RL Tutorial 2

Eric DeWitt

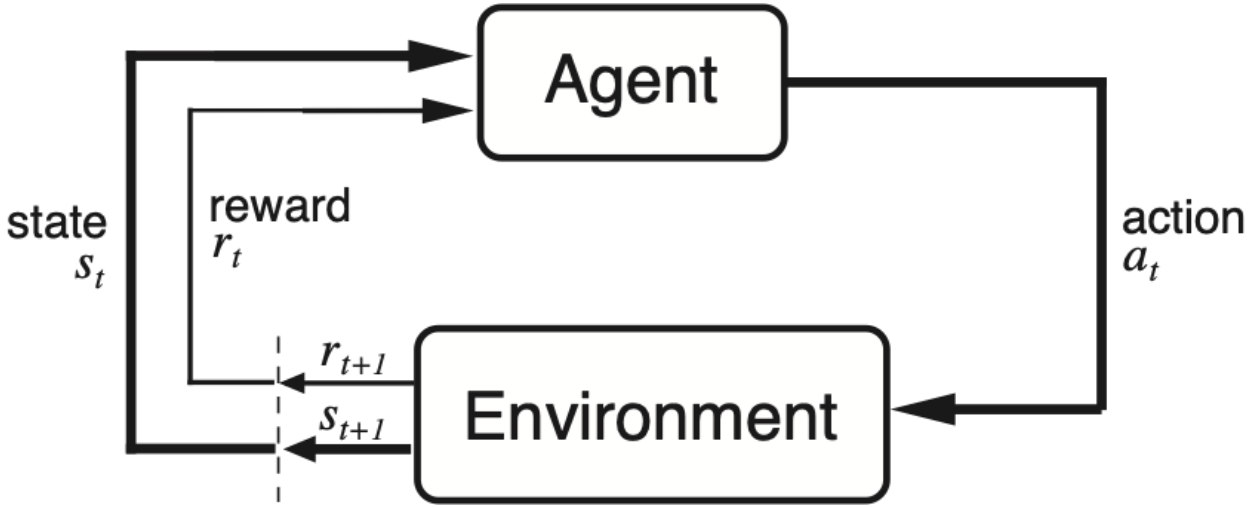
How do RL agents learn to act?

You should have an intuition from Tutorial 1 for how dopamine might implement a reward prediction error to learn future expected value.

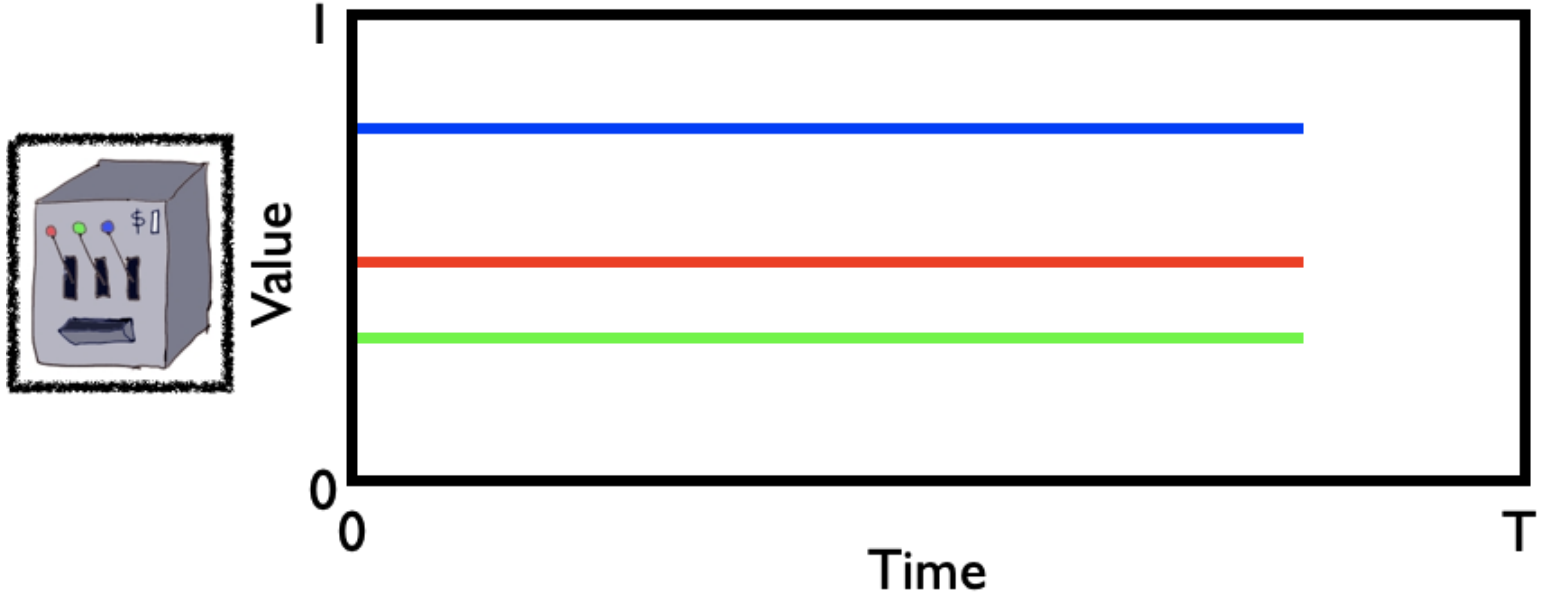
But how does the brain use expected values to learn ***how*** to act?

- A policy is the formal description of how an agent chooses an action based on the observed state of the world
- A fundamental problem with learning how to act is exploration vs exploitation
- When you have states and actions, you need to use a representation of value that informs the policy—or an action-value representation—if you want both the value function and the actions to produce the optimal behavior

Why is reinforcement learning interesting?



The n-armed bandit: explore or exploit?

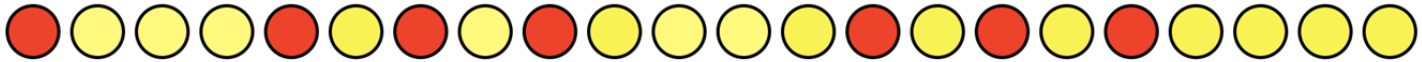


The n-armed bandit: explore or exploit?

Explore then exploit

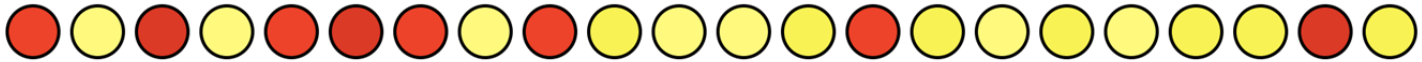
...

ε-Greedy



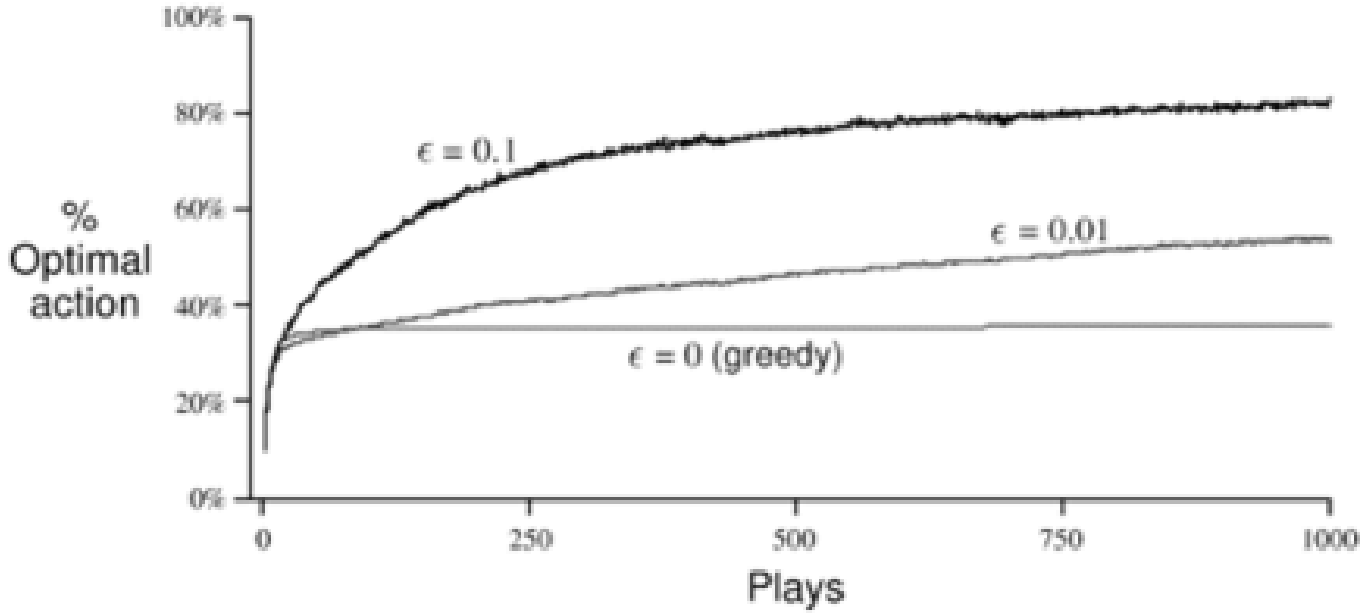
...

ε-Greedy with decaying **ε**

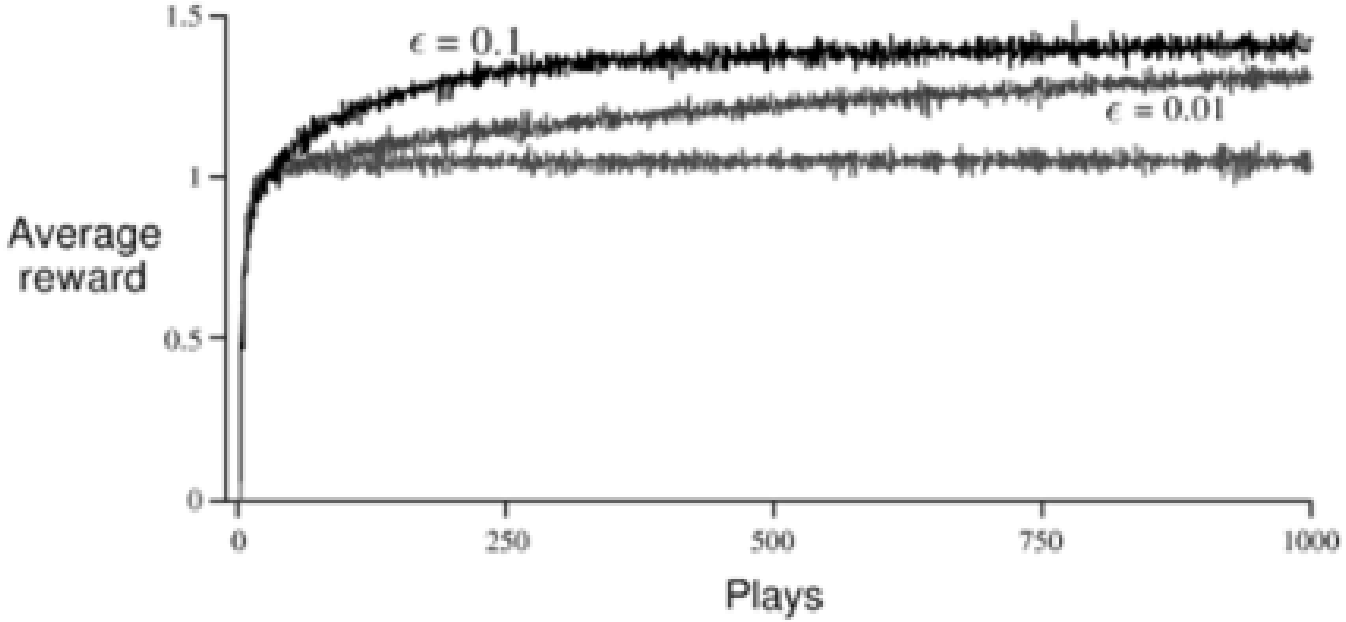


...

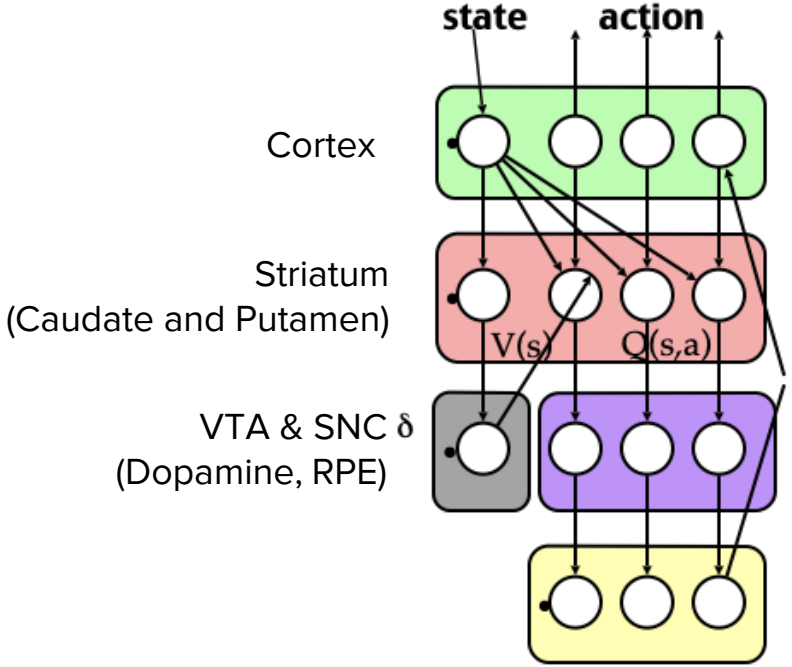
The n-armed bandit: explore or exploit?



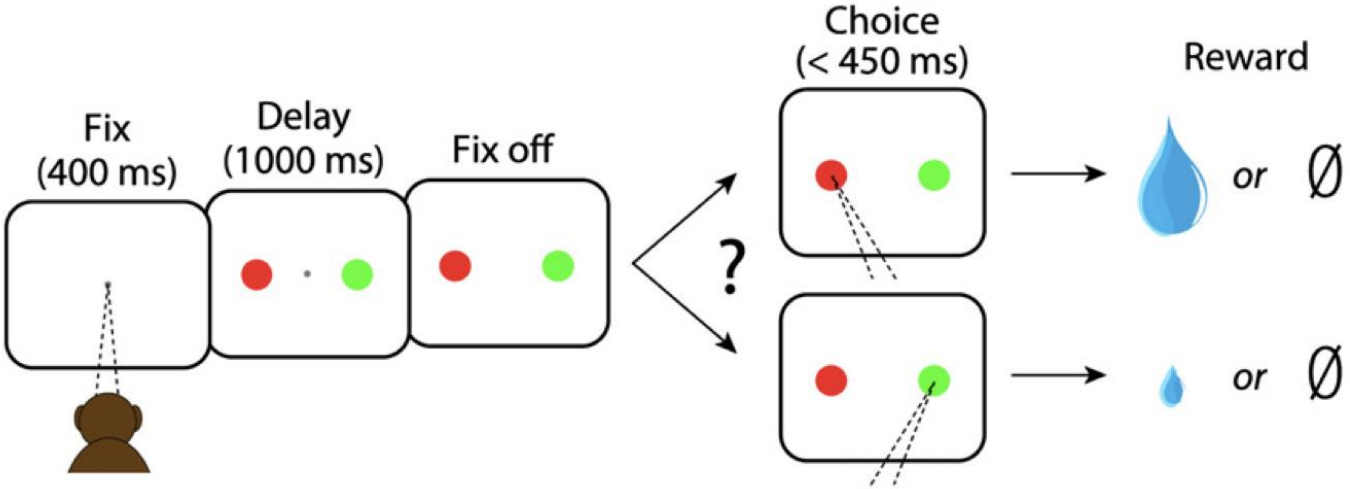
The n-armed bandit: explore or exploit?



Action values in biology

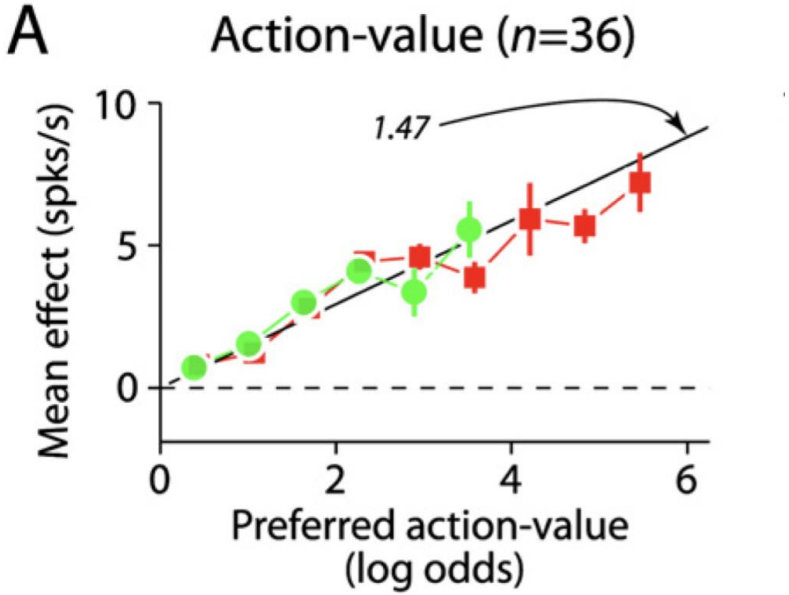


Action values in biology



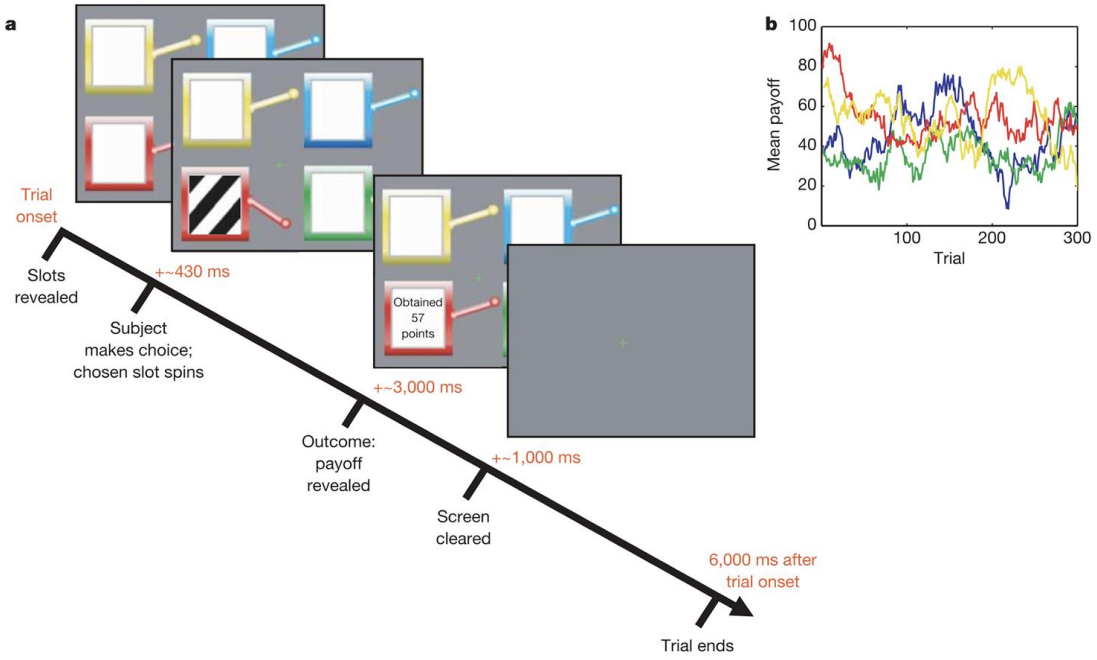
Lau & Glimcher (2008)

Action values in biology



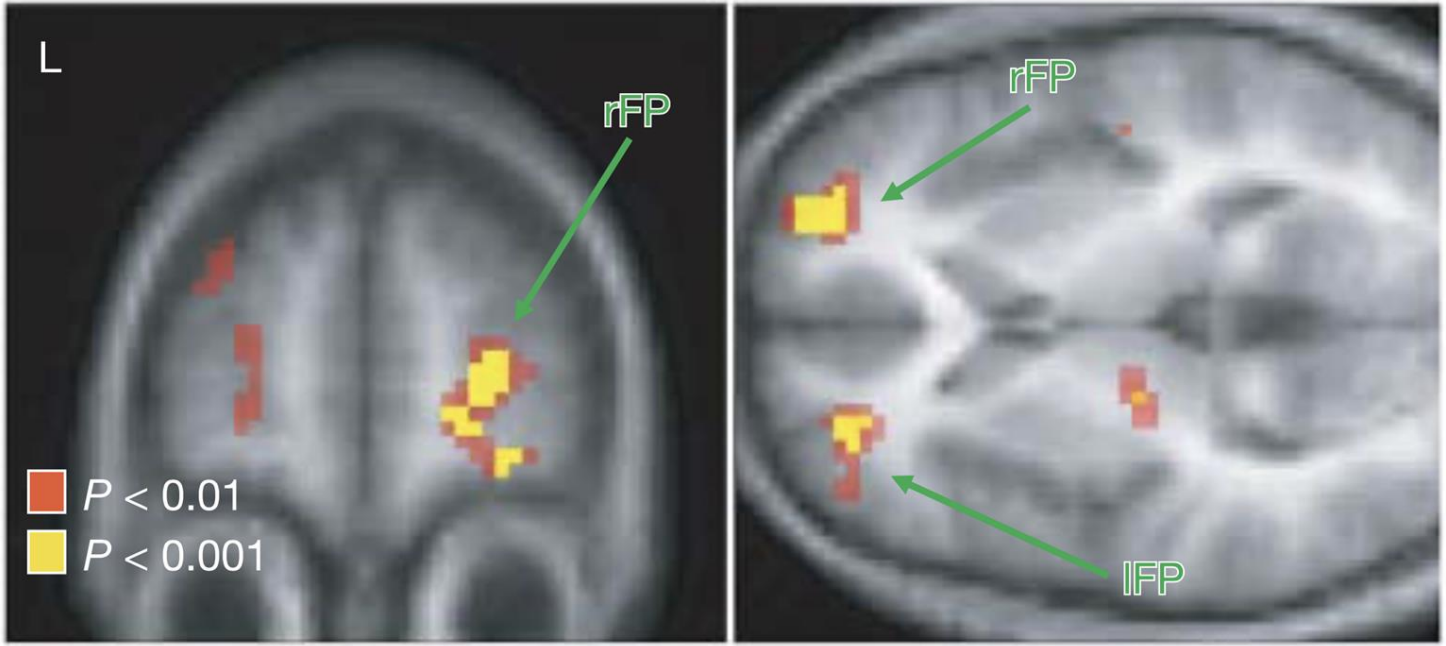
Lau & Glimcher (2008)

The n-armed bandit: explore or exploit?



Daw et al (2006)

The n-armed bandit: explore or exploit?



Daw et al (2006)

Exercises 1, 2 and 3

You will use n-armed bandits study action values representations and to build an intuition for the exploration exploitation trade-off.

- In Exercise 1 you will implement $\boldsymbol{\varepsilon}$-Greedy
- In Exercise 2 you will implement an action-value update
- In Exercise 3 you will explore how different values of the ‘exploration’ parameter ($\boldsymbol{\varepsilon}$) and the learning rate ($\boldsymbol{\alpha}$) effect how well the agents learns