Classical Conditioning IV: TD learning in the brain

PSY/NEU338: Animal learning and decision making: Psychological, computational and neural perspectives

recap: Marr's levels of analysis

David Marr (1945-1980) proposed three levels of analysis:

1. the problem (Computational Level)
2. the strategy (Algorithmic Level)
3. how its actually done by networks of neurons (Implementational Level)

lets start over, this time from the top...

The problem: optimal prediction of future reinforcement

$$V_t = E\left[\sum_{i=t+1}^{\infty} r_i\right]$$

want to predict expected sum of future reinforcement

$$V_t = E\left[\sum_{i=t+1}^{\infty} \gamma^{i-t-1} r_i\right]$$

want to predict expected sum of *discounted* future reinf. ($0<\gamma<1$)

$$V_t = E\left[\sum_{i=t+1}^{T} r_i\right]$$

want to predict expected sum of future reinforcement in a trial/episode

lets start over, this time from the top...

The problem: optimal prediction of future reinforcement

$$\begin{aligned} V_t &= E\left[r_{t+1} + r_{t+2} + \ldots + r_T\right] \\ &= E\left[r_{t+1}\right] + E\left[r_{t+2} + \ldots + r_T\right] \\ &= E\left[r_{t+1}\right] + V_{t+1} \end{aligned}$$

(note: *t* indexes time within a trial)

$$V_t = E\left[\sum_{i=t+1}^{T} r_i\right]$$

want to predict expected sum of future reinforcement in a trial/episode

lets start over, this time from the top...

The problem: optimal prediction of future reinforcement

$$V_t = E\left[r_{t+1} + r_{t+2} + \ldots + r_T\right]$$
$$= E\left[r_{t+1}\right] + E\left[r_{t+2} + \ldots + r_T\right]$$
$$= E\left[r_{t+1}\right] + V_{t+1}$$

(note: *t* indexes time within a trial)

Think football…

What would be a sensible learning rule here?

How is this different from Rescorla-Wagner?

Temporal Difference (TD) learning

Marr's 3 levels:

The problem: optimal prediction of future reinforcement

The algorithm: $V_t = E\left[r_{t+1}\right] + V_{t+1}$

(note: *t* indexes time within a trial, T indexes trials)

$$V_t^{new} = V_t^{old} + \eta(\underbrace{r_{t+1} + V_{t+1}^{old} - V_t^{old}})$$

temporal difference prediction error δ(t+1)

compare to: $V^{T+1} = V^T + \eta\left(r^T - V^T\right)$

Sutton & Barto 1983, 1990

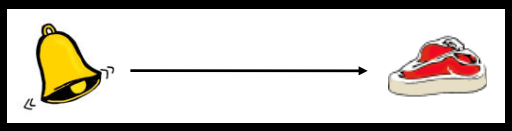
Temporal Difference (TD) learning

Marr's 3 levels:

The problem: optimal prediction of future reinforcement

The algorithm: $V_t^{new} = V_t^{old} + \eta(r_{t+1} + V_{t+1}^{old} - V_t^{old})$

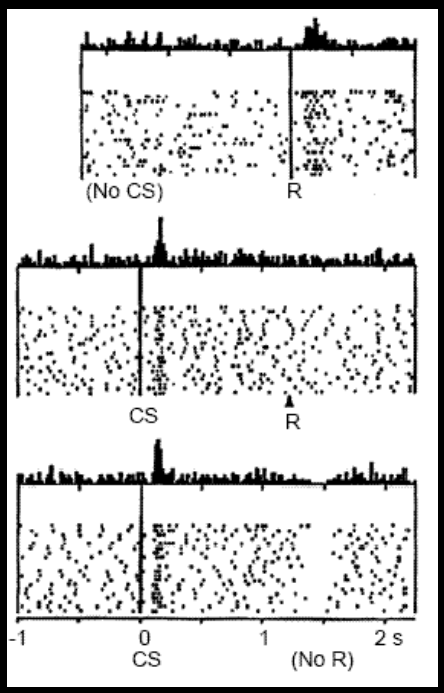
prediction error $\delta(t+1)$



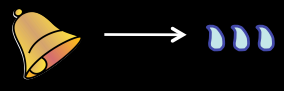
beginning of trial	middle of trial	end of trial
$r_t = 0$	$r_t = 0$	$V_t = 0$
$V_{t-1} = 0$		
$\delta(t) = V_t$	$\delta(t) = V_t - V_{t-1}$	$\delta(t) = r_t - V_{t-1}$

Sutton & Barto 1983, 1990

dopamine



$\delta(t) = r_t$

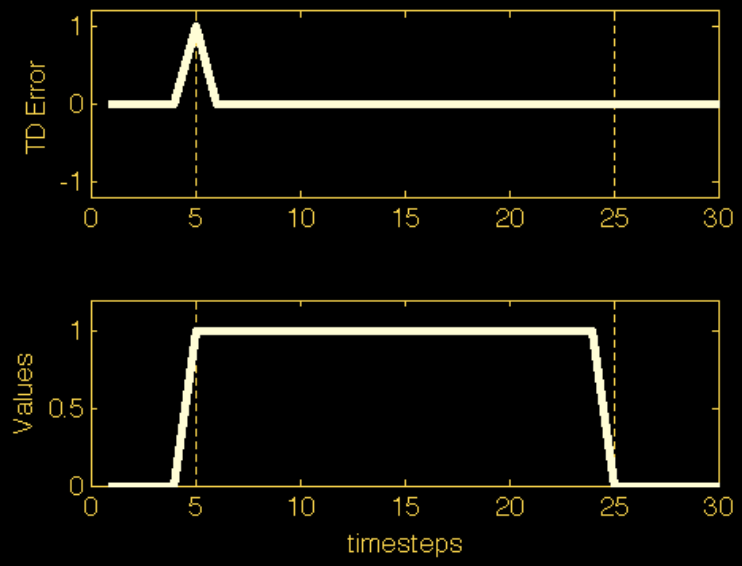


$\delta(t) = V_t$ $\delta(t) = r_t - V_{t-1}$

$\delta(t) = V_t$ $\delta(t) = 0 - V_{t-1}$

Schultz et al, 1997

simulation



what would happen with partial reinforcement?
what would happen in second order conditioning?

A note on time book-keeping

$$V_t^{new} = V_t^{old} + \eta(\underbrace{r_{t+1} + V_{t+1}^{old} - V_t^{old}}_{\text{prediction error } \delta_{t+1}})$$

- the prediction error δ can be defined at any time point, but can only be based on information the animal already has (not information that it will get in the future!)
- so, we can say $\delta_{t+1} = r_{t+1}+V_{t+1}-V_t$ as above, because at time (t+1) information regarding r_{t+1} and V_{t+1} is already known
- we can also say $\delta_t = r_t+V_t-V_{t-1}$ in pretty much the same way
- we *cannot* say $\delta_t = r_{t+1}+V_{t+1}-V_t$ as it just would not make logical sense!
- importantly, in all cases, δ is used to update the preceding prediction, that is, δ_{t+1} is used to update V_t and δ_t is used to update V_{t-1}

what does the theory explain?

	R-W	TD
acquisition	√	√
extinction	√	√
blocking	√	√
overshadowing	√	√
temporal relationships	X	√
overexpectation	√	√
2nd order cond.	X	√

Summary so far...

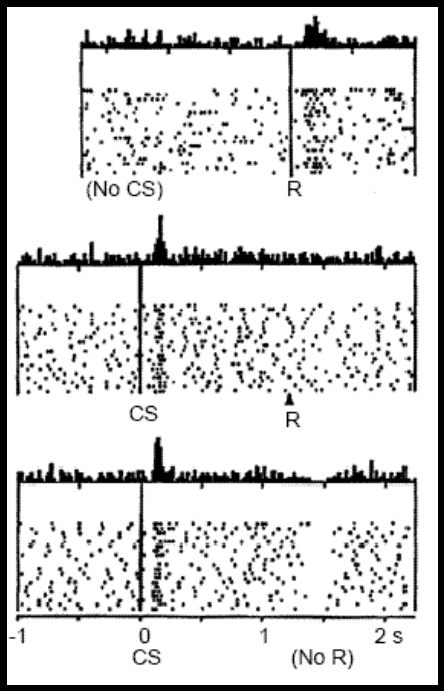
- Temporal difference learning is a "better" version of Rescorla-Wagner learning
- derived from first principles (from definition of problem)
- explains everything that R-W does, and more (eg. 2nd order conditioning)
- basically a generalization of R-W to real time

break!

Back to Marr's three levels

The problem: optimal prediction of future reinforcement
The algorithm: temporal difference learning
Neural implementation: does the brain use TD learning?

we already saw this:



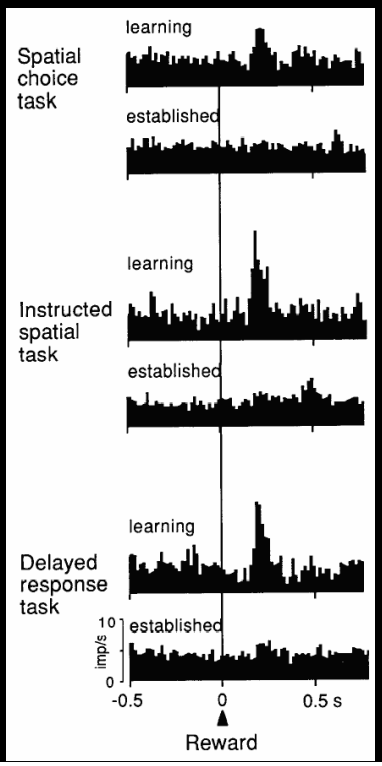
$\delta(t) = r_t$

$\delta(t) = V_t$ $\delta(t) = r_t - V_{t-1}$

$\delta(t) = V_t$ $\delta(t) = 0 - V_{t-1}$

Schultz et al, 1997

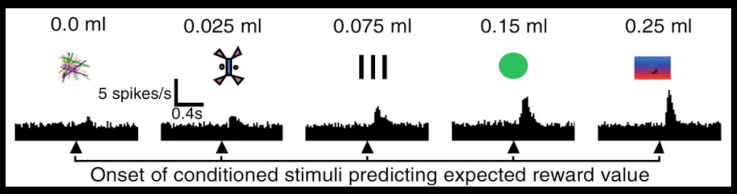
prediction error hypothesis of dopamine



Schultz et al, 1993

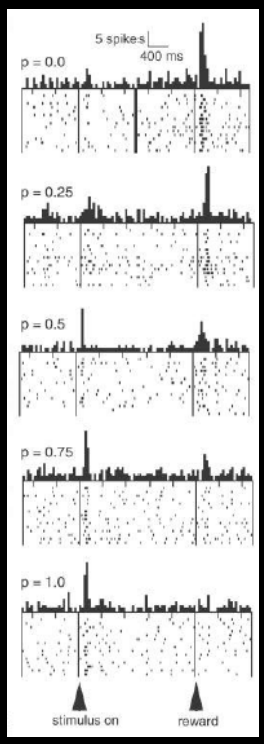
The idea: Dopamine encodes a temporal difference reward prediction error

(Montague, Dayan, Barto mid 90's)

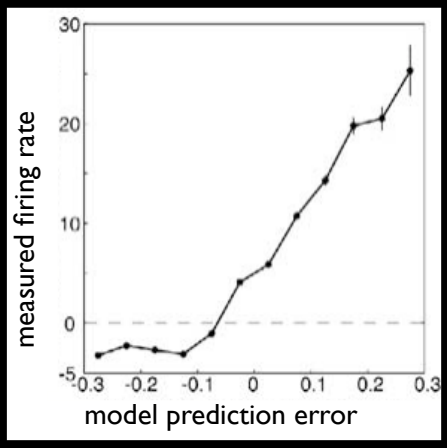


Tobler et al, 2005

prediction error hypothesis of dopamine



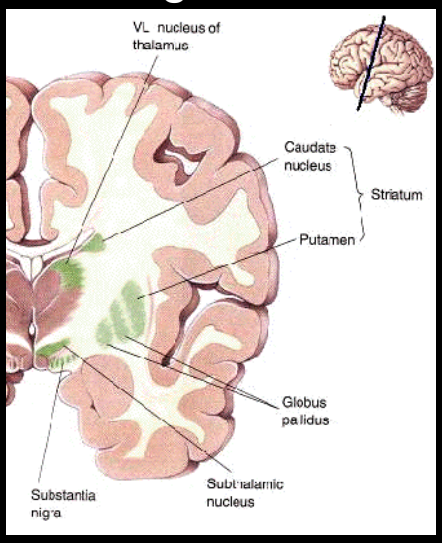
Fiorillo et al, 2003

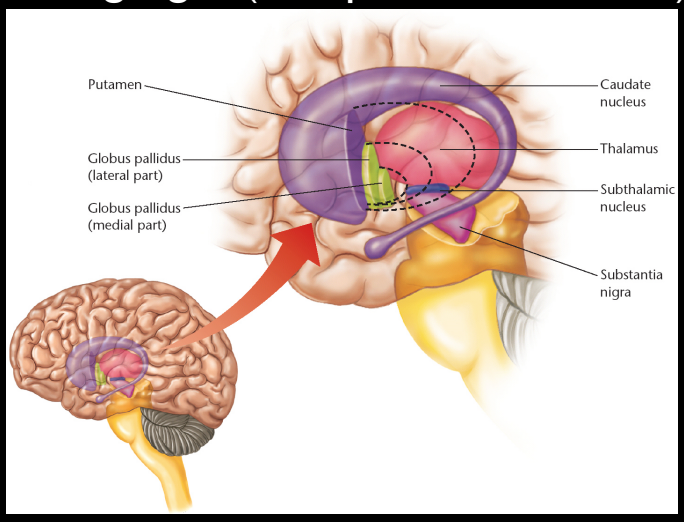


Bayer & Glimcher (2005)

where does dopamine project to?

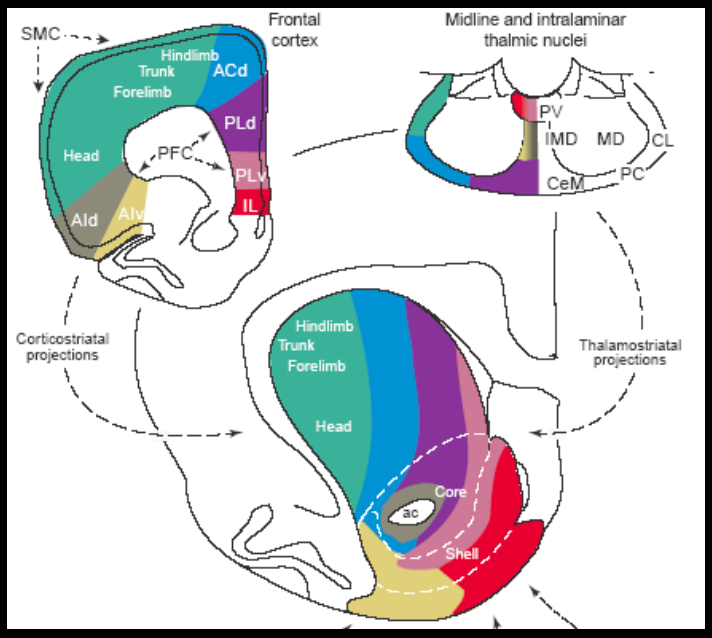
main target: striatum in basal ganglia (also prefrontal cortex)





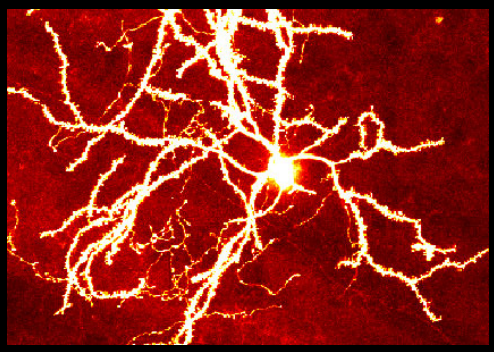
the basal ganglia: afferents

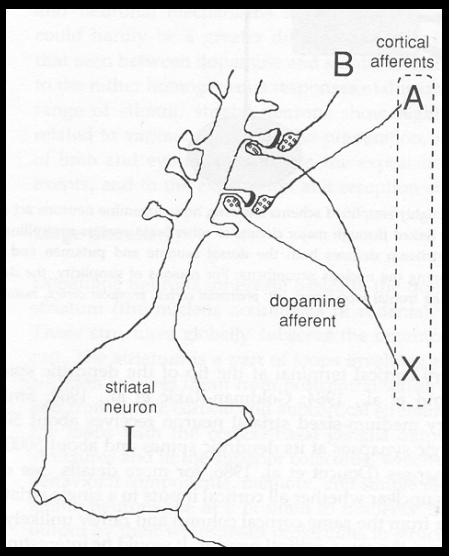
inputs to striatum are from all over the cortex
(and they are topographic)



Voorn et al, 2004

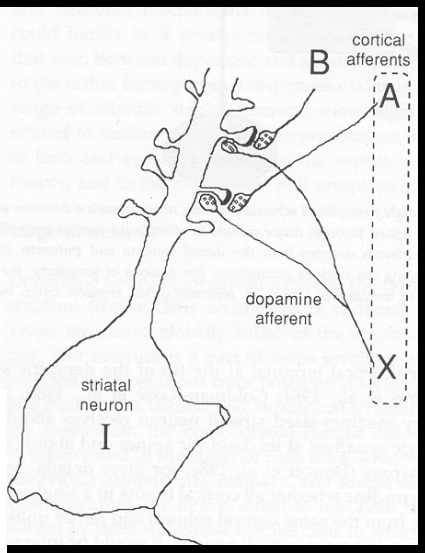
a precise microstructure

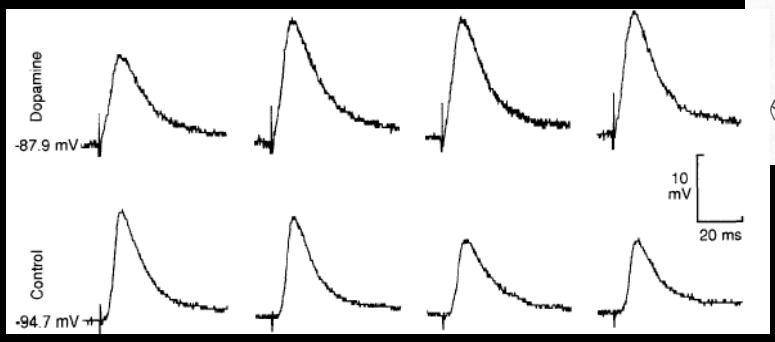




dopamine and synaptic plasticity

- prediction errors are for learning…
- cortico-striatal synapses show dopamine-dependent plasticity
- three-factor learning rule: need presynaptic+postsynaptic+dopamine





Wickens et al, 1996

summary

Classical conditioning can be viewed as prediction learning

- The problem: prediction of future reward
- The algorithm: temporal difference learning
- Neural implementation: dopamine dependent learning in BG

⇒ A computational model of learning allows us to look in the brain for "hidden variables" postulated by the model

⇒ Precise (normative!) theory for generation of dopamine firing patterns

⇒ Explains anticipatory dopaminergic responding, 2nd order conditioning

⇒ Compelling account for the role of dopamine in classical conditioning: prediction error drives prediction learning

if you are confused or intrigued: additional reading

- Rescorla & Wagner (1972) - *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement* - the original chapter that is so well cited (and well written!)
- Sutton & Barto (1990) - *Time derivative models of Pavlovian reinforcement* - shows step by step why TD learning is a suitable rule for modeling classical conditioning
- Niv & Schoenbaum (2008) - *Dialogues on prediction errors* - a guide for the perplexed
- Barto (1995) - *adaptive critic and the basal ganglia* - very clear exposition to TD learning in the basal ganglia

(all will be on BlackBoard)