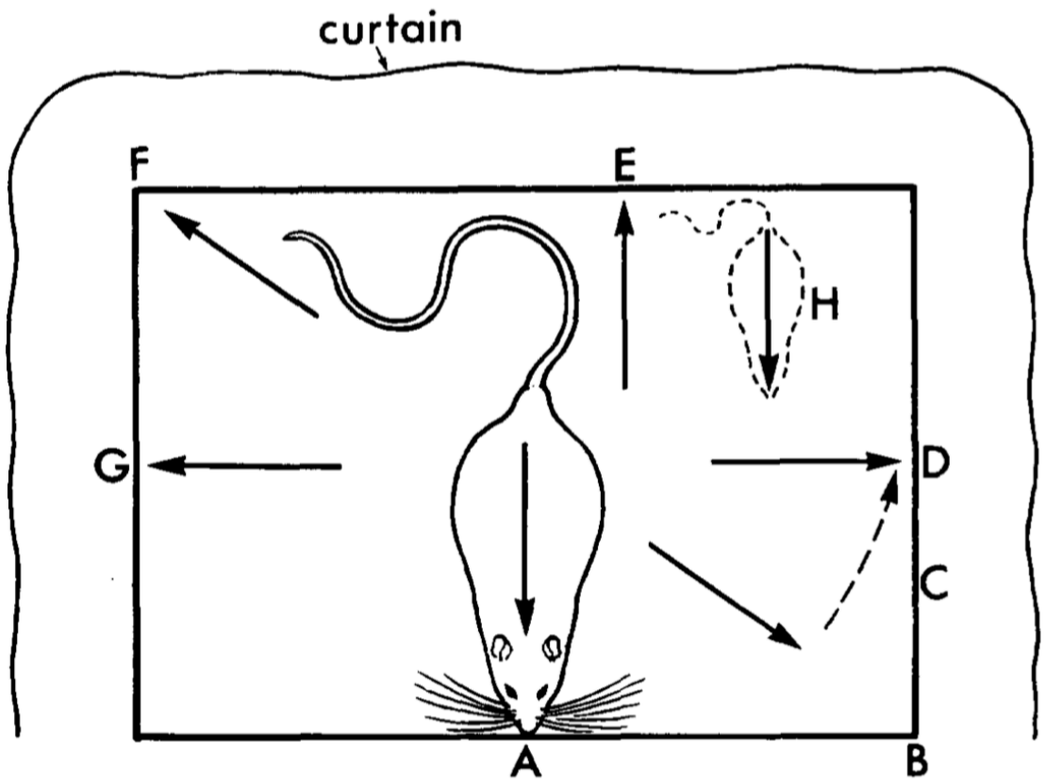
Toward the biological model of the hippocampus as the successor representation agent

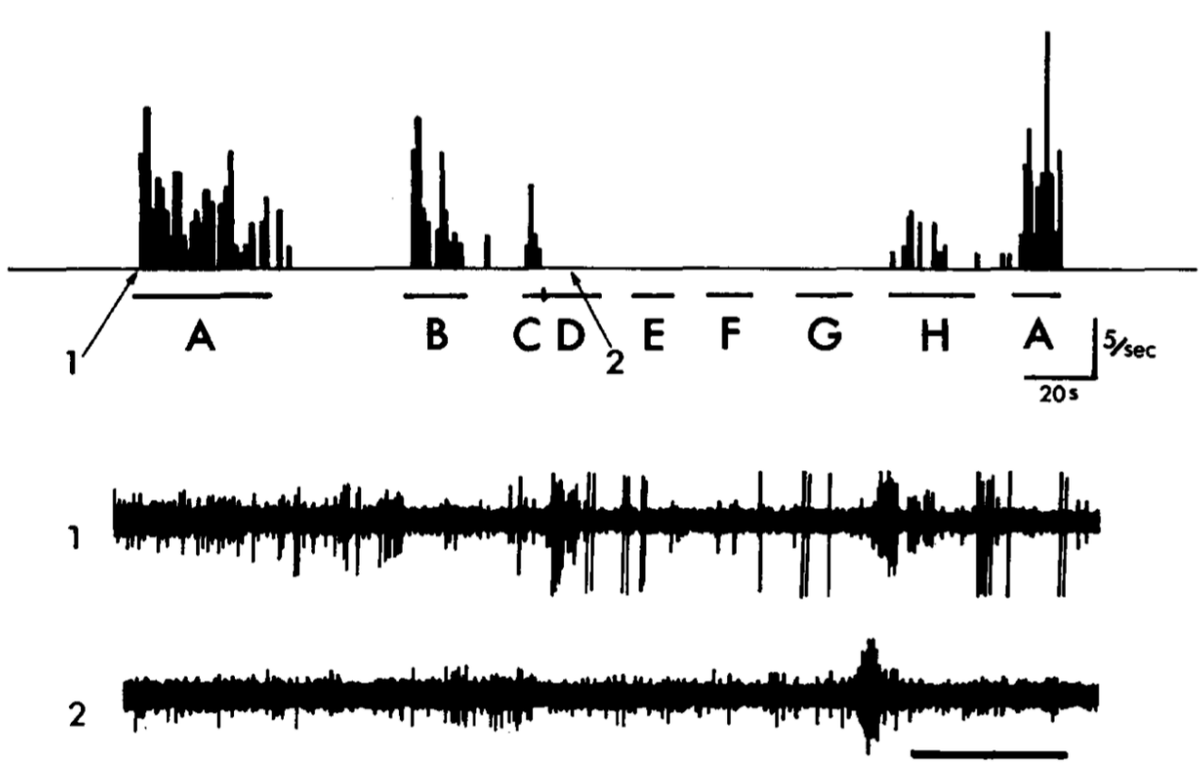
Hyunsu Lee

Department of Anatomy School of Medicine Keimyung University

How do we perceive space?

- After the discovery of hippocampal place cells and grid cells, place cells were believed to response to the simple Gaussian place filed.

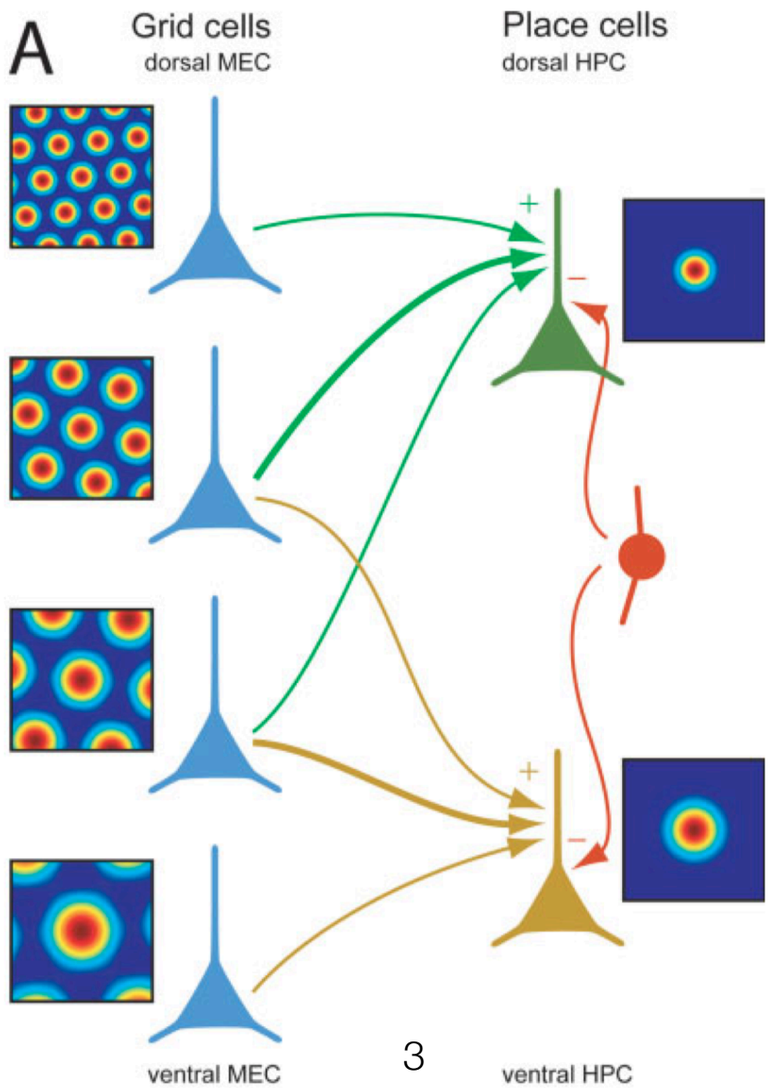




O`keefe and Dostrovsky. 1971

How do we perceive space?

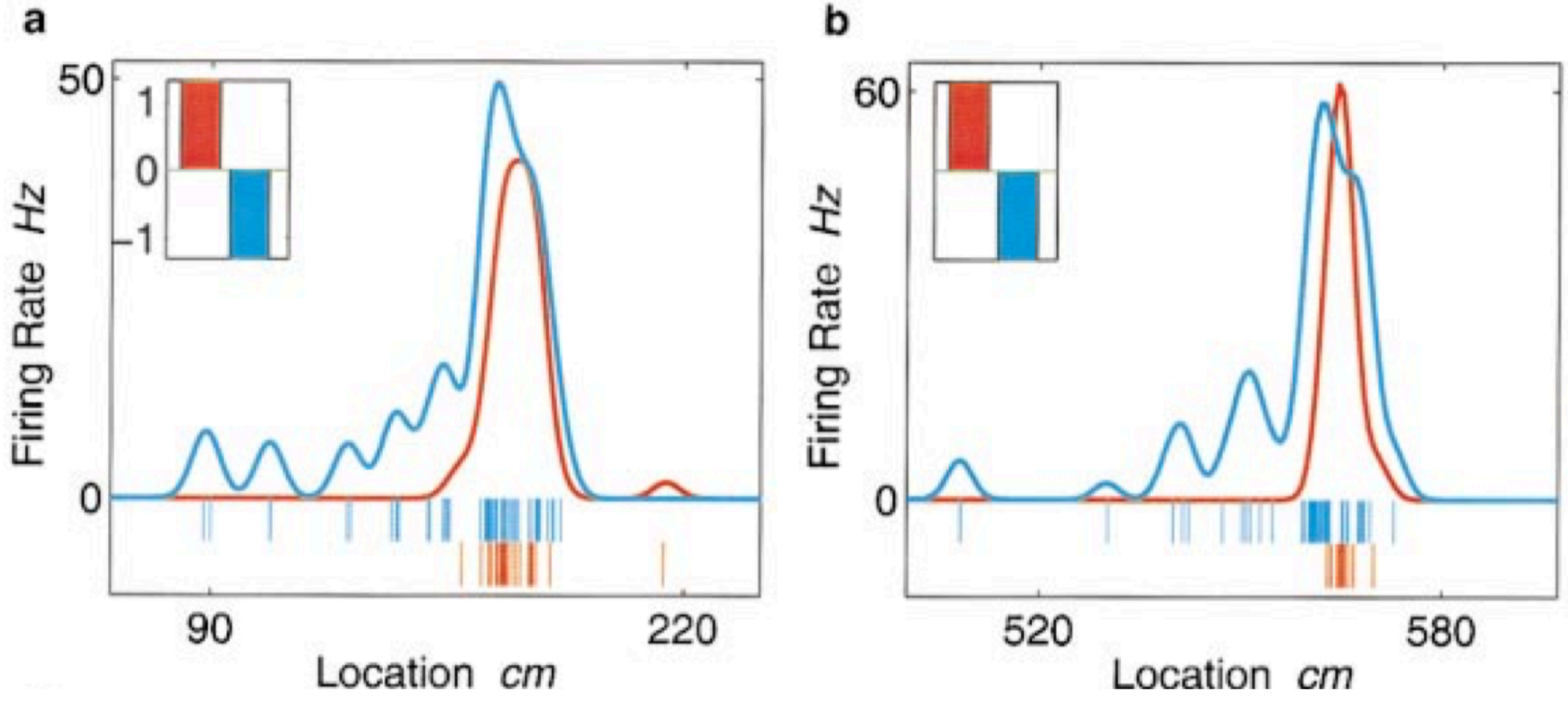
- After the discovery of hippocampal place cells and grid cells, place cells were believed to response to the simple Gaussian place filed.



Solstad et al. 2006

Gaussian place field? However..

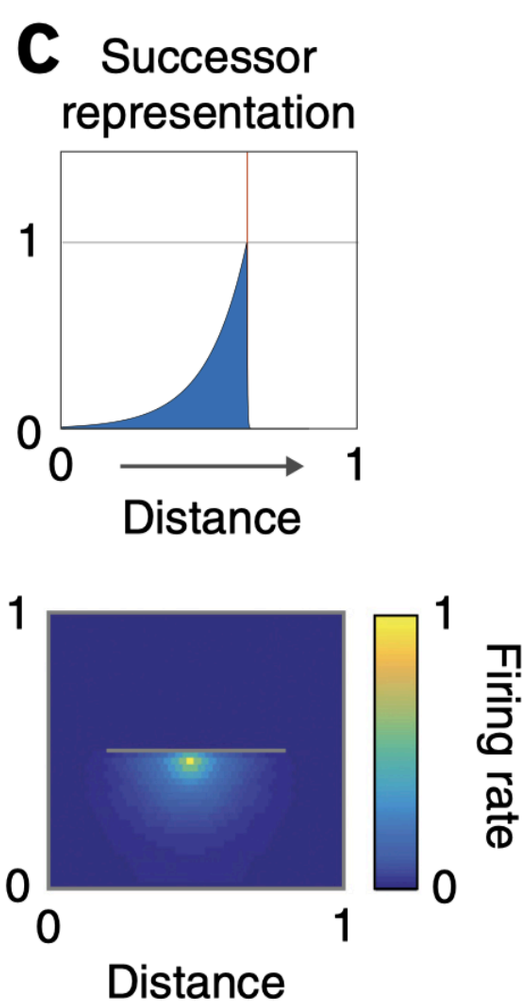
- Even in the 1D-shape maze, the place field was asymmetric after learning.



Mehta et al. 2000

Predictive map theory

- Based on the successor representation model, the predictive map theory was proposed.



Stachenfeld et al. 2017

The SR model, details

- It was based on the TD model.

$$V(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t)]$$

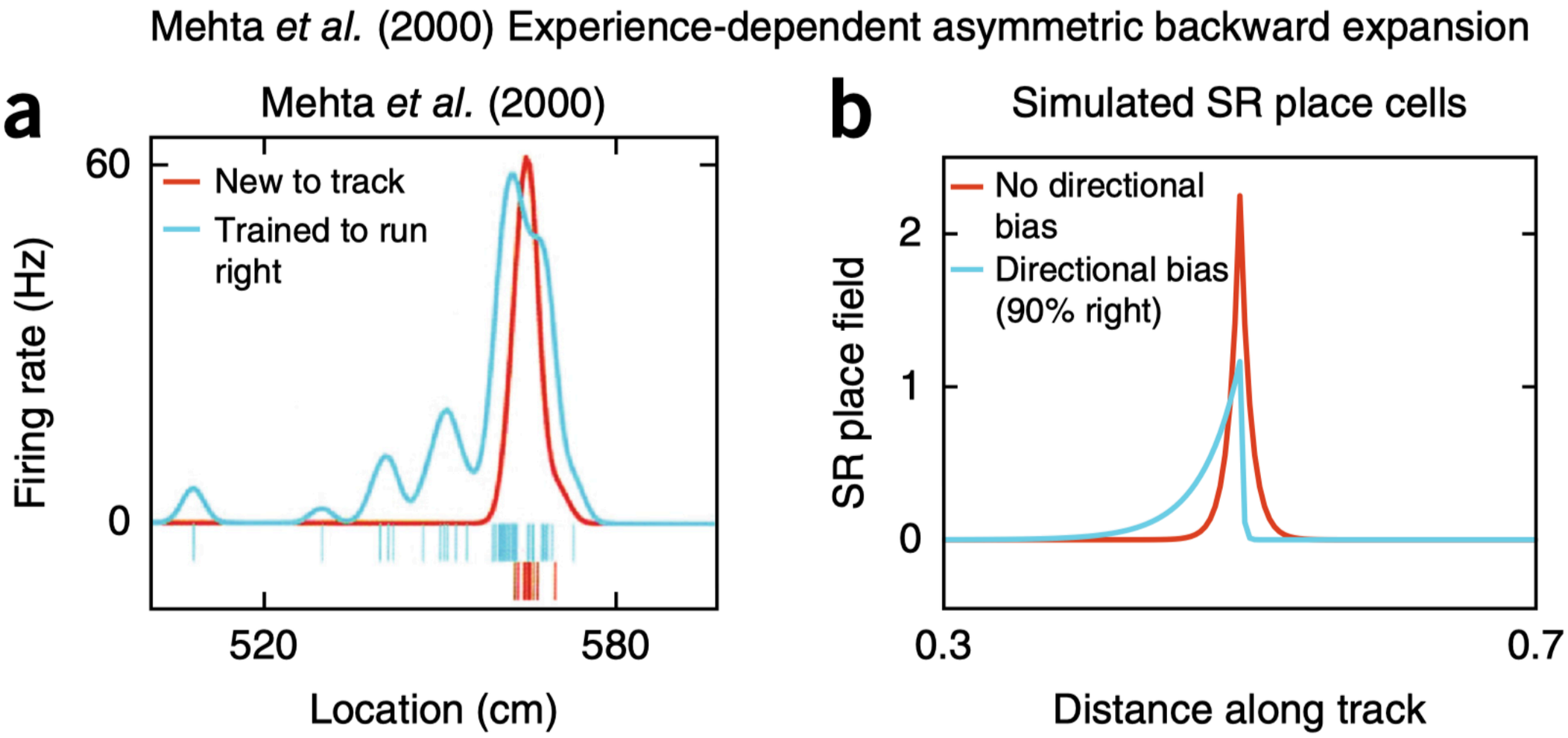
- Decompose the value function into reward function and predictive representation of the state.

$$V(s) = \sum_{s'} M(s, s')R(s')$$

- M indicates expected discounted future state (s') occupancy

$$M(s, s') = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t \mathbb{I}(s_t = s')|s_0 = s]$$

M similar with place field



Stachenfeld et al. 2017

How the SR agent learn?

- Temporal difference learning

$$M_{t+1}(s_t, s') = M_t(s_t, s') + \eta[\mathbb{I}(s_t = s') + \gamma M_t(s_{t+1}, s') - M_t(s_t, s')]$$

Question

$$M_{t+1}(s_t, s') = M_t(s_t, s') + \eta[\mathbb{I}(s_t = s') + \gamma M_t(s_{t+1}, s') - M_t(s_t, s')]$$

- What are the $M_t(S_{t+1}, S')$ and $M_t(S_t, S')$ in the animal?

Toward biological plausible model

- What are the $M_t(S_{t+1}, S')$ and $M_t(S_t, S')$ in the animal?
- If M is treated as an artificial neuron,

$$M_t(s_t, s') = W_t \cdot Pre_t$$

- the inner product of presynaptic input with synaptic weight.

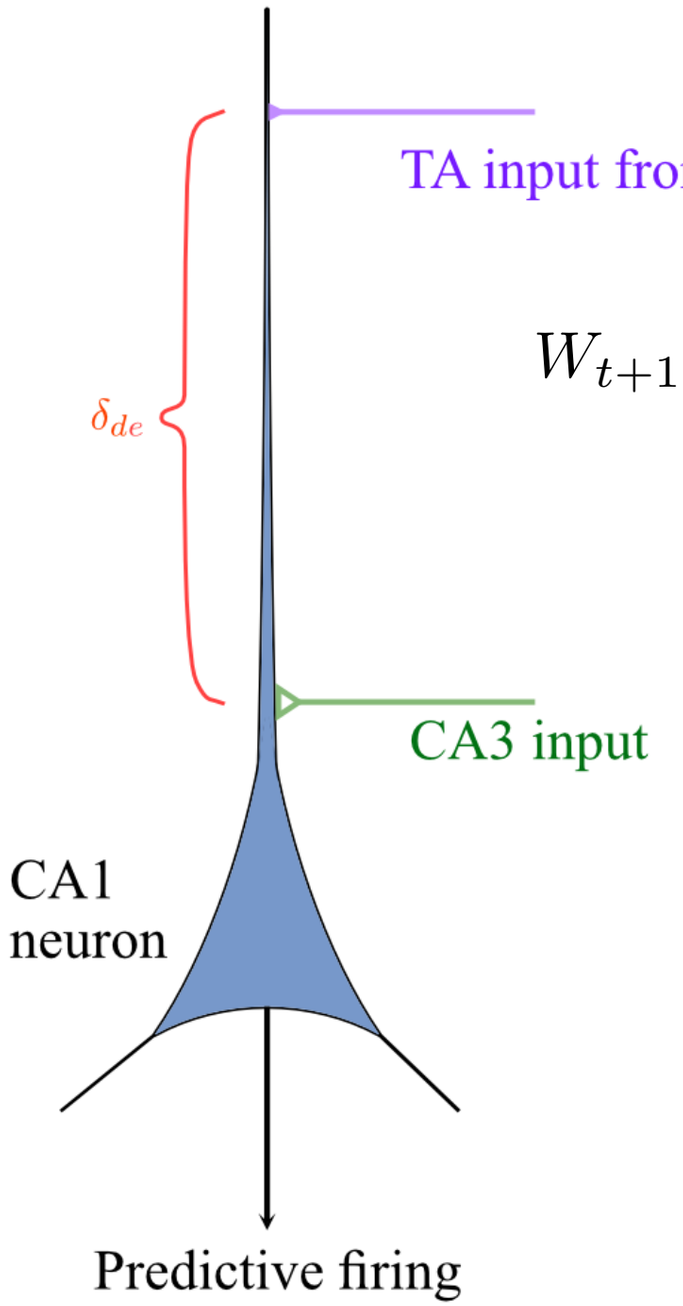
Transformation

$$M_{t+1}(s_t, s') = M_t(s_t, s') + \eta[\mathbb{I}(s_t = s') + \gamma M_t(s_{t+1}, s') - M_t(s_t, s')]$$

$$M_t(s_t, s') = W_t \cdot Pre_t$$

$$W_{t+1} \cdot Pre(s_t) = W_t \cdot Pre(s_t) + \eta[Pre(s_t) + \gamma W_t \cdot Pre(s_{t+1}) - W_t \cdot Pre(s_t)]$$

Based on CA1 connectivity



$$W_{t+1} \cdot Pre^{CA3}(s_t) =$$
$$W_t \cdot Pre^{CA3}(s_t) +$$
$$\eta[Pre^{CA3}(s_t) + \gamma W^{TA} \cdot Pre^{TA}(s_{t+1}) - W_t \cdot Pre^{CA3}(s_t)]$$

Derived heterosynaptic plasticity

- In transforming the synaptic update rule from the SR model, heterosynaptic plasticity rule was revealed.

$$\Delta W \cdot Pre^{CA3}(s_t) = \eta[Pre^{CA3}(s_t) + \gamma W^{TA} \cdot Pre^{TA}(s_{t+1}) - W_t \cdot Pre^{CA3}(s_t)]$$

$$\Delta W = \eta[\gamma W^{TA} \cdot Pre^{TA}(s_{t+1}) \cdot inv(Pre^{CA3}(s_t)) - W_t + 1]$$

Comparison

error term of TD learning

$$\delta_{TD} = R_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)$$

error term of SR learning

$$\delta_{SR} = \mathbb{I}(s_t = s') + \gamma M_t(s_{t+1}, s') - M_t(s_t, s')$$

The dendritic error term of CA1 derived from the SR model includes the heterosynaptic plasticity rule

$$\begin{aligned}
\delta_{de} &= Pre^{CA3}(s_t) + \gamma W^{TA} \cdot Pre^{TA}(s_{t+1}) - W_t \cdot Pre^{CA3}(s_t) \\
&= \Delta W \cdot Pre^{CA3}(s_t) \\
\Delta W &= \gamma W^{TA} \cdot \underline{Pre^{TA}(s_{t+1}) \cdot inv(Pre^{CA3}(s_t))} - W_t + 1
\end{aligned}$$

Discussion

- Conjunctive activation of TA and CA3 inputs drives dendritic plateau potential of CA1 place cells and forming new place fields.(Bittner et al. 2015)

- This biological mechanism is comparable with our heterosynaptic plasticity rule derived from the TD algorithm.

References

- O’Keefe J, Dostrovsky J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 1971;34:171-175.

- Solstad T, Moser EI, Einevoll GT. From grid cells to place cells: A mathematical model. Hippocampus. 2006;16:1026-1031.

- Mehta MR, Quirk MC, Wilson MA. Experience-dependent asymmetric shape of hippocampal receptive fields. Neuron. 2000;25:707-715.

- Stachenfeld KL, Botvinick MM, Gershman SJ. The hippocampus as a predictive map. Nat Neurosci. 2017;7:1951.

- Bittner KC, Grienberger C, Vaidya SP et al. Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. Nat Neurosci. 2015;18:1133.