CSCN8040 – Week 1

Learning Activities

Supervised Learning

Labeled data: (x, y)

data → x, y ← label

Size

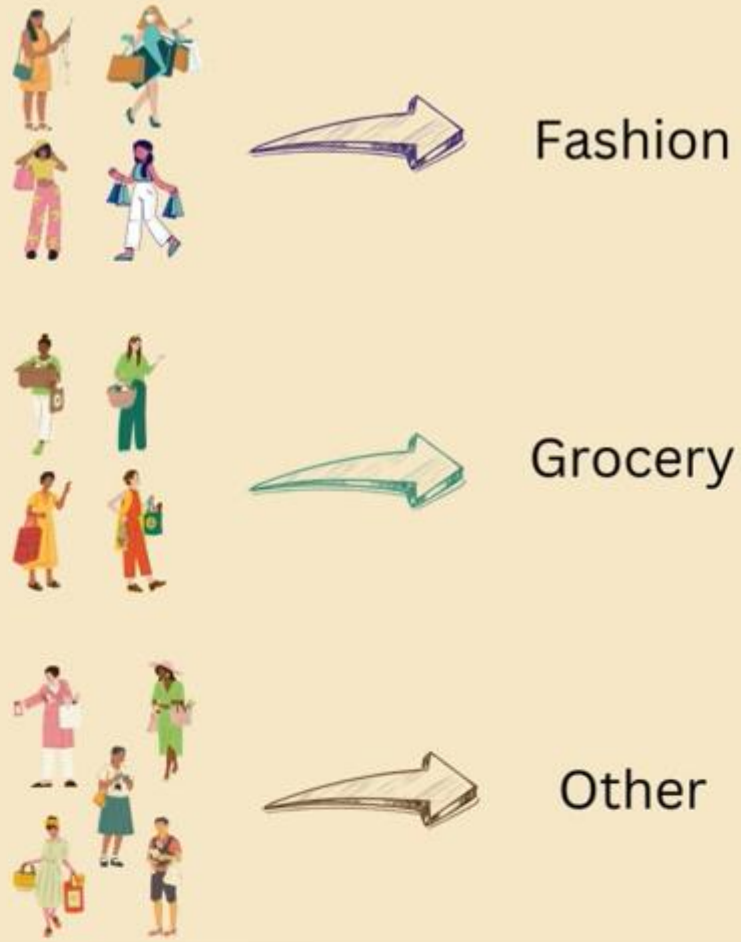
Location → $ 420,870

Number of rooms

Unsupervised Learning

Unlabeled data: x

x is data, no labels

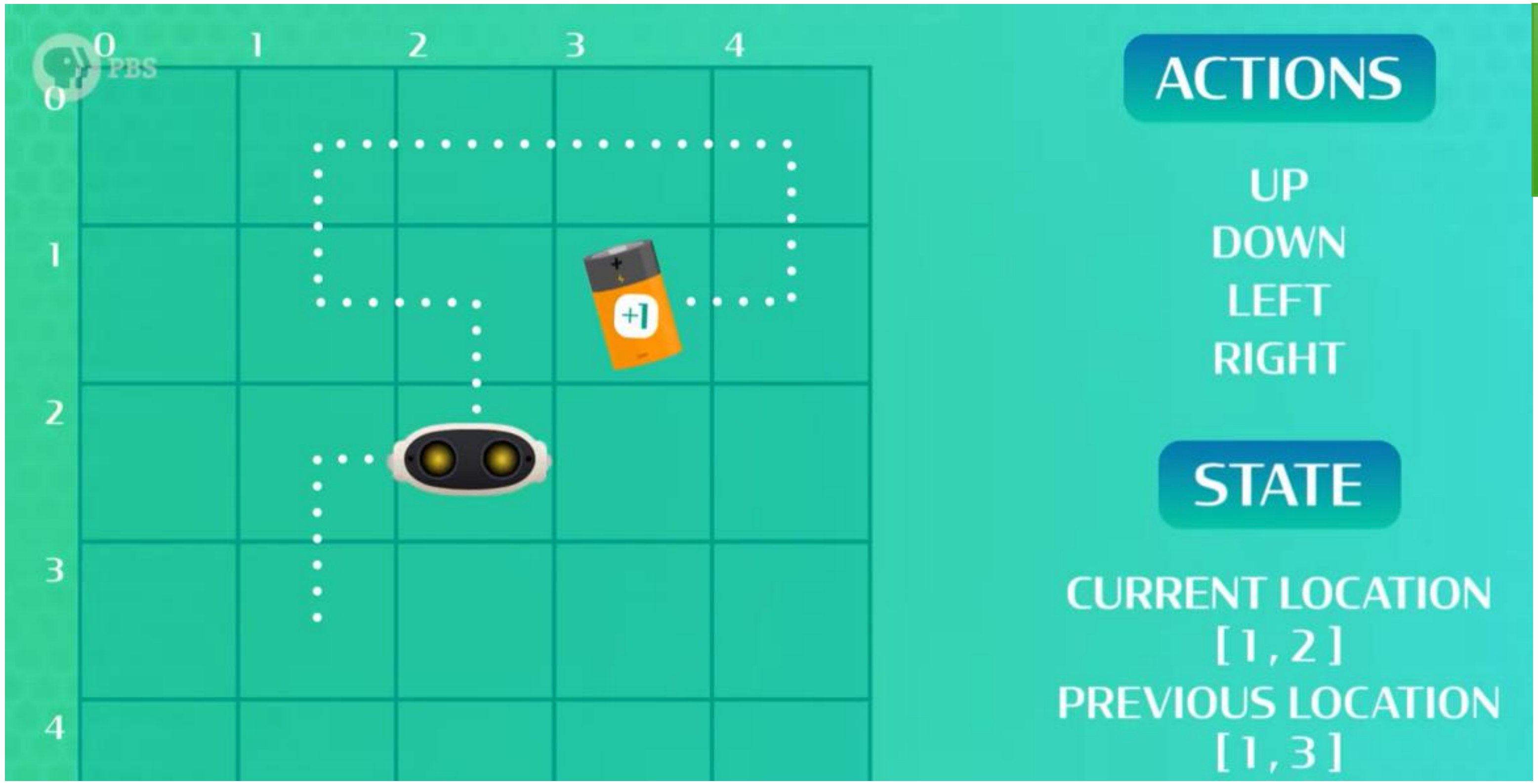


Reinforcement Learning

Data: state-action pairs

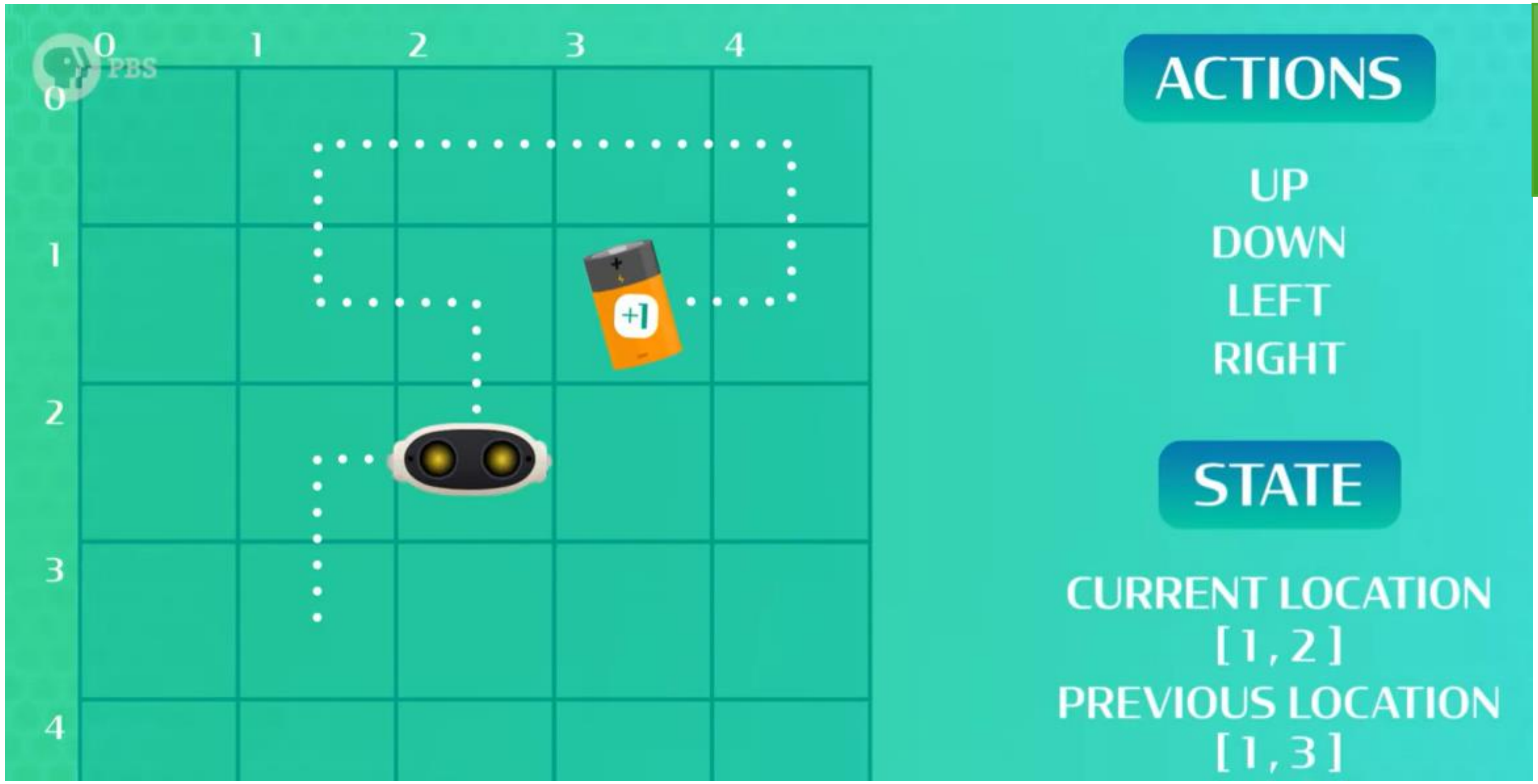
Source: https://www.youtube.com/@TheOpenSourceChannel https://youtu.be/nIgIv4IfJ6s?si=7asNBGkh6V81vdWO

State Machine



$$A = \{A_1, A_2, A_3, A_4\}$$

State Machine



ACTIONS

UP
DOWN
LEFT
RIGHT

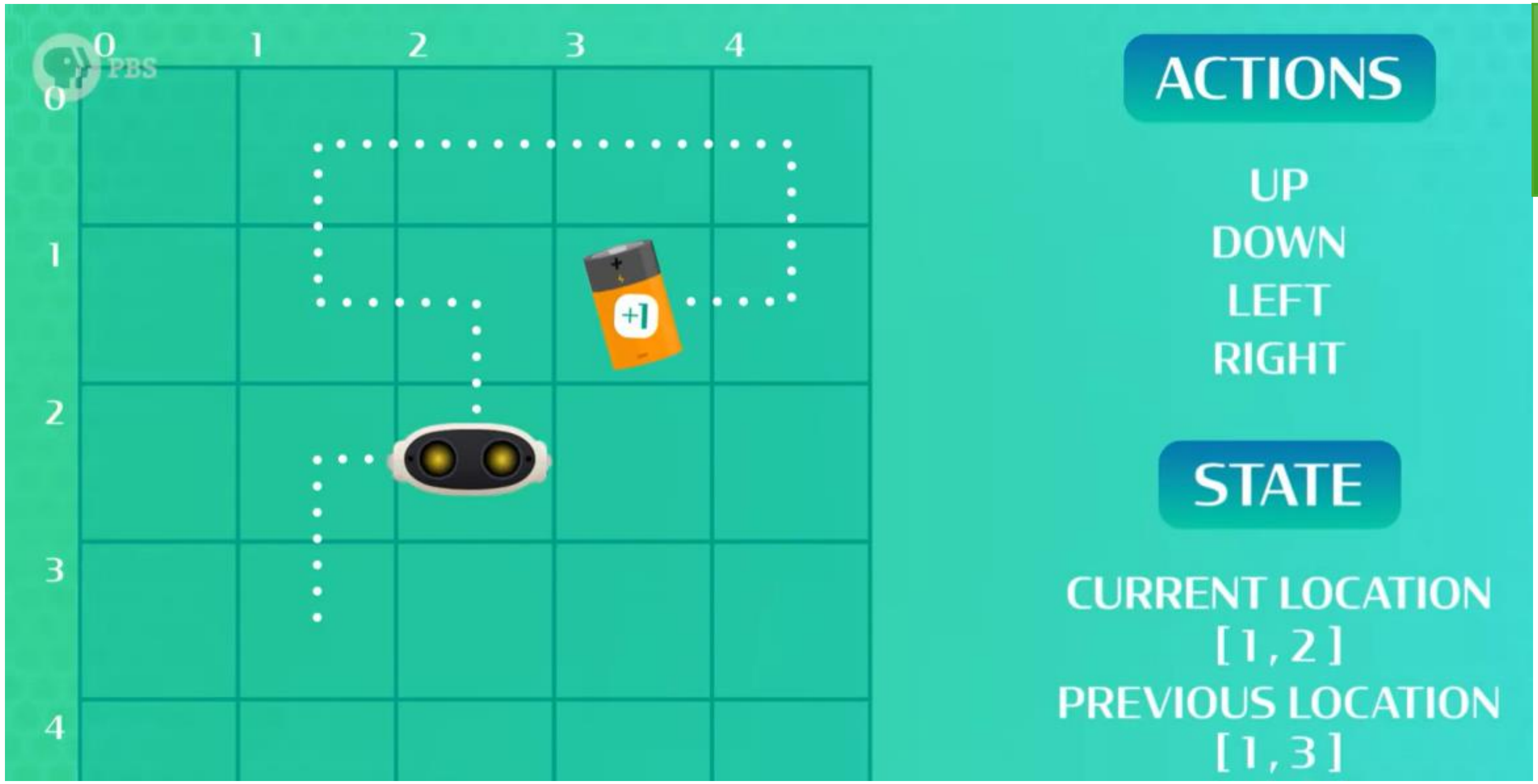
STATE

CURRENT LOCATION
[1, 2]
PREVIOUS LOCATION
[1, 3]

$$S = \{S_1, S_2, \dots, S_t\},\ t{=}11$$

and $\exists Sn|S = [x, y]$

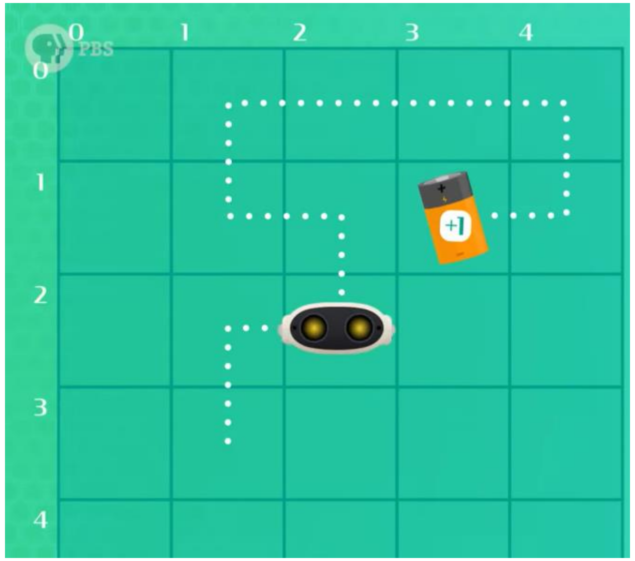
State Machine

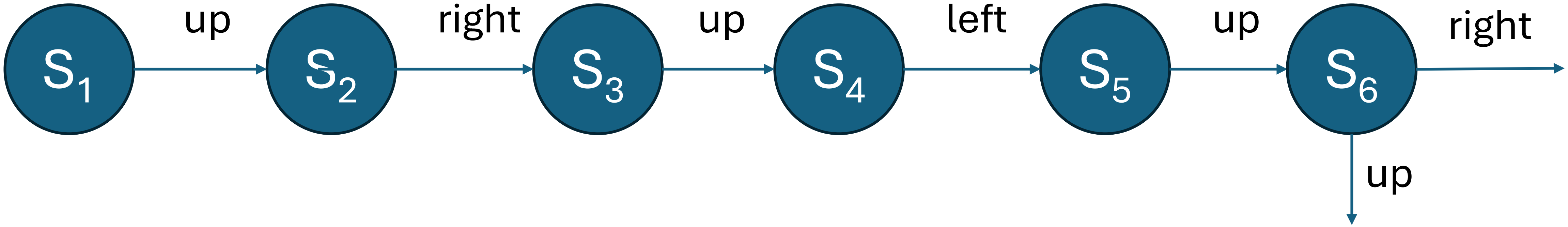


$$\partial(S_1, A_1) = S_2$$

where: $S_t \in S$ and: $A_t \in A$

State Diagram





Exercise

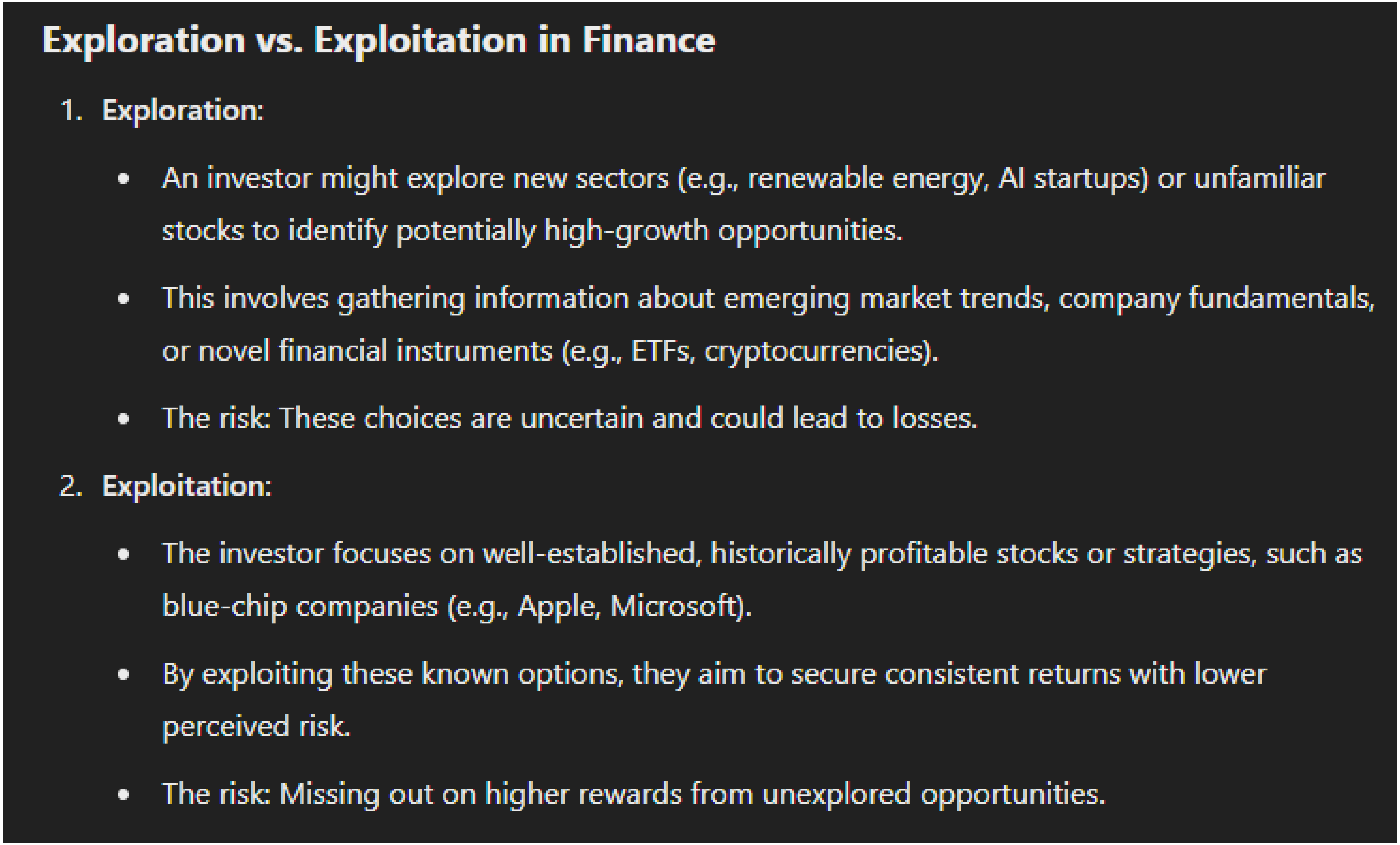
State Machine

Get together as teams

You are designing the state machine for a vending machine that accepts coins and dispenses snacks

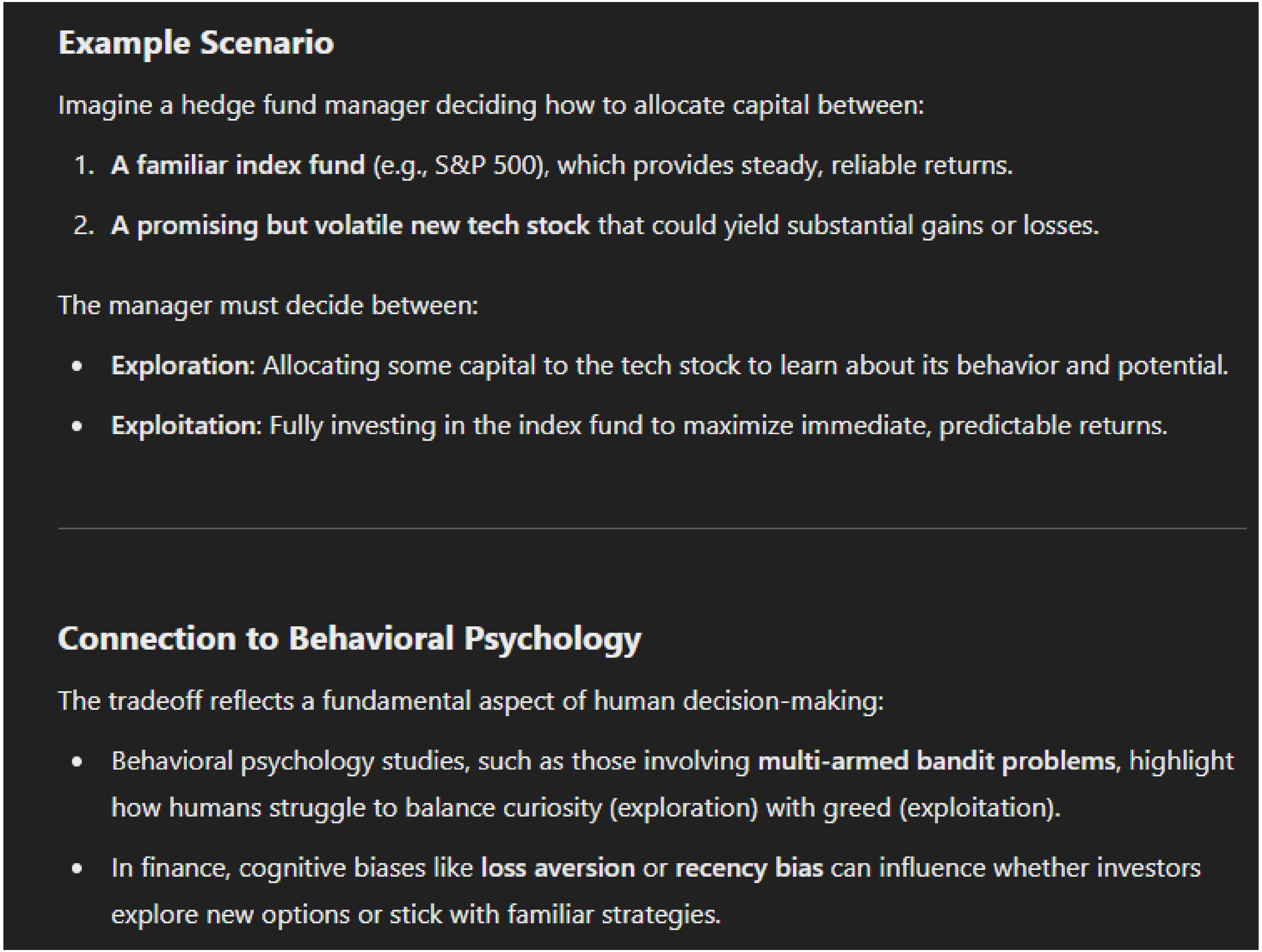
States: Idle, Coin Inserted, Selection Made, Dispensing.
Inputs: Insert coin, make selection, dispense item, return to idle.

→ Define states and name them.
→ Identify inputs that trigger transitions.
→ Draw a state diagram showing transitions between states.
→ Include arrows for transitions and label them with the input doing the transition.



Exploration vs. Exploitation in Finance

1. **Exploration:**
   - An investor might explore new sectors (e.g., renewable energy, AI startups) or unfamiliar stocks to identify potentially high-growth opportunities.
   - This involves gathering information about emerging market trends, company fundamentals, or novel financial instruments (e.g., ETFs, cryptocurrencies).
   - The risk: These choices are uncertain and could lead to losses.
2. **Exploitation:**
   - The investor focuses on well-established, historically profitable stocks or strategies, such as blue-chip companies (e.g., Apple, Microsoft).
   - By exploiting these known options, they aim to secure consistent returns with lower perceived risk.
   - The risk: Missing out on higher rewards from unexplored opportunities.



Example Scenario

Imagine a hedge fund manager deciding how to allocate capital between:

1. **A familiar index fund** (e.g., S&P 500), which provides steady, reliable returns.
2. **A promising but volatile new tech stock** that could yield substantial gains or losses.

The manager must decide between:

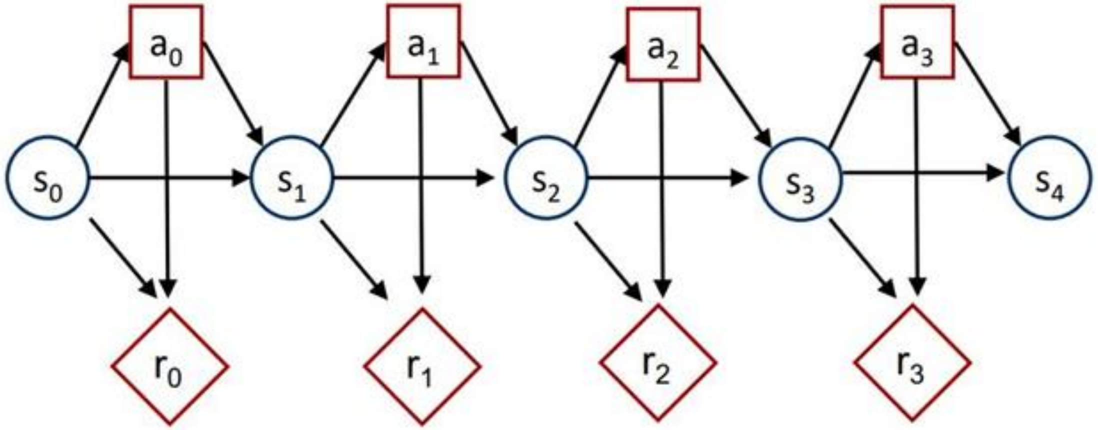
- **Exploration:** Allocating some capital to the tech stock to learn about its behavior and potential.
- **Exploitation:** Fully investing in the index fund to maximize immediate, predictable returns.

Connection to Behavioral Psychology

The tradeoff reflects a fundamental aspect of human decision-making:

- Behavioral psychology studies, such as those involving **multi-armed bandit problems**, highlight how humans struggle to balance curiosity (exploration) with greed (exploitation).
- In finance, cognitive biases like **loss aversion** or **recency bias** can influence whether investors explore new options or stick with familiar strategies.

Markov Decision Process (MDPs)



$$p(s', r|s, a) \doteq \Pr\{S_t = s', R_t = r'|S_{t-1} = s, A_{t-1} = a\}$$

This is the transition probability in an MDP. It explains the likelihood of moving to a new state s' and receiving a reward r, given that the current state is s and the chosen action is a.

States S: Think of a robot in a maze. Each location in the maze is a state.

Actions A: The robot can move up, down, left, or right.

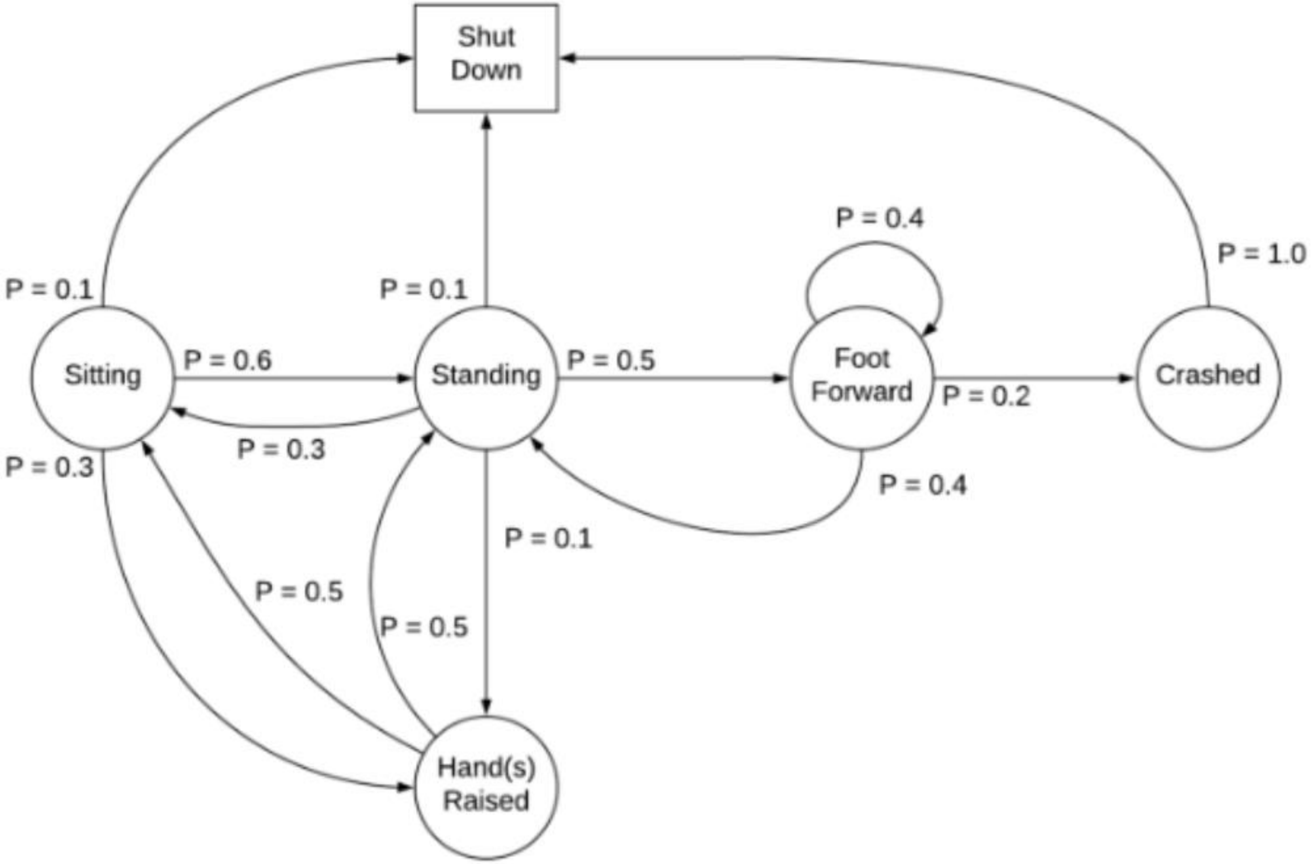
Transition Probability: The formula tells us how likely the robot is to end up in a new location (s') with a specific reward (r) if it takes an action (a) from a given starting point (s).

Markov Decision Process (MDP)

$$\sum_{s' \in S} \sum_{r \in R} p(s', r \mid s, a) = 1 \; \forall \; s \in S, a \in A(s)$$

The sum of all possible probabilities of transitioning to any state s' and receiving any reward r, given a state s and action a, must equal 1.

This formula ensures probability conservation. In other words, when you take an action, something is guaranteed to happen—either you transition to a specific state with a reward or to another state, but the total probability over all possibilities must add up to 1.



Markov Decision Process (MDP)

Try at home

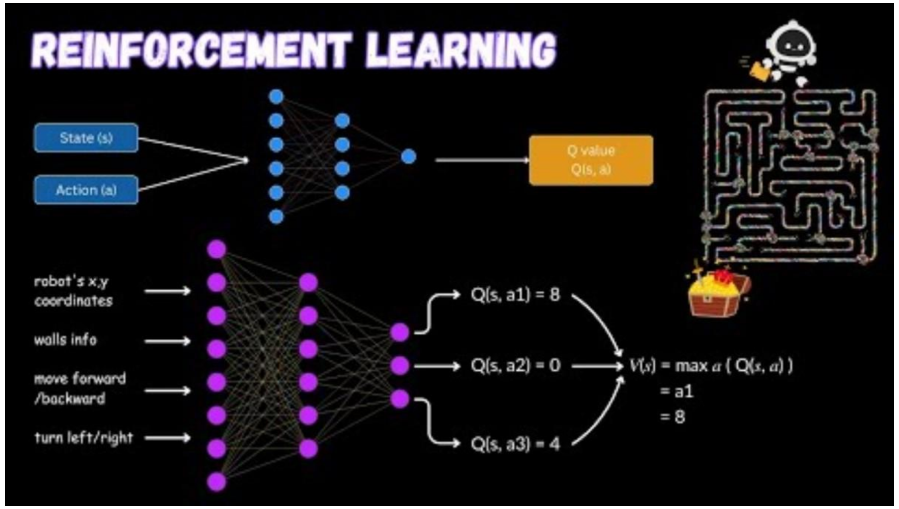
1. **Maze Navigation Game:**
   - Create a simple grid maze and assign probabilities for moving between grid cells (states) based on different actions (e.g., moving left, right, up, down).
   - Add a small reward (e.g., +1) or penalty (e.g., -1) for certain transitions.

2. **Hands-on Calculation:**
   - Calculate p(s’, r | s, a) for given state-action pairs.
   - Verify that the probabilities for all possible transitions sum to 1 for any specific state and action.

Quiz next week

Resources

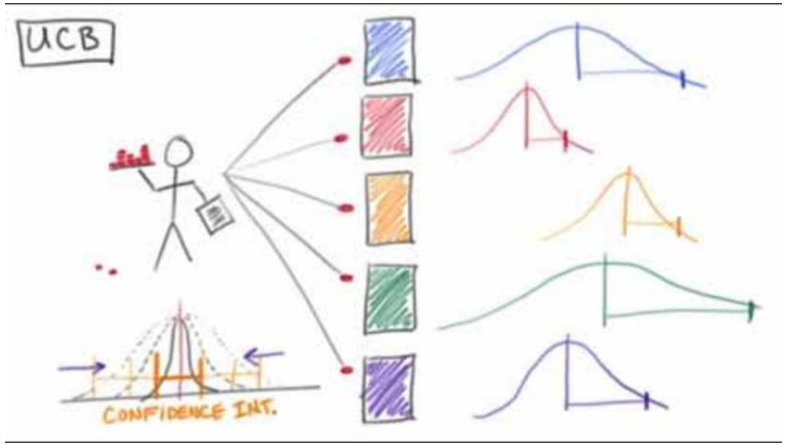
https://youtu.be/vufTSJbzKGU?si=5Mn5FuXlyPbi5LOf



https://youtu.be/vw0Zy_oCWxE?si=wasqU0gpbvLeI1pH

Resources (2)

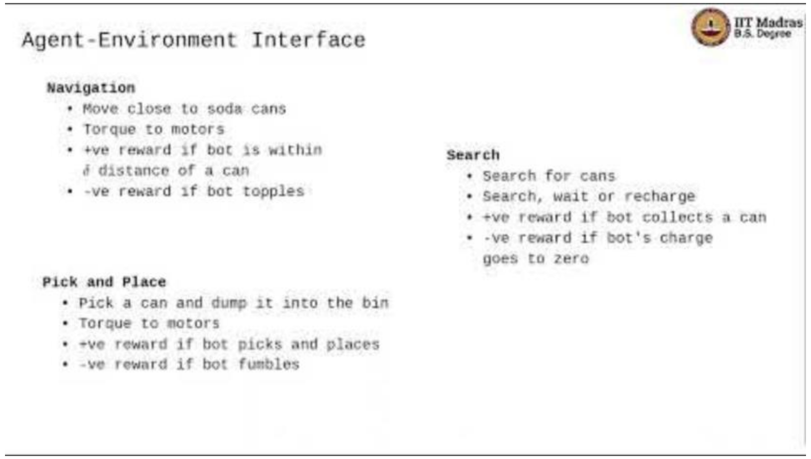
The Multi-Armed Bandit Problem



https://youtu.be/bkw6hWvh_3k?si=jQBl9Aejj_fE63VL

https://youtu.be/my207WNoeyA?si=j9QnT_dyxhRtwQc9

Resources (3)



Agent-Environment Interface

Navigation

- Move close to soda cans
- Torque to motors
- +ve reward if bot is within δ distance of a can
- -ve reward if bot topples

Pick and Place

- Pick a can and dump it into the bin
- Torque to motors
- +ve reward if bot picks and places
- -ve reward if bot fumbles

Search

- Search for cans
- Search, wait or recharge
- +ve reward if bot collects a can
- -ve reward if bot's charge goes to zero

https://youtu.be/CHpR3KVMLzU