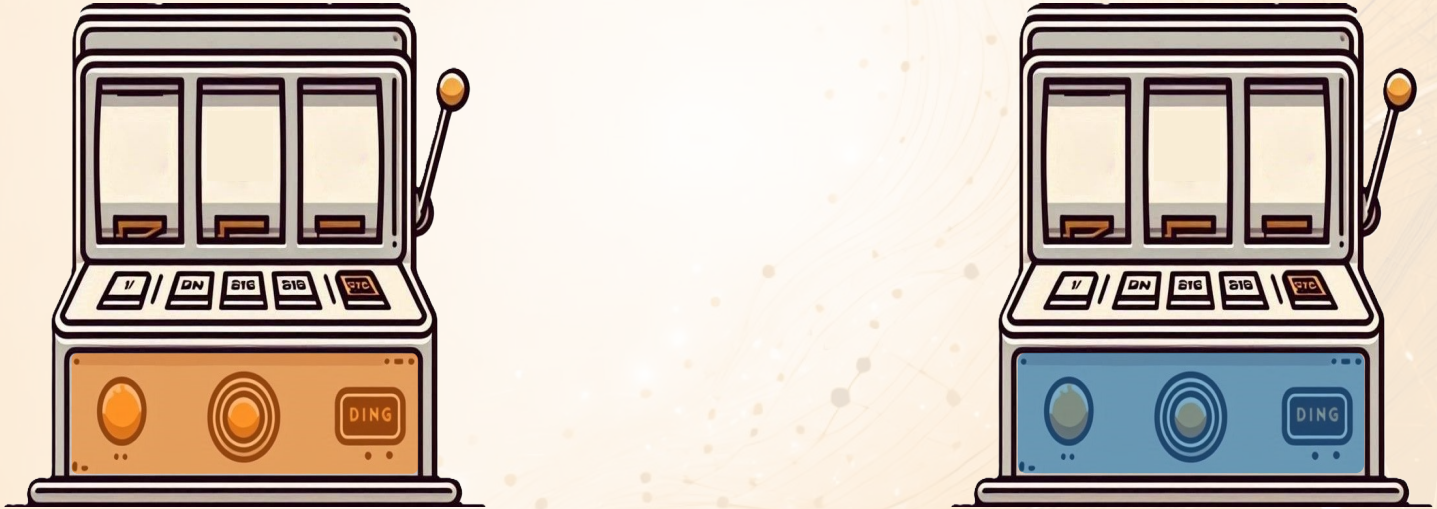
Lesson 1:
Two-armed Bandit Task

Multi Armed Bandit

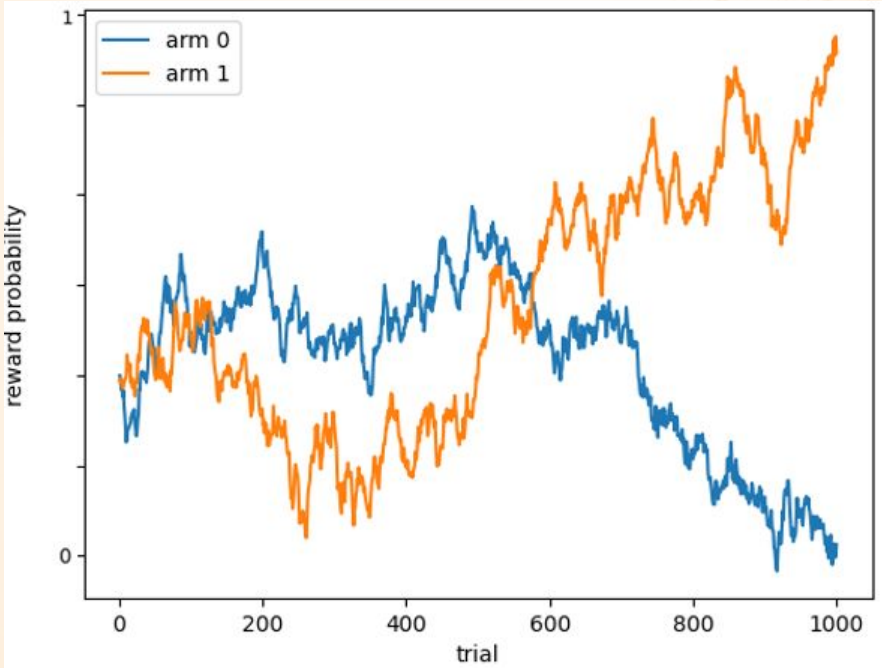


Multi Armed Bandit

Learn the action policy to maximize reward

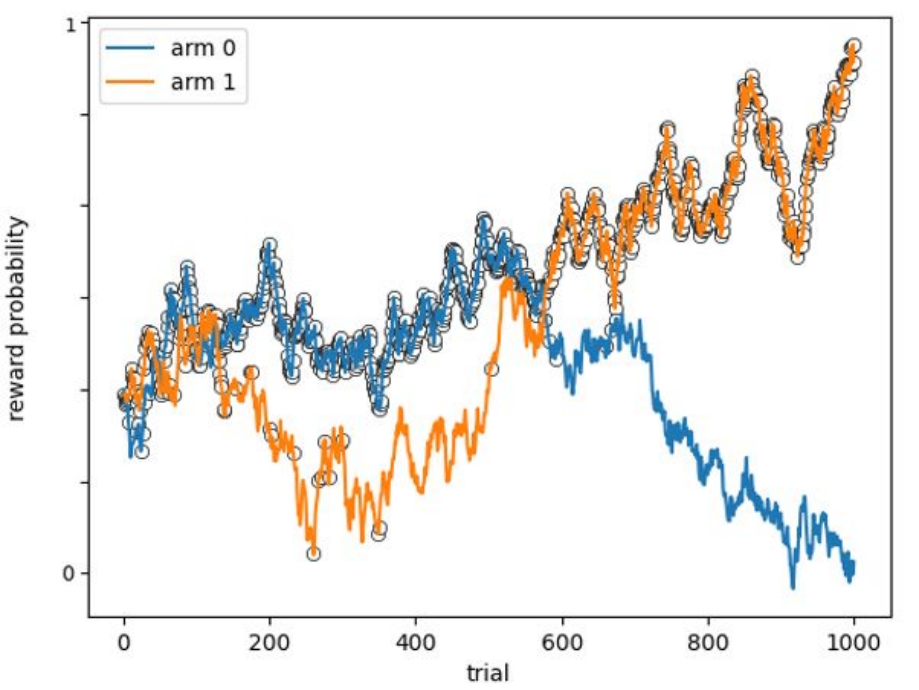
Multi Armed Bandit

Learn the action policy to maximize reward



Multi Armed Bandit

Learn the action policy to maximize reward



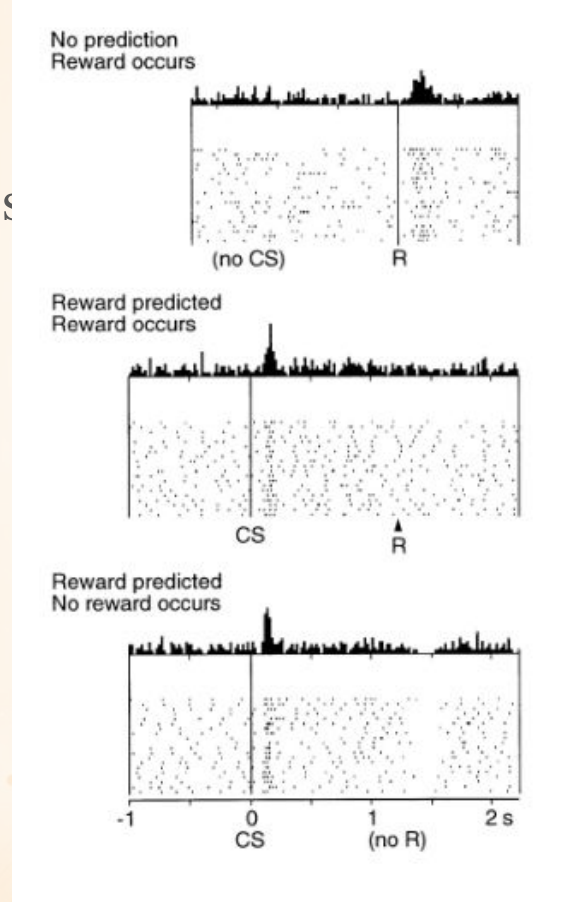
Multi Armed Bandit

Prediction Error.
Difference between observed and expected outcome

Multi-Armed Bandit

Prediction Error.

Difference between obs tcome



Schultz, 1998

Multi Armed Bandit

Prediction Error.

Difference between observed and expected outcome

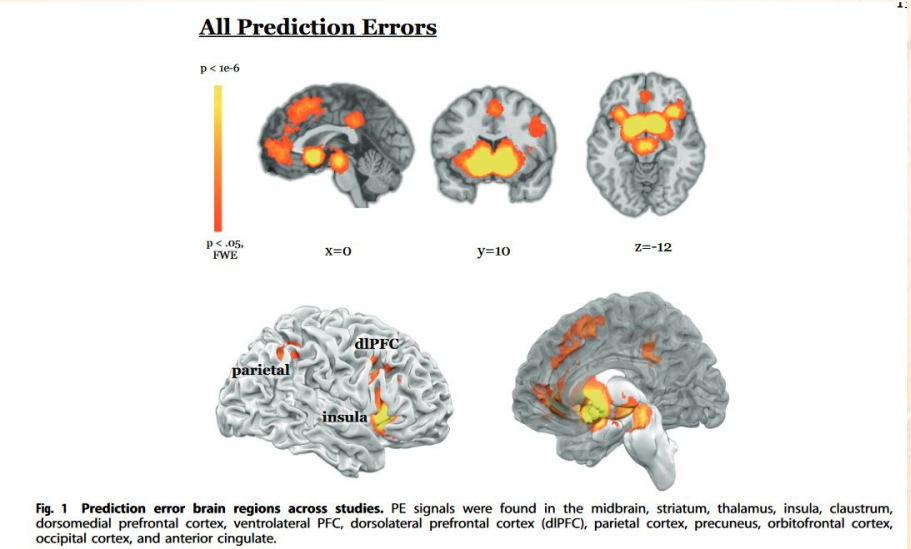
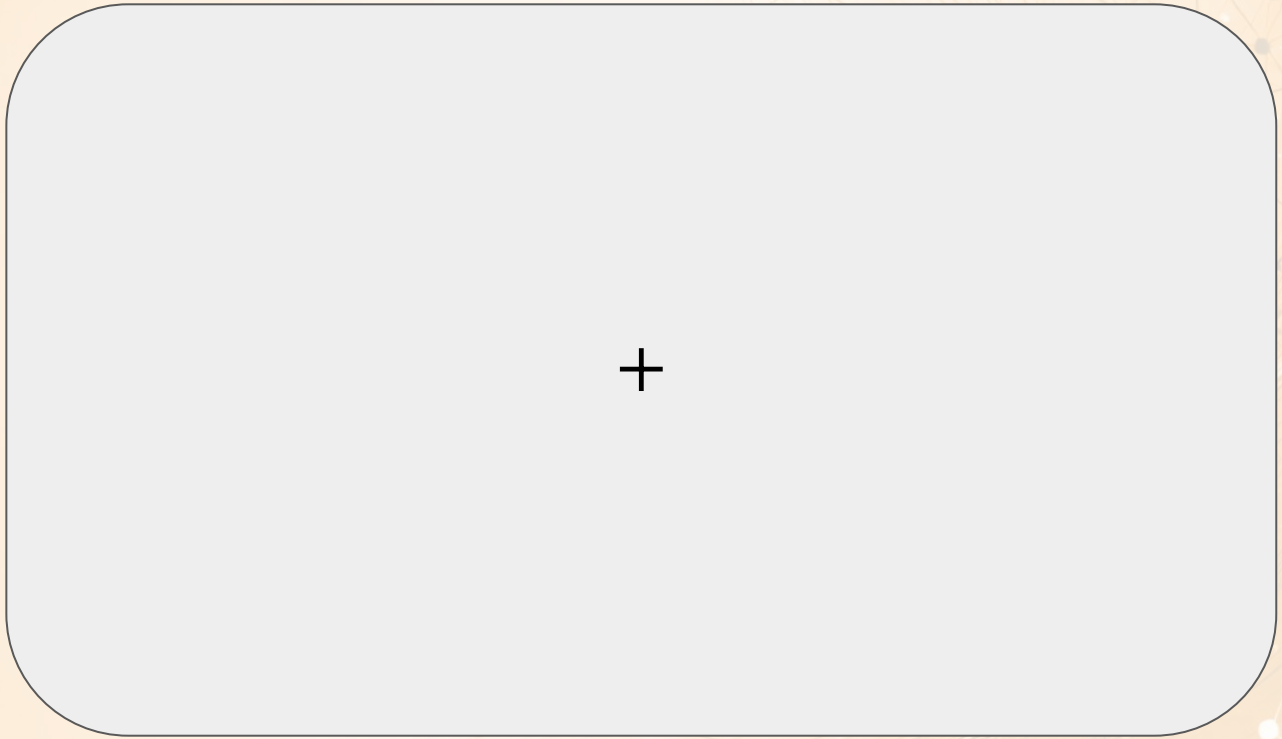


Fig. 1 Prediction error brain regions across studies. PE signals were found in the midbrain, striatum, thalamus, insula, claustrum, dorsomedial prefrontal cortex, ventrolateral PFC, dorsolateral prefrontal cortex (dlPFC), parietal cortex, precuneus, orbitofrontal cortex, occipital cortex, and anterior cingulate.

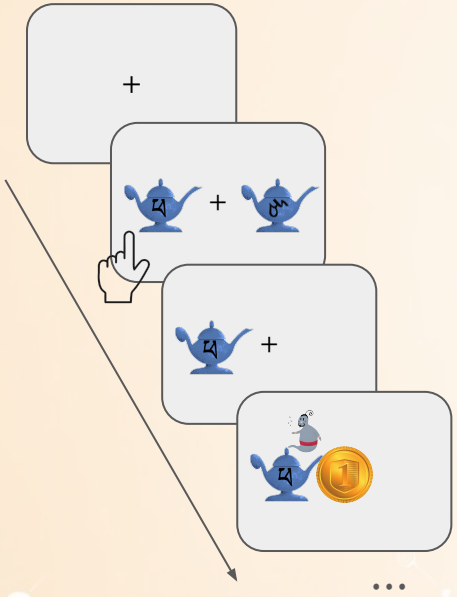
(Corlett, Mollick & Kober 2022)

Q-Learning Model



+

Q-Learning



Variables

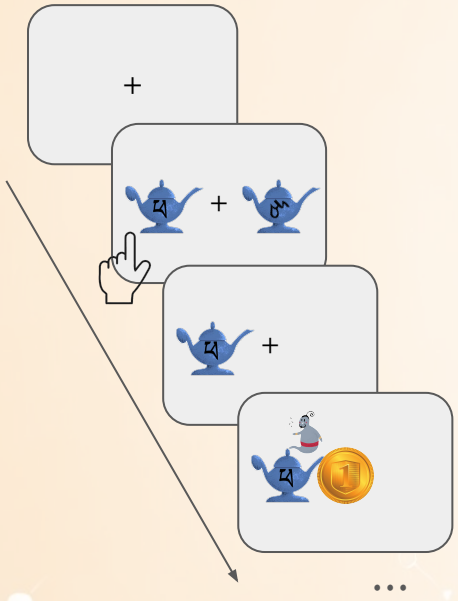
$a_t \in \{1, 2\}$ $r_t \in \{0, 1\}$ $Q_{t(a)}$

Parameters

α - learning rate

β - noise parameter

Q-Learning



Updating action values

$$Q_{t+1(a)} = Q_{t(a)} + \alpha \cdot (r_t - Q_{t(a)})$$

Action selection

$$P(a_t = 1) = \frac{e^{\beta \cdot Q_{t(a1)}}}{e^{\beta \cdot Q_{t(a1)}} + e^{\beta \cdot Q_{t(a2)}}}$$

softmax demo

Q-Learning

Step-by-step:

1

2

Q-Learning

Step-by-step:

1. simulating artificial behavior
2.

Q-Learning

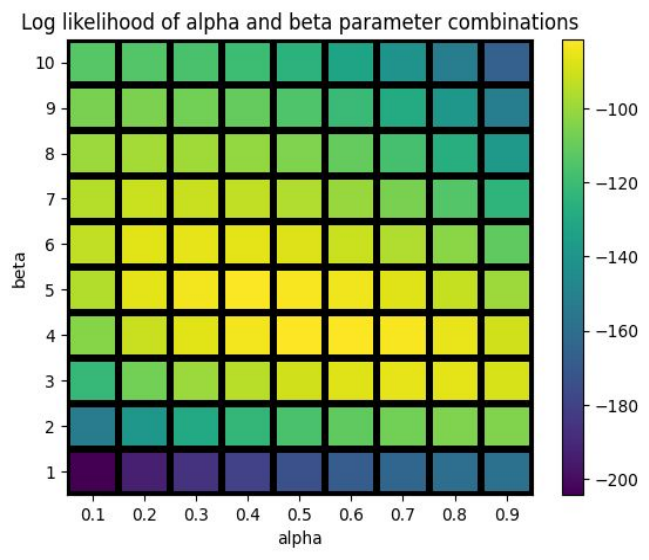
Step-by-step:

1. simulating artificial behavior
2. estimate parameters

Q-Learning

$\alpha = 0.5, \beta = 5$

Ntrials = 100



Ntrials = 1000