Bonus Material
TD-learning in Biological System

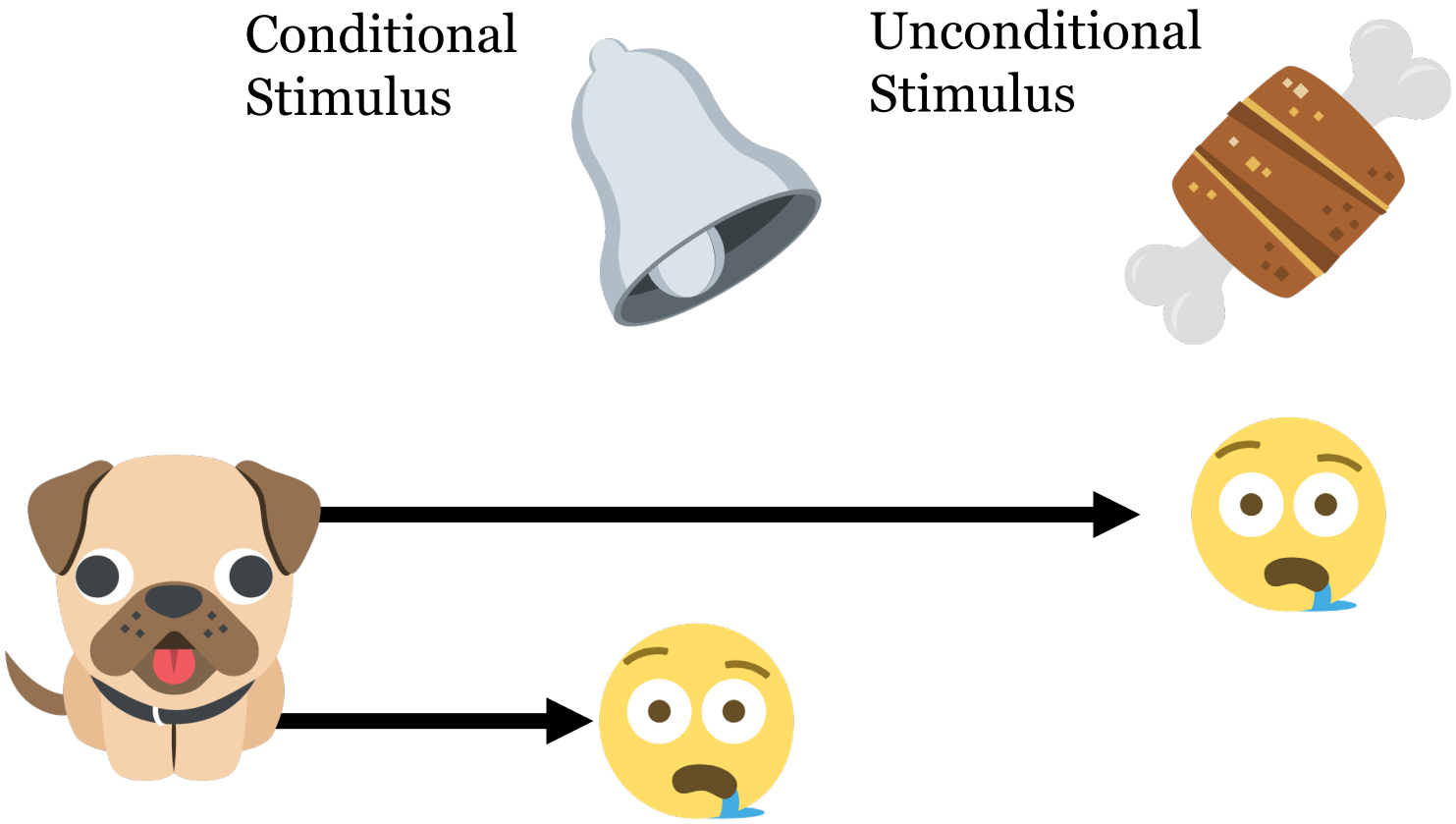
Understand learning in biology

"The neural basis of RL in mammalian systems is arguably one of the best understood system in system neuroscience

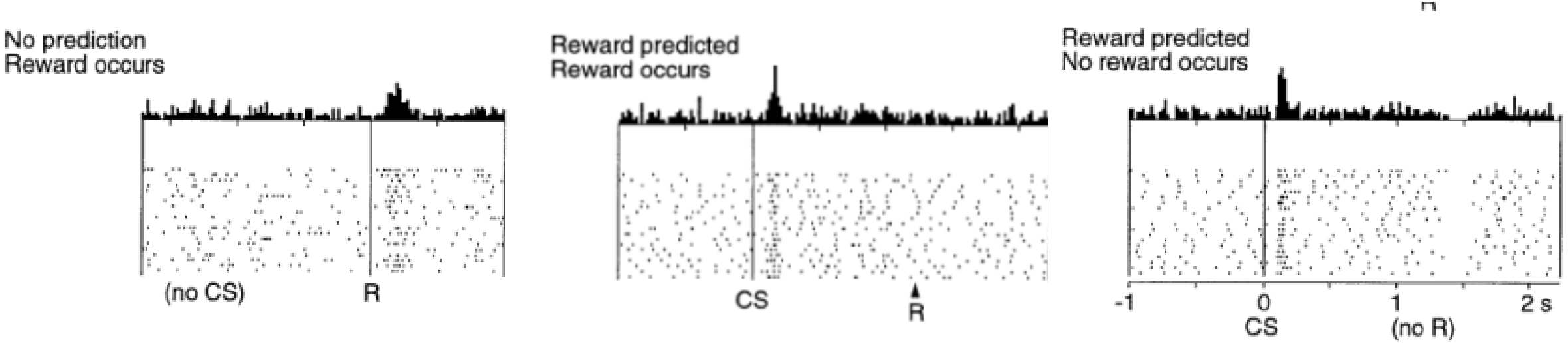
[Thanks to TD-learning] "

Neftci and Averbeck 2019

Classical conditioning



Phasic Dopamine neuron [Shultz 1998]



Surprise!

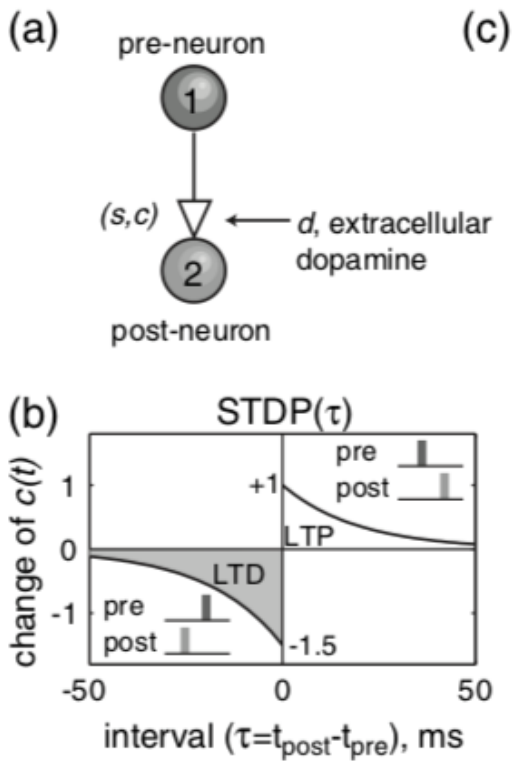
No surprise!

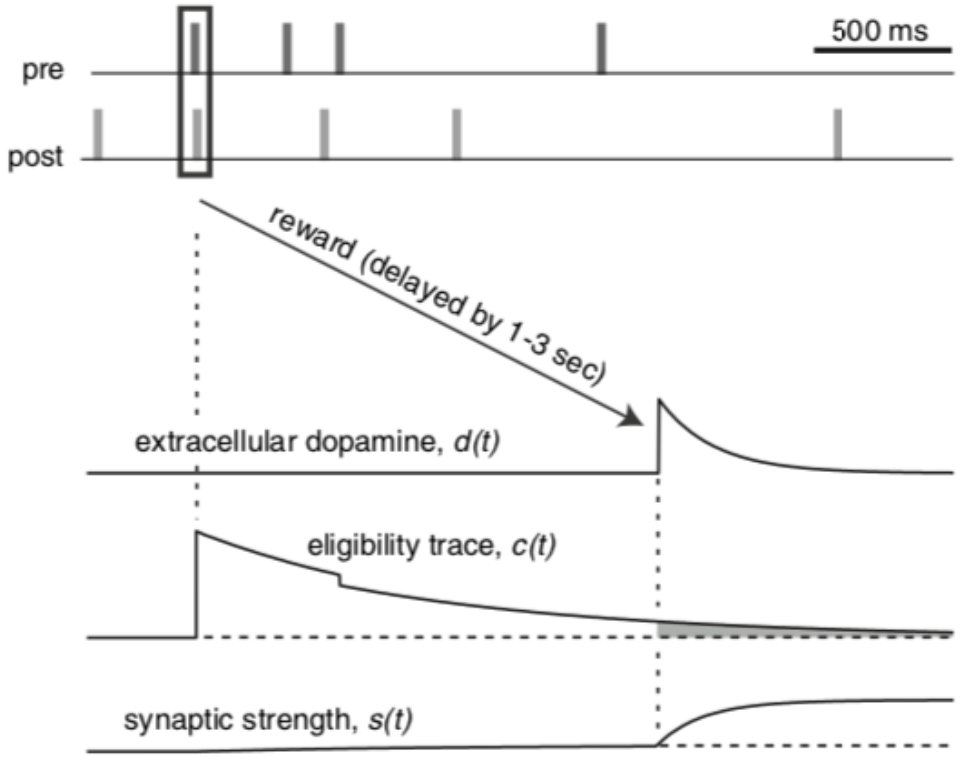
Negative TD?

Prediction error = r - v

- From this experiment by Shultz 1998,
  we think that phasic dopamine serves as reward prediction error.

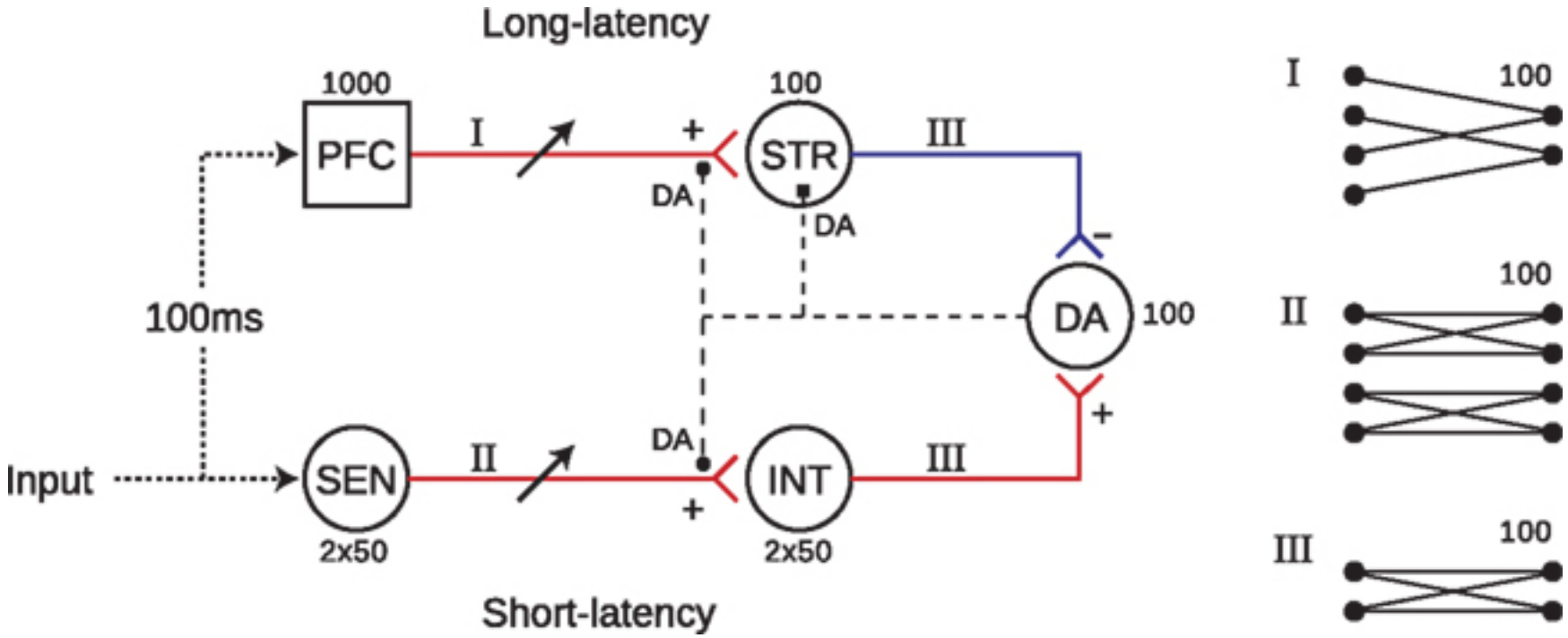
Possible Mechanism of how dopamine effect the synaptic strength



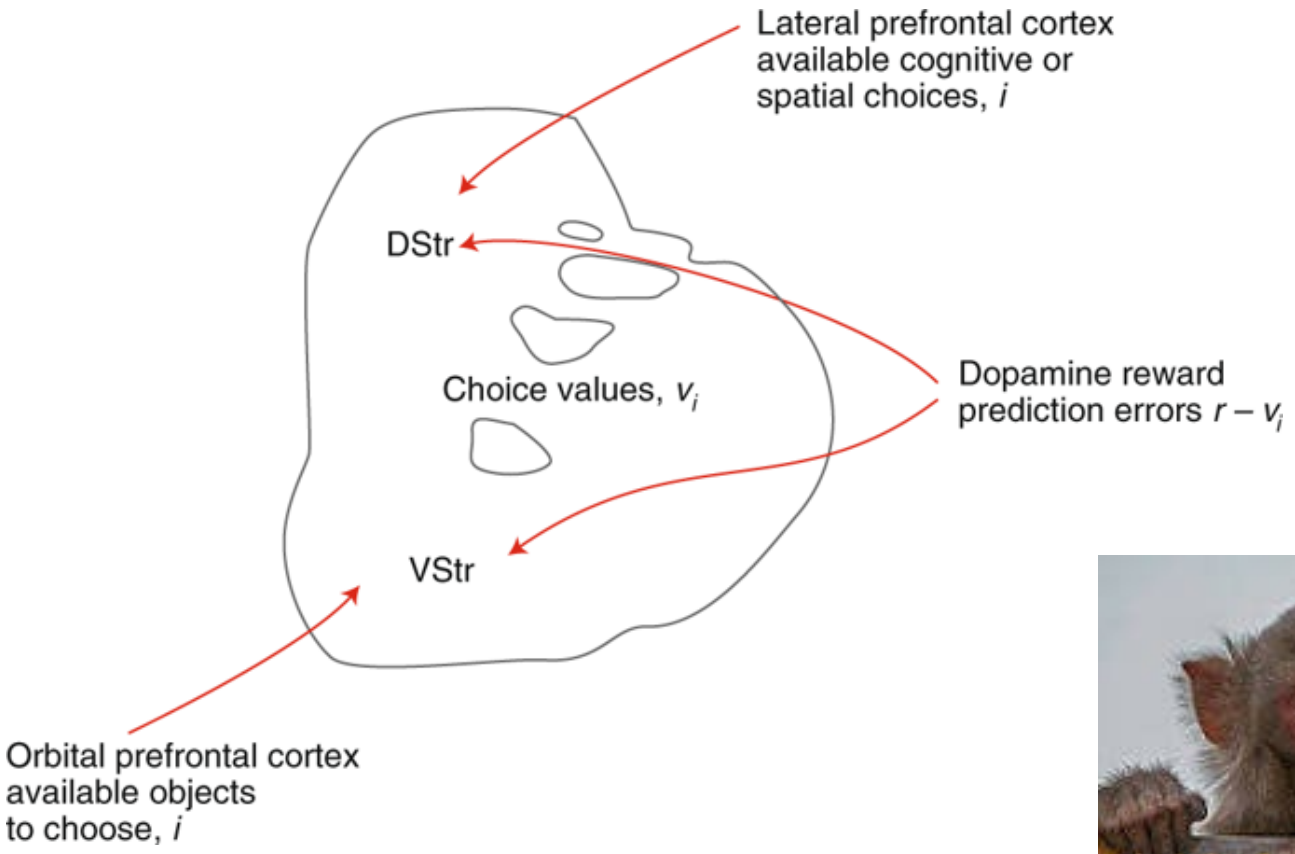


$\delta \leftarrow r + \gamma Q(s', a^*) - Q(s, a)$
$e(s, a) \leftarrow e(s, a) + 1$
For all s, a:
$Q(s, a) \leftarrow Q(s, a) + \alpha \delta e(s, a)$
If $a' = a^*$, then $e(s, a) \leftarrow \gamma \lambda e(s, a)$
else $e(s, a) \leftarrow 0$
$s \leftarrow s'$; $a \leftarrow a'$

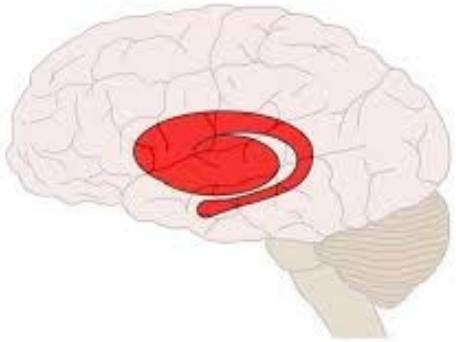
Chorley and Seth's Circuit



People have been looking at the activity of dopamine neurons and its effect on behavior.

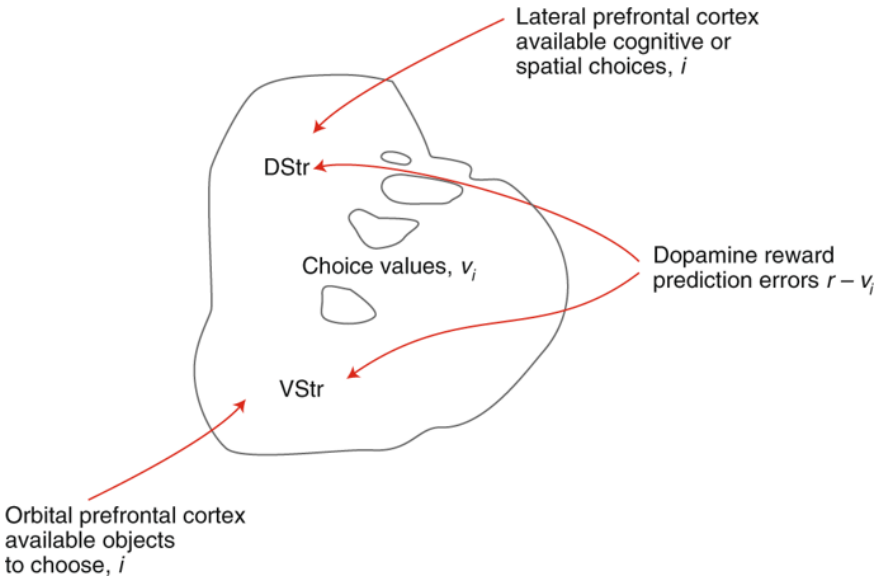


This is a model of the rhesus monkey striatum



According to one model, they think that …

- Cortex represents the set of available choices (i.e. actions).
- The activity of striatal cells encodes information about the values.
- Dopamine encode reward prediction error.



What are other algorithms the brain might be using ?