Nonlinear neural simulation of error-driven learning

Sarah Oh*, Jennifer Senta* & Milena Rmus*

equal contributions

Motivation & background

The relationship between formal reinforcement learning (RL) algorithms and neural mechanisms of learning (e.g. dopaminergic signaling of VTA neurons)[1] has established RL as an important connection between computer science and neuroscience.

However, many neural network (NN) models of RL leverage optimization algorithms such as gradient descent and backpropagation, which are not biologically plausible[2,3].

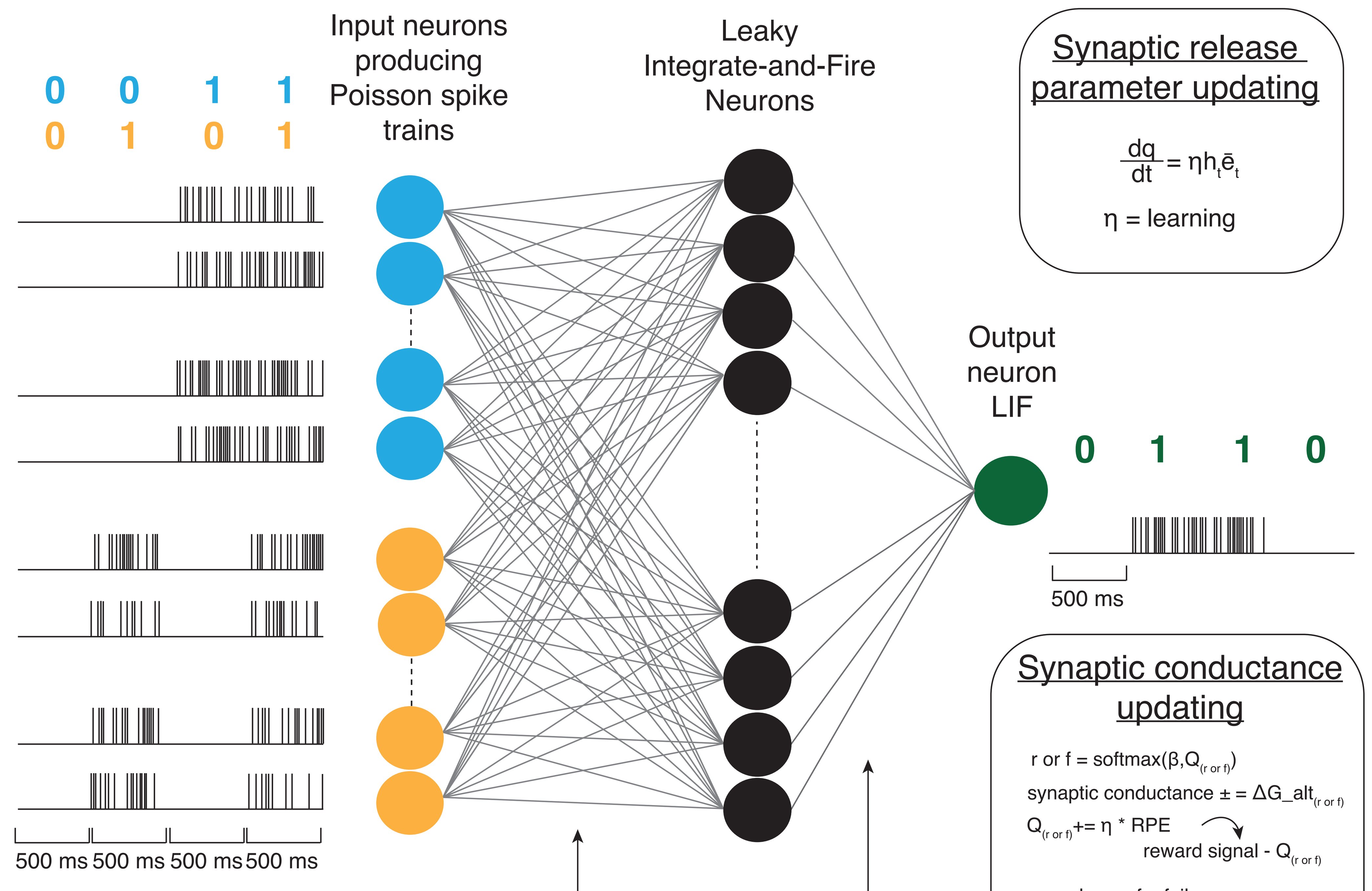
We attempt to use an alternative, more biologically realistic spiking neural network combined with synaptic plasticity to capture reward based learning on a commonly used training benchmark[4,5].

Benchmark

Exclusive-or (XOR) problem

Input	Output
(0,0)	0
(0,1)	1
(1,0)	1
(1,1)	0

Neural network structure



Network figure adapted from Seung (2003)

Synaptic release parameter updating

$$\frac{dq}{dt} = \eta h_t \bar{e}_t$$

η = learning

Synaptic conductance updating

r or f = softmax(β, $Q_{(r\ or\ f)}$)

synaptic conductance ± = ΔG_alt$_{(r\ or\ f)}$

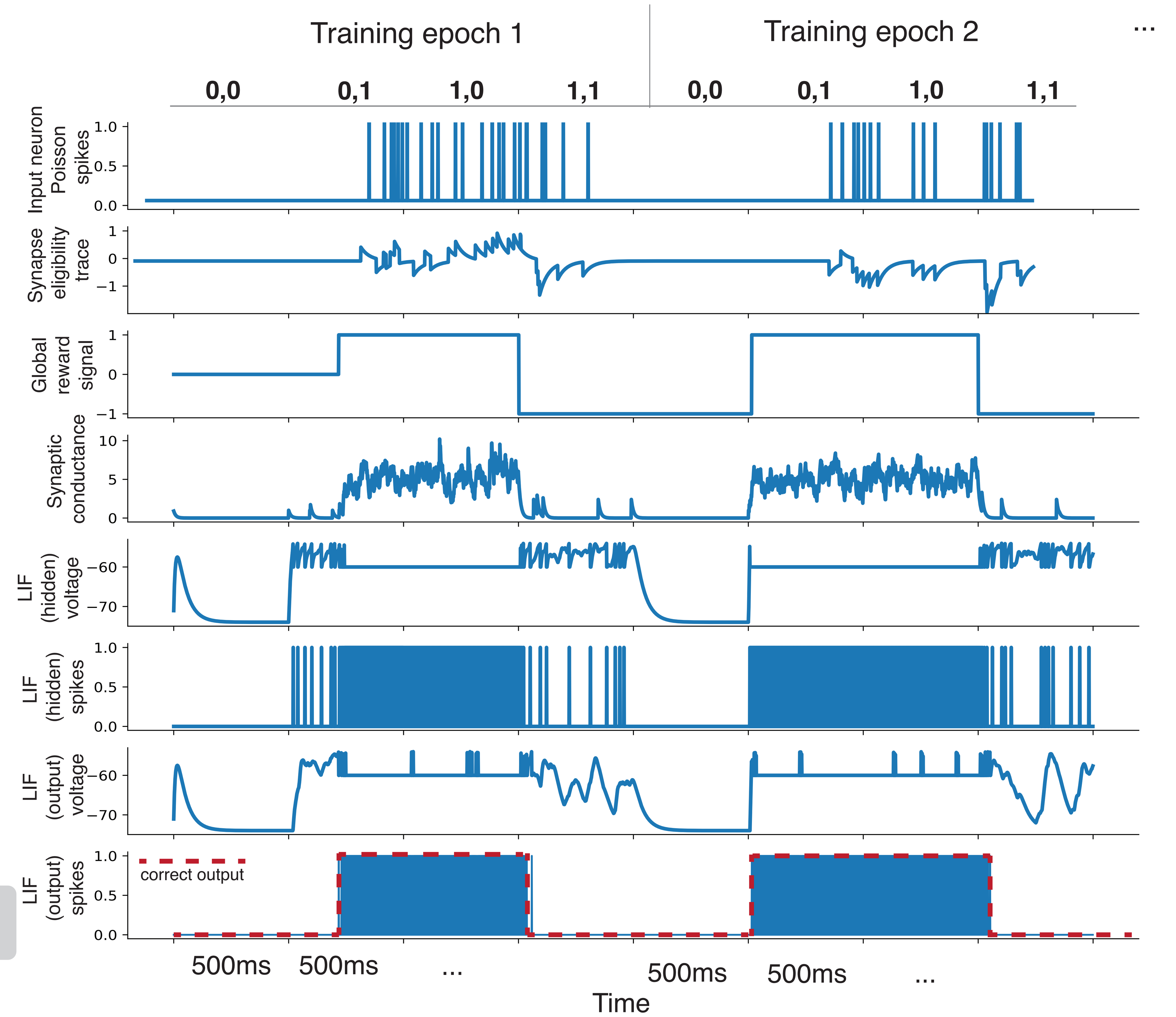
$Q_{(r\ or\ f)}$ += η * RPE

reward signal - $Q_{(r\ or\ f)}$

r = release; f = failure

G_alt is a hyperparameter

Results



Dynamics

Input layer: Poisson spike train neurons

$p_{spike} = \delta_t$ * rate; rate = 40Hz

$$\text{spike} = \begin{cases} \text{binomial}(1, p_{spike}) & \text{if } 1, \\ 0 & \text{if } 0 \end{cases}$$

Hidden layer: Leaky integrate-and-fire neurons

$$C\frac{dV_i}{dt} = -g_L(V_i - V_L) - \Sigma G_{i,j}(V_i - V_{i,j}) + I_{tonic}$$

Parameters

V_L = -74 mV $\quad$ V_θ = -54 mV

g_L = 25 nS $\quad$ V_{reset} = -60 mV

C = 500 pF $\quad$ refractory time = 0.001 s

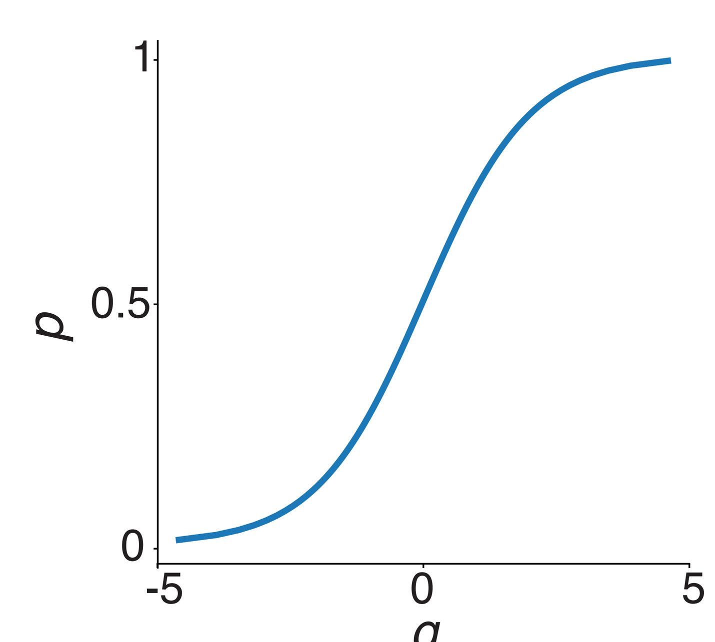
η = 0.3

$I_{tonic} \sim N(\mu = 425\ \text{pA}, \sigma = 200\ \text{pA})$

Synapse

Release probability (p) $\quad p = \dfrac{1}{1 + e^{-q}}$

Release parameter (q) $\quad q = \log \dfrac{p}{1 - p}$



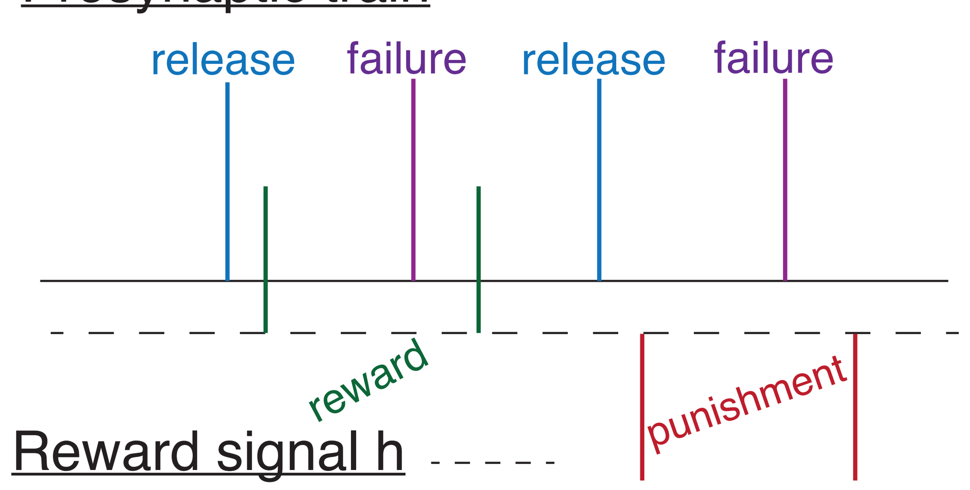
Spikes:

$$\Delta\bar{e} = \begin{cases} 1-p & \text{if release,} \\ -p & \text{if failure} \end{cases}$$

No spikes (exponential decay):

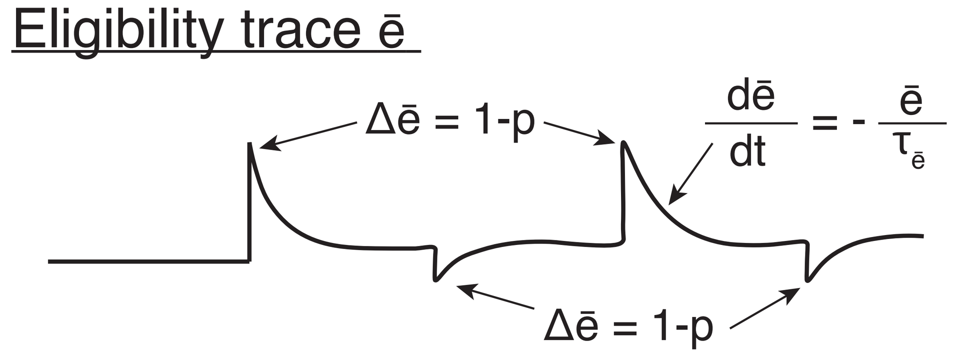
$$\frac{d\bar{e}}{dt} = -\frac{\bar{e}}{\tau_{\bar{e}}}$$

Presynaptic train



Reward signal h

Eligibility trace ē



Conclusions

Hedonistic learning at synapse conductance level provides a biologically realistic multilevel neural network model capable of learning XOR function

Eligibility trace for each synaptic connection stores cumulative memory of recent actions by synapse

Future directions

Consider inhibitory vs excitatory inputs to network from distinct brain regions to expand similarity to biological system

Implement alternative neural model instead of leaky-integrate-and-fire, such as more complex Hodgkins-Huxley neurons

References

1.Schultz W, Dayan P, Montague PR (1997). A neural substrate of prediction and reward. Science 275: 1593–1599.

2.Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533-536.

3. Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

4. Seung, H. S. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. Neuron, 40(6), 1063-1073.

5. Vasilaki, E., Frémaux, N., Urbanczik, R., Senn, W., & Gerstner, W. (2009). Spike-based reinforcement learning in continuous state and action space: when policy gradient methods fail. PLoS computational biology, 5(12), e1000586.