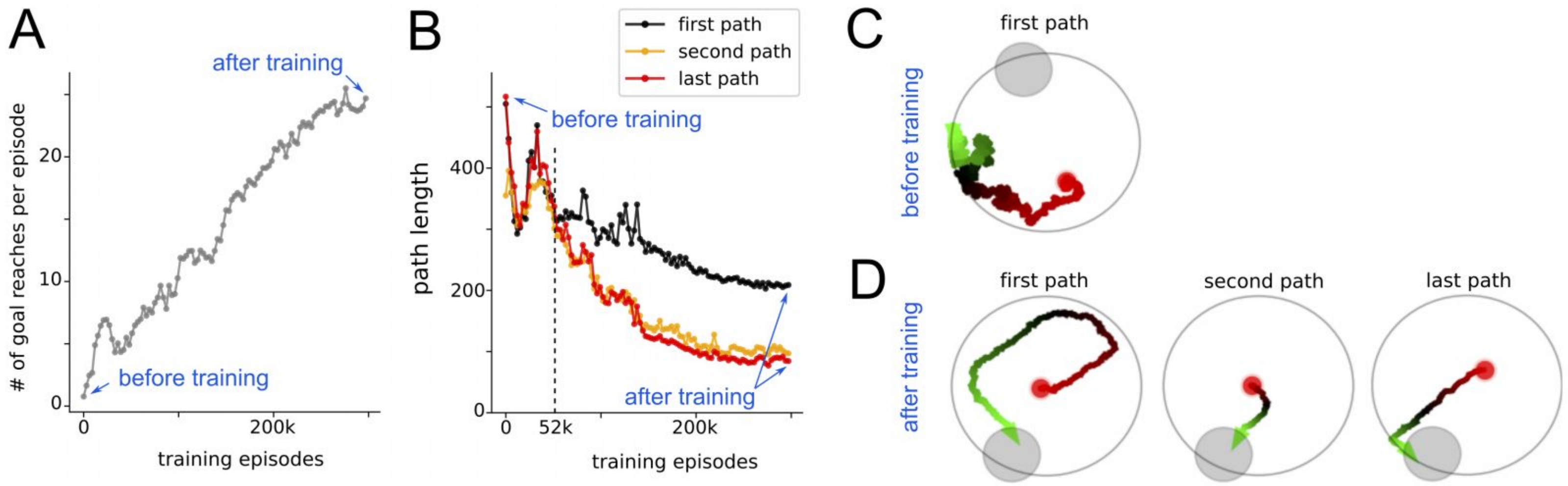




Long short-term memory and learning-to-learn in networks of spiking neurons. Bellec et al. 2018 NeurIPS

Moving towards biological realism

- Learning to learn from reward with spiking neural networks
- Morris water maze task



Long short-term memory and learning-to-learn in networks of spiking neurons. Bellec et al. 2018 NeurIPS

Why move toward biological realism?

- To get out of local optimums

Why move toward biological realism?

- To get out of local optimums
- To try to emulate what the brain and biology does best: solve problems under uncertainty, finite computation, decomposable situations, and structured environments

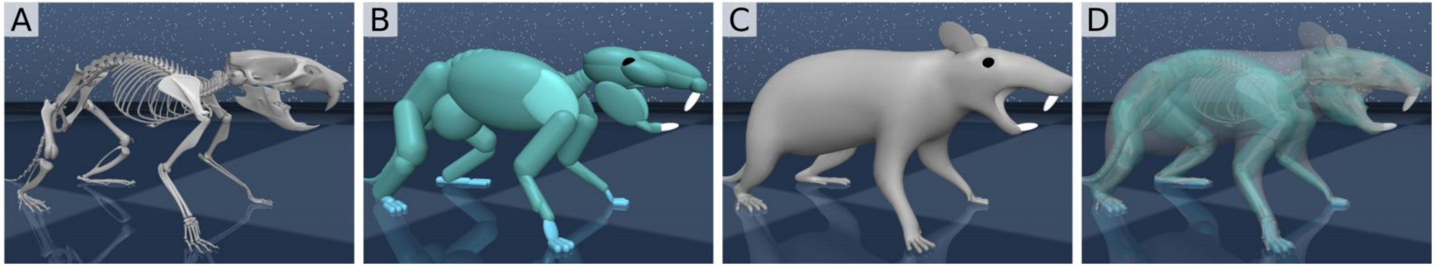
Why move toward biological realism?

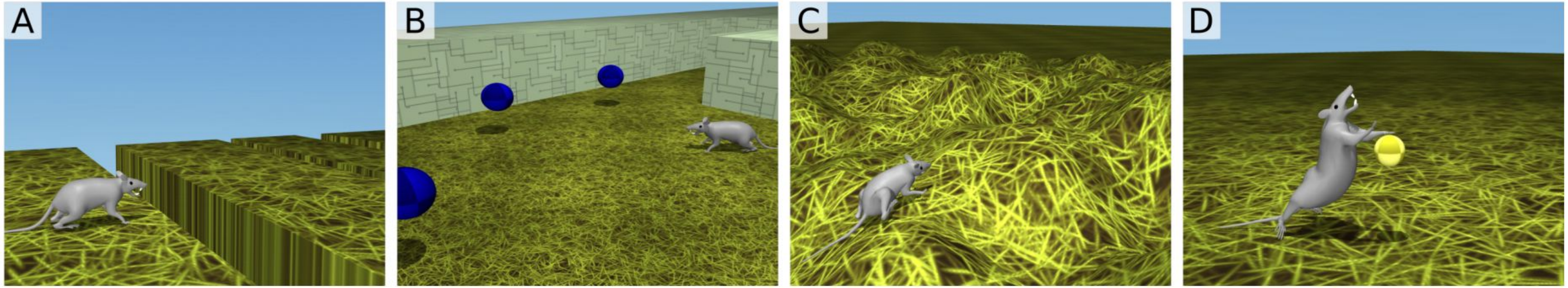
- To get out of local optimums
- To try to emulate what the brain and biology does best: solve problems under uncertainty, finite computation, decomposable situations, and structured environments
- To be able to solve more real-world problems

Why move toward biological realism?

- To get out of local optimums
- To try to emulate what the brain and biology does best: solve problems under uncertainty, finite computation, decomposable situations, and structured environments
- To be able to solve more real-world problems
- To get additional clues about what problems biology is trying to solve

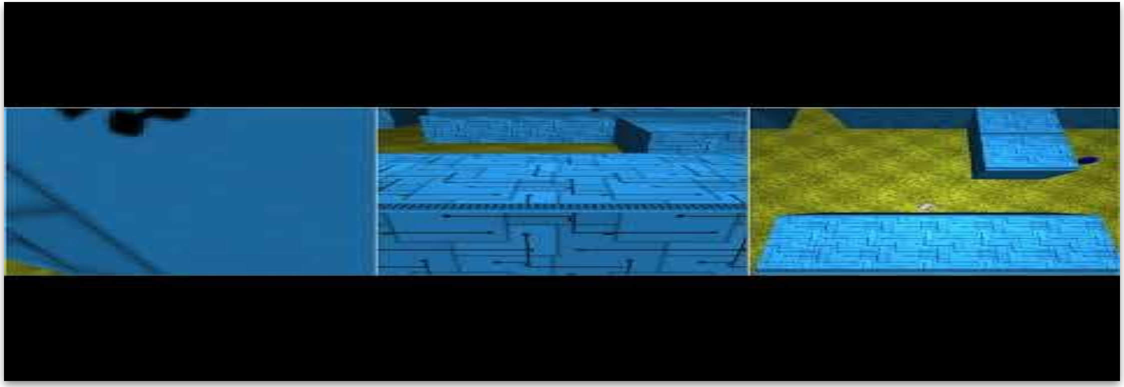
Artificial models of embodied control

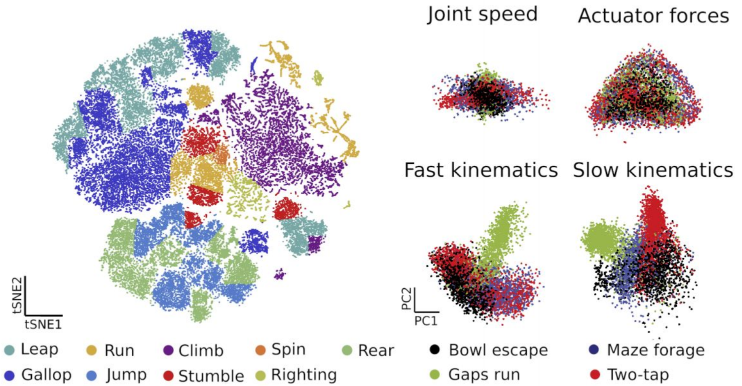




Deep neuroethology of a virtual rodent. Merel et al. 2020 ICLR

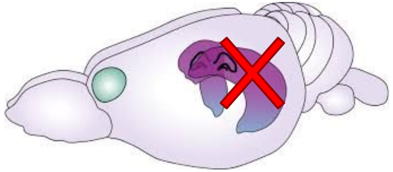
Artificial models of embodied control

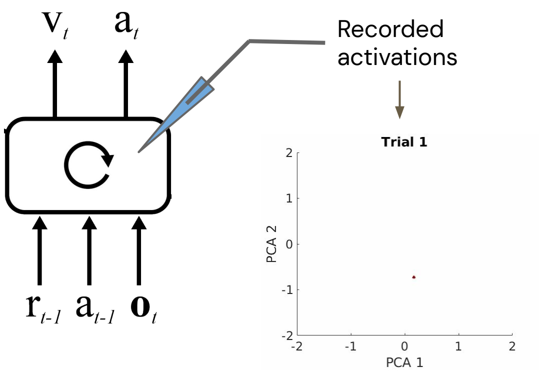


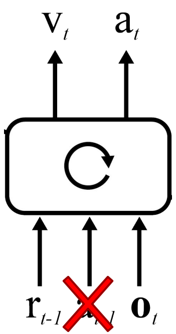


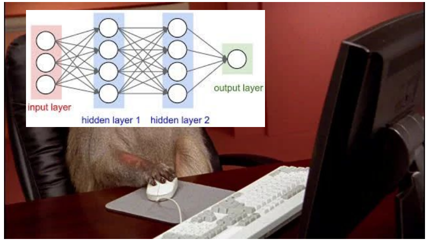
Deep neuroethology of a virtual rodent. Merel et al. 2020 ICLR

Understanding DNNs the way we understand brains



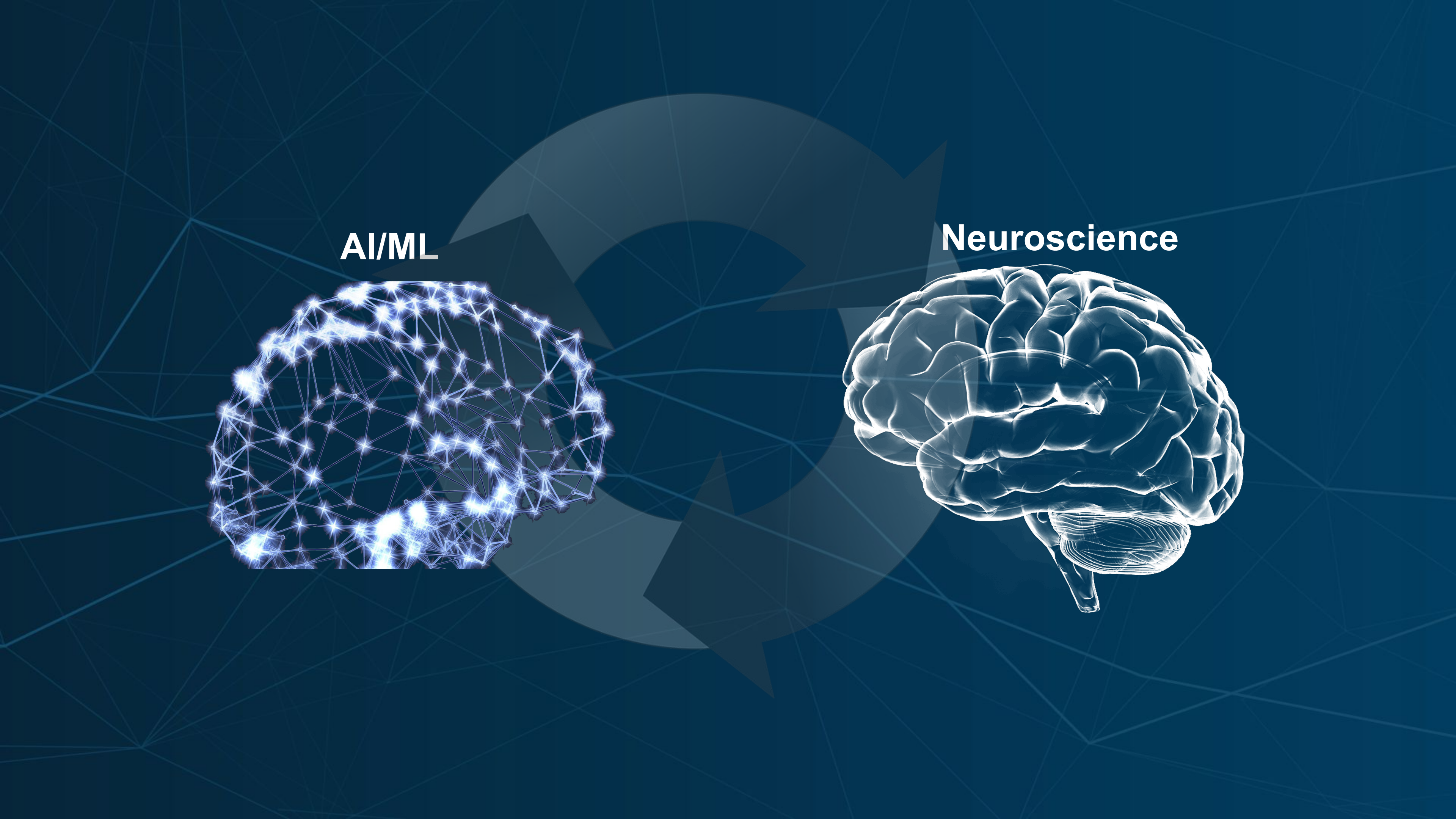


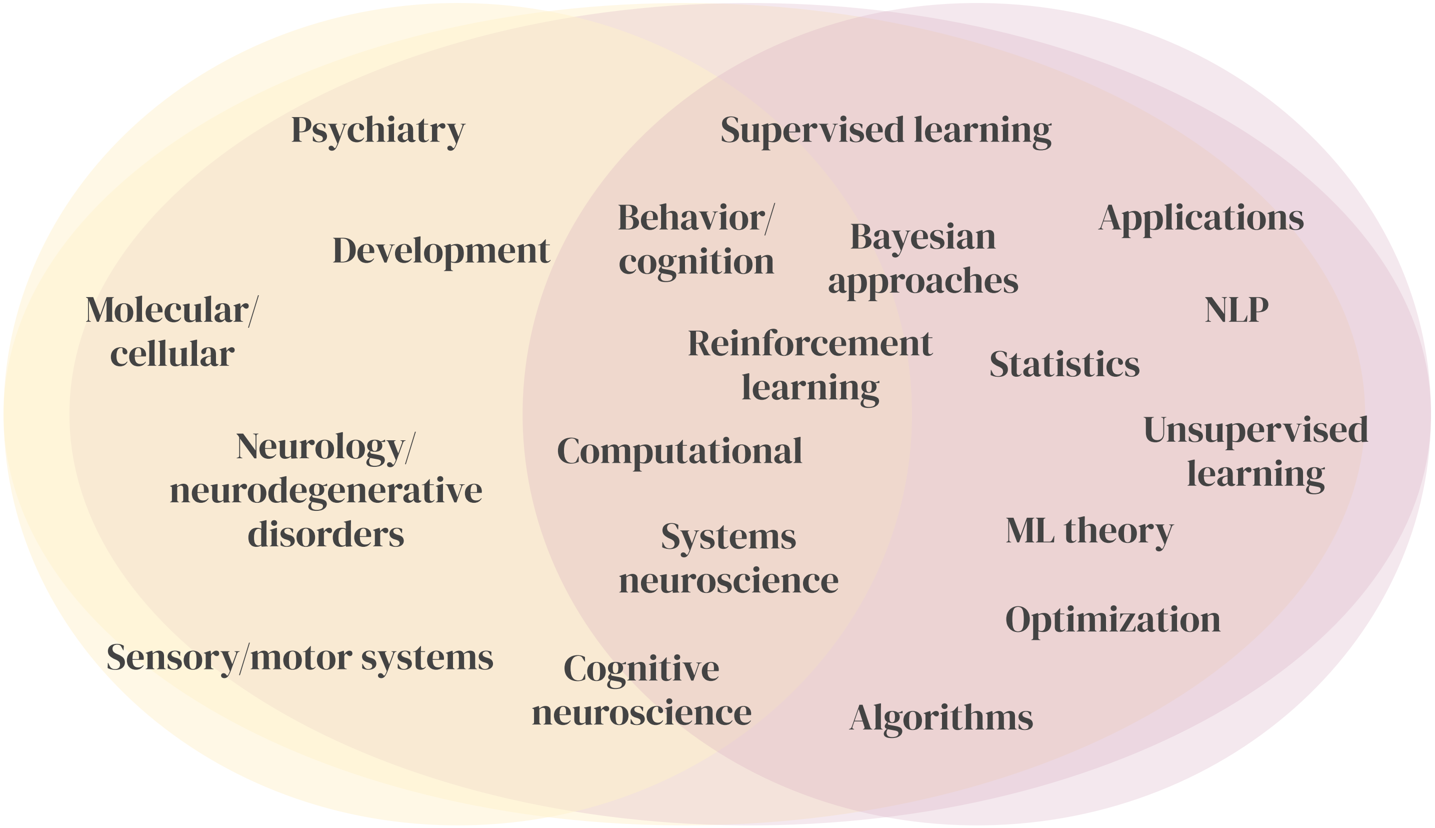




Understanding DNNs the way we understand brains

1. “Visualizing and understanding atari agents.” Greydanus et al, 2018 ICML
2. “Analyzing biological and artificial neural networks: challenges with opportunities for synergy?” Barrett et al. 2019 Curr Opin Neurobiol
3. “On the importance of single directions for generalization.” Morcos et al, 2018 ICLR
4. “Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability.” Raghu et al, 2017 NeurIPS
5. “Explain Your Move: Understanding Agent Actions Using Specific and Relevant Feature Attribution.” Gupta et al. 2020 ICLR
6. “Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks.” Sussillo et al, 2013
7. “Universality and individuality in neural dynamics across large populations of recurrent networks.” Maheswaranathan et al, 2019 NeurIPS


AI/ML
Neuroscience


Psychiatry
Supervised learning
Development
Behavior/
cognition
Bayesian
approaches
Applications
Molecular/
cellular
NLP
Reinforcement
learning
Statistics
Neurology/
neurodegenerative
disorders
Computational
Unsupervised
learning
ML theory
Systems
neuroscience
Optimization
Sensory/motor systems
Cognitive
neuroscience
Algorithms

Thanks for your attention!

Website: https://sites.google.com/view/neurips-2020-tutorial-neurosci/home

Submit questions: https://app.sli.do/event/92gy6nuo

DeepMind

UCL