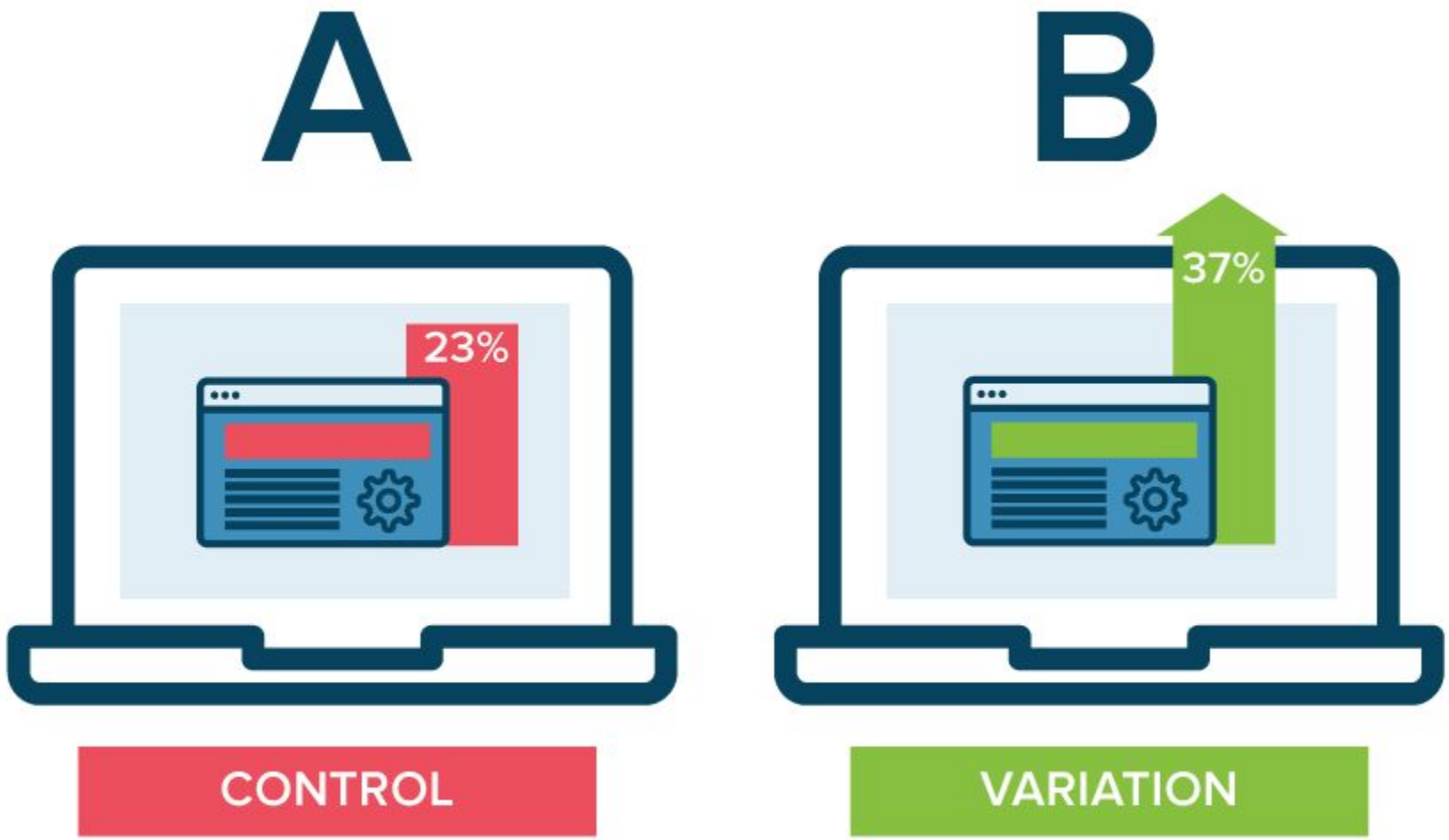
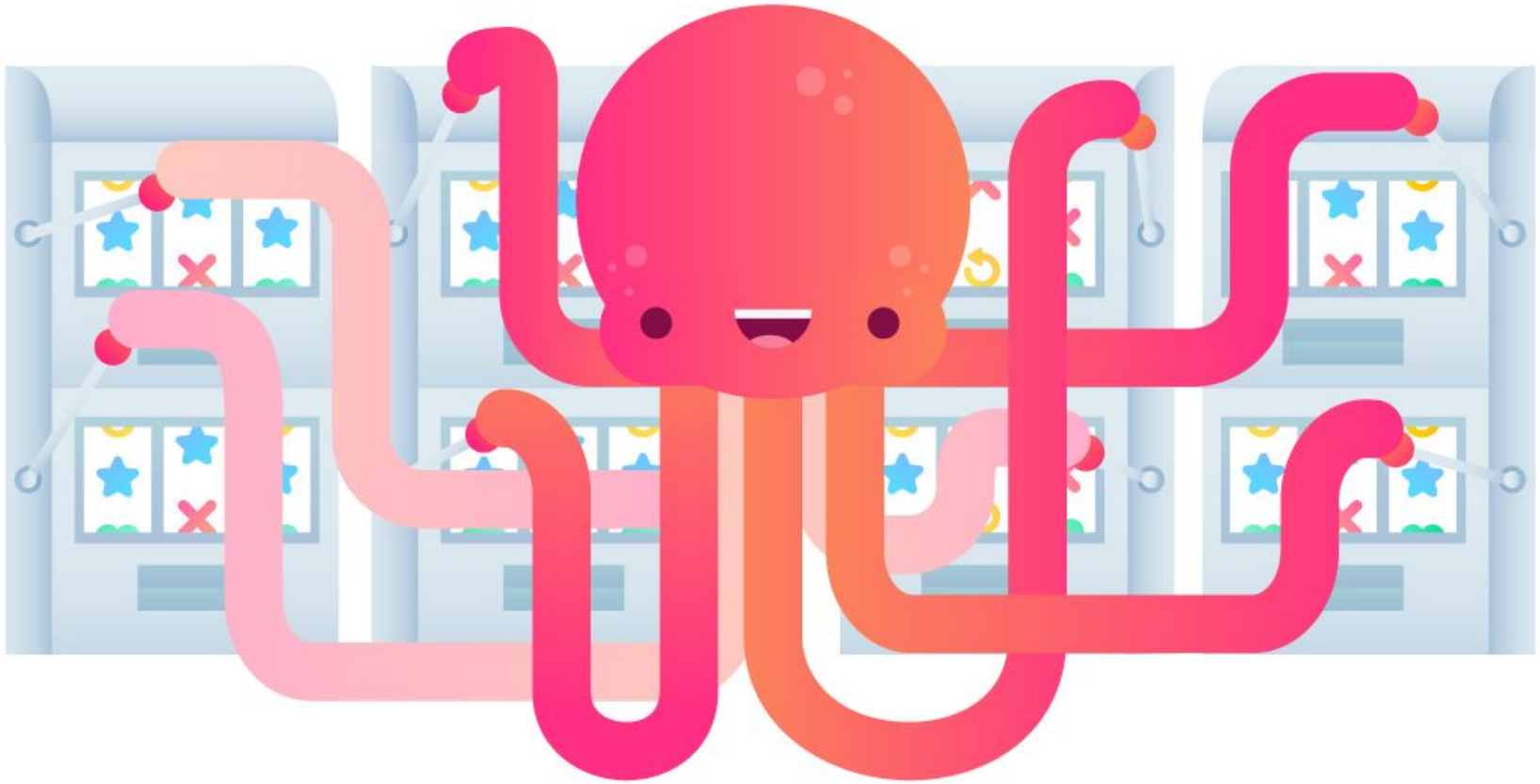
Multi-Armed Bandit vs A/B Test

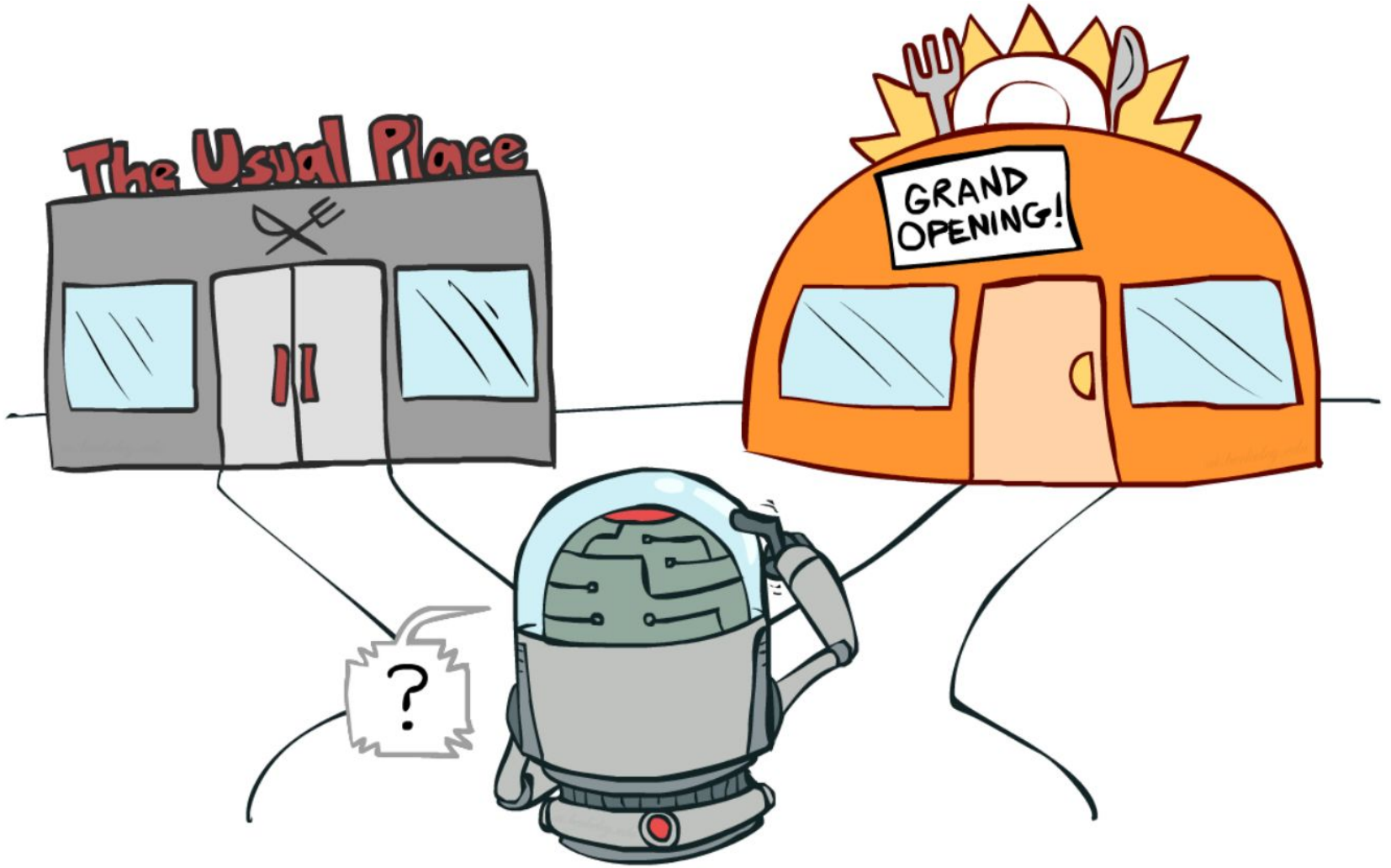
Classical A/B Testing



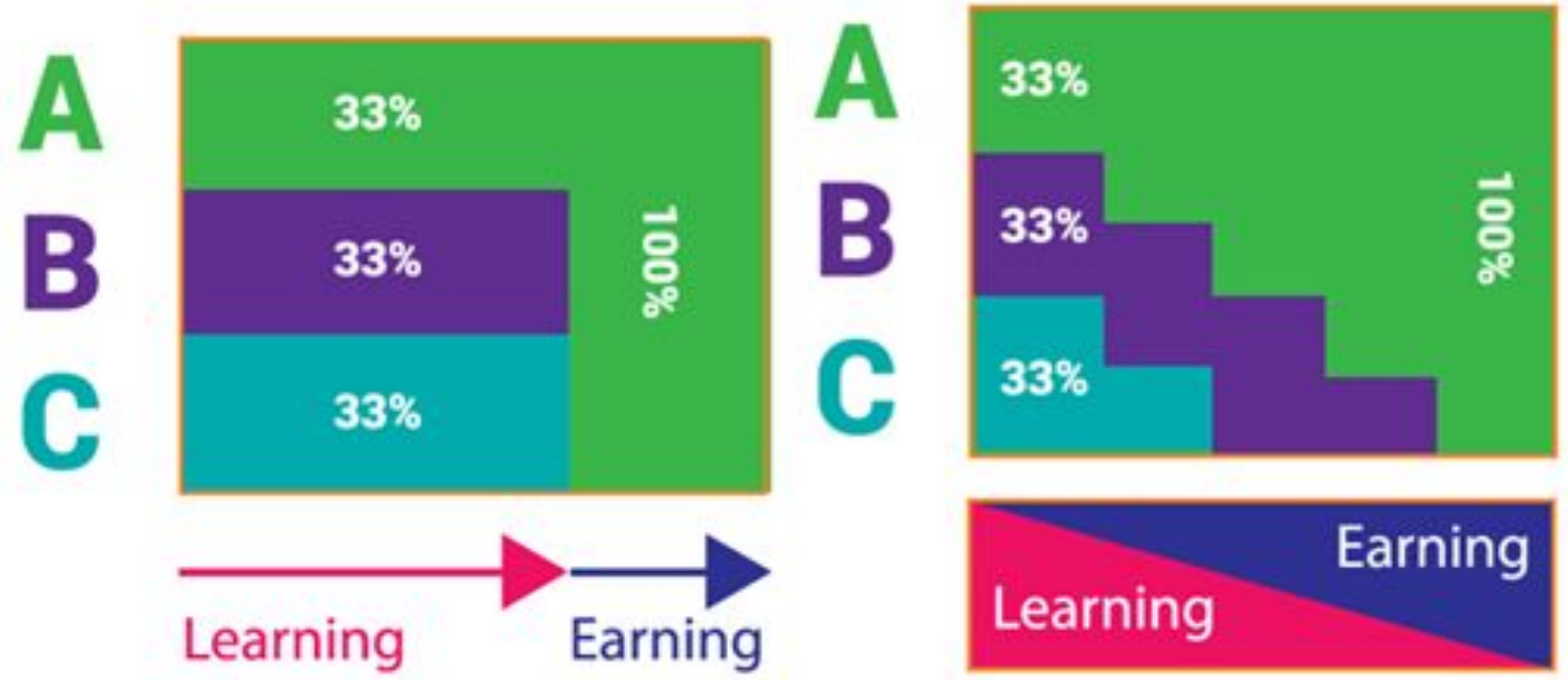
Multi-Armed Bandit



Exploration vs Exploitation



Earn or Learn



Agility

	KPI Rate	Sample Size needed for Significance
A/B Test	A: 5% B: 4%	223 Days
Multi-Armed Bandit	A: 5% B: 4%	31 Days

Epsilon-Greedy Agent

- Action
- Reward
- Action Value

A simple bandit algorithm

Initialize, for $a = 1$ to k:

$Q(a) \leftarrow 0$

$N(a) \leftarrow 0$

Loop forever:

$$A \leftarrow \begin{cases} \arg\max_a Q(a) & \text{with probability } 1-\varepsilon \quad \text{(breaking ties randomly)} \\ \text{a random action} & \text{with probability } \varepsilon \end{cases}$$

$R \leftarrow bandit(A)$

$N(A) \leftarrow N(A) + 1$

$Q(A) \leftarrow Q(A) + \frac{1}{N(A)}\left[R - Q(A)\right]$

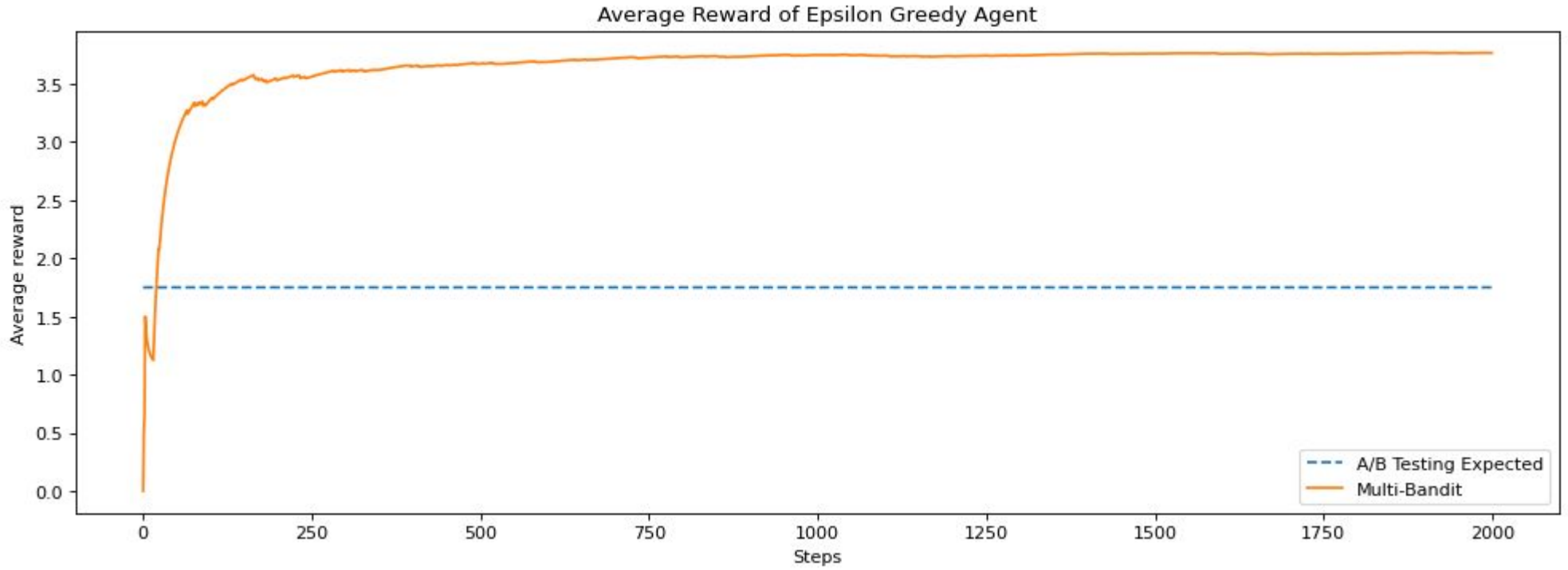
Parameters

- Epsilon Rate
- Initial State
- Learning Rate
- Step Definition
- Count of Steps

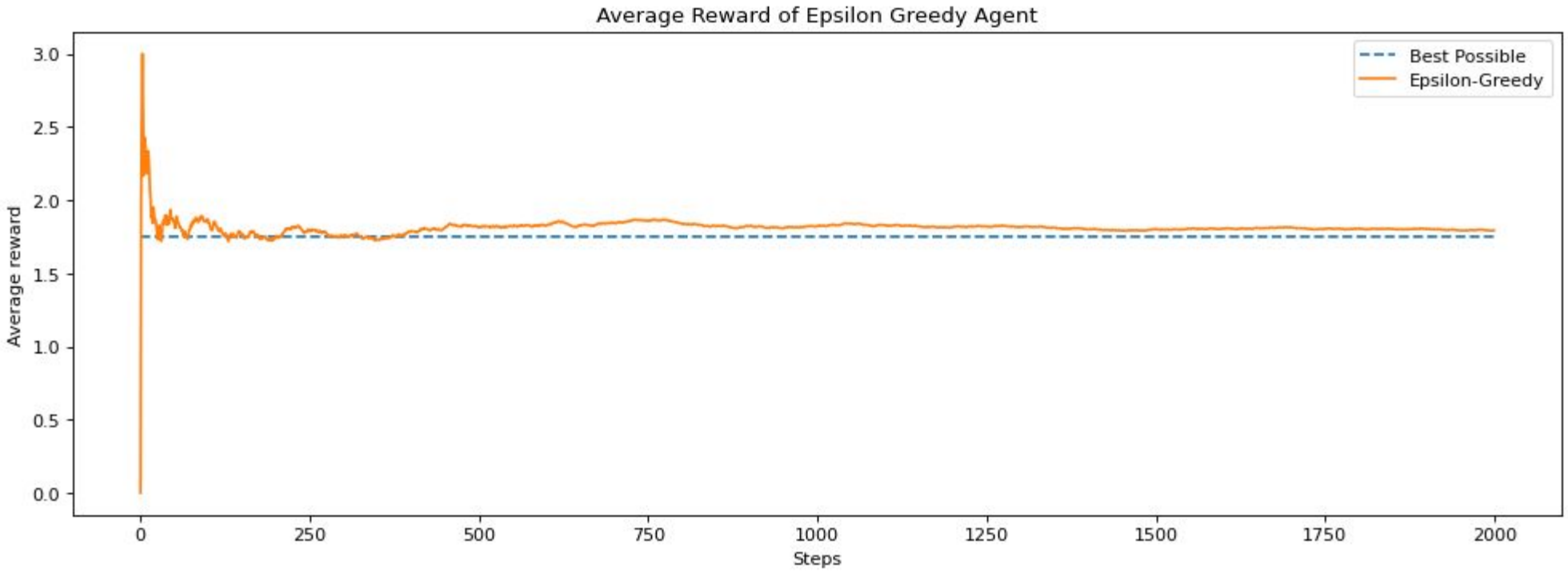
Simulation

[Related Notebook]

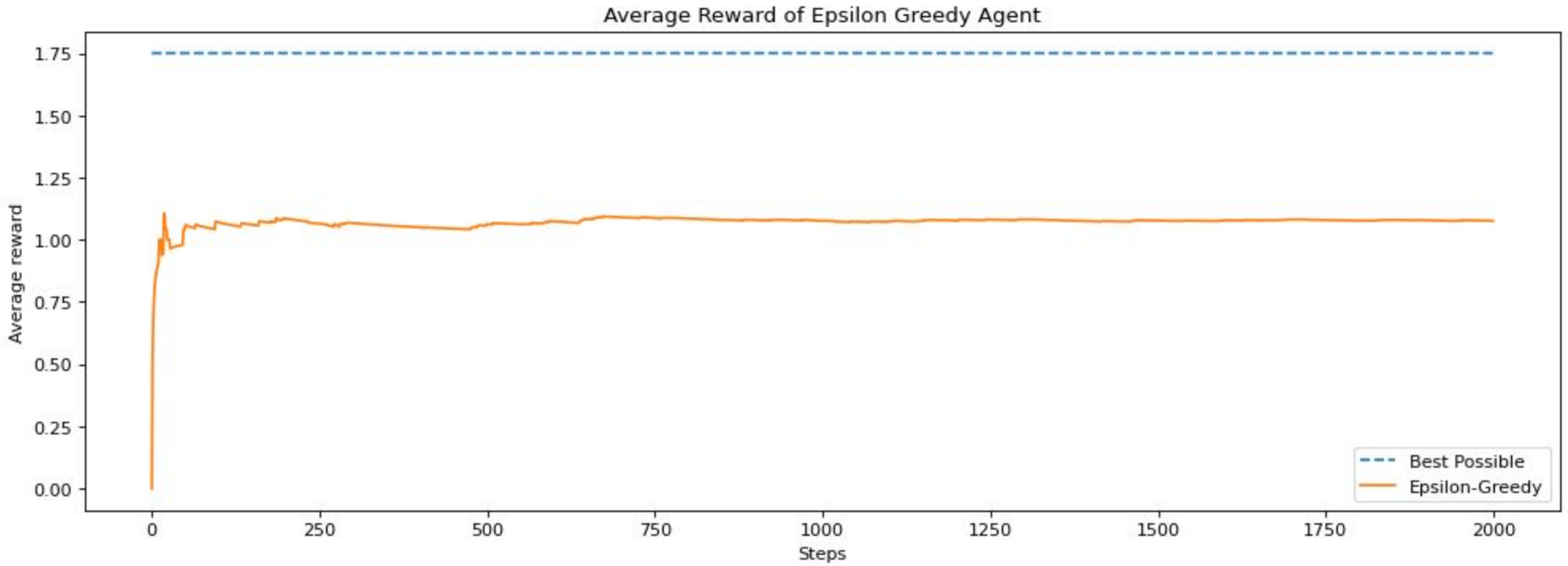
An agent has **Four** actions and their rewards are [0,1,4,2], We want to find best action by Epsilon-Greedy agent. Here assign set epsilon rate equal to 5% and learning rate equal to 0.2



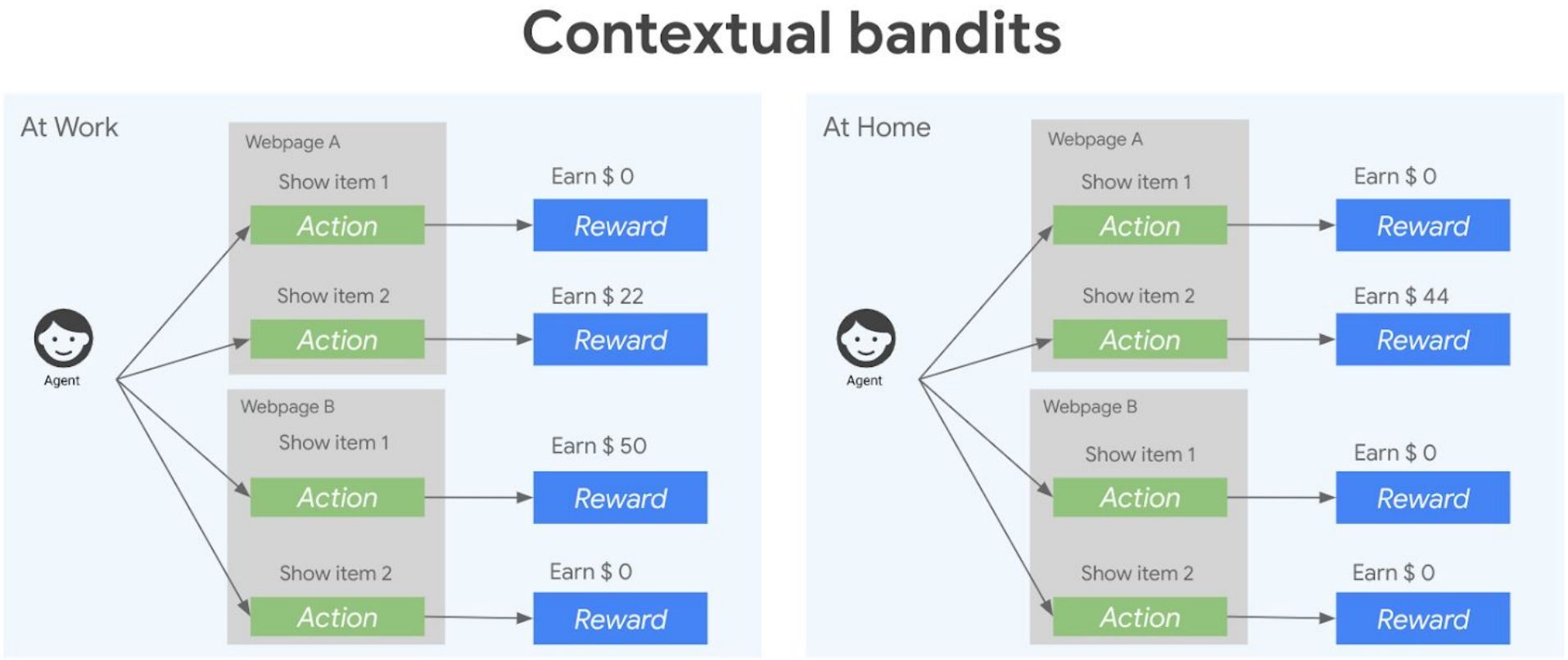
Epsilon = 100% (Same as A/B Test)



Not Suitable Learning Rate (=0.001 or 10)



Contextual Multi-Armed Bandit



Reward is conditional to the state of the environment:

Rewards vary according to the **state** or **context** that the agent is operating. The agent has more data points to analyze to decide which action to take.

Advantages

- The loss is affordable in some solutions
- Multi solutions not only two
- Agility to reach result in smaller sample size situation
- Automation
- Robustness to changing

Disadvantages

- Hard Implementation
- Complicated Analysis
- Inconsistent User Experience

What's the point?

- Headlines Testing
- Short-Term Campaigns
- Long-Term Dynamic Changes

Thank You!