FPT Education
FPT UNIVERSITY

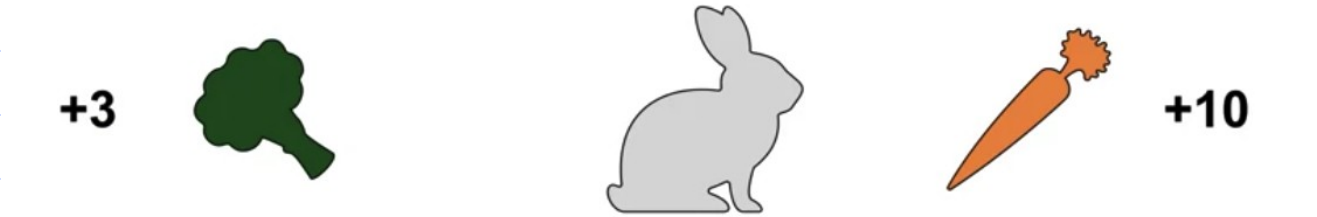
Markov Decision Processes

Objectives

- Understand Markov Decision Process
- Describe how the dynamic of MDP and defined

Markov Decision Processes

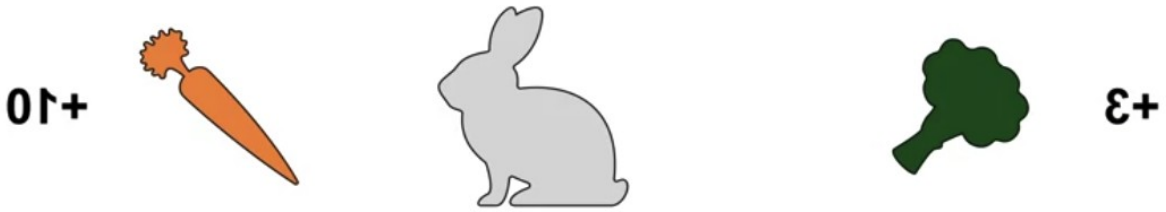
- Example: rabbit is wandering around in a field looking for food in a situation where
  - Carrot on its right: eating the carrot generates a reward of plus 10
  - Broccoli on its left: eating broccoli only plus three



- Rabbit prefers carrots→ go right!

Markov Decision Processes

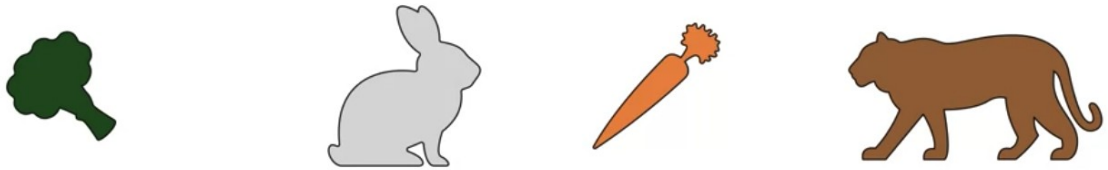
- Example: but what if later the rabbit finds itself in another situation, where there's broccoli on the right and carrot on the left. Here, the rabbit would clearly prefer to go left instead of right.



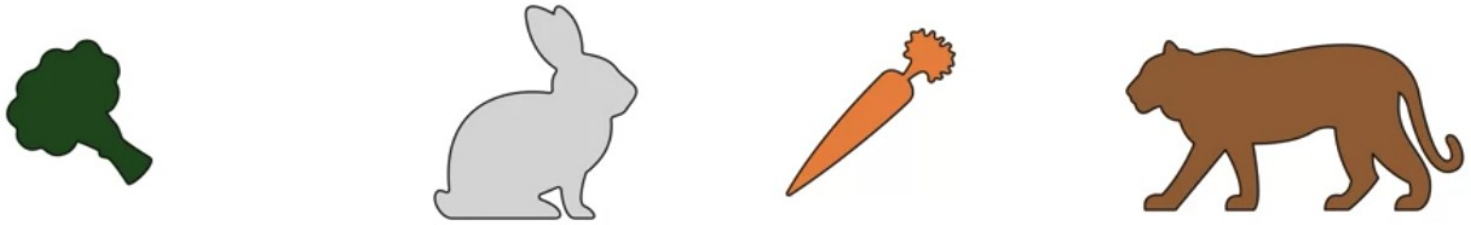
- The k-Armed Bandit problem does not account for the fact that different situations call for different actions.

Markov Decision Processes

- Other example:
  - the rabbit would like to go right to get the carrot. However, on the right of the carrot there is a tiger. If the rabbit moves right, it gets to eat the carrot. But afterwards it may not be fast enough to escape the tiger.
  - If we account for the long-term impact of our actions, the rabbit should go left and settle for broccoli to give itself a better chance to escape



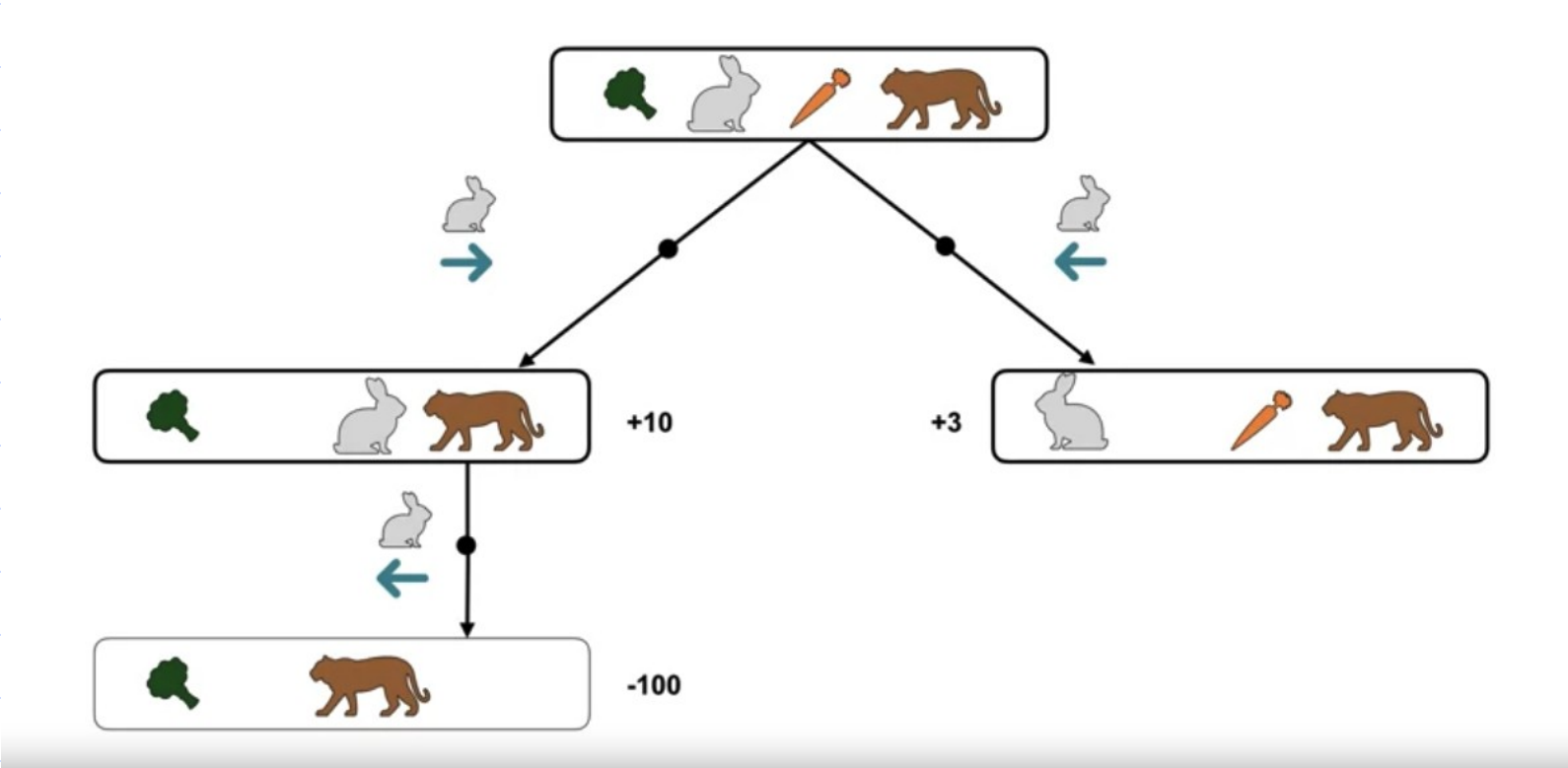
Markov Decision Processes



- A bandit rabbit would only be concerned about immediate reward and so it would go for the carrot.
- But a better decision can be made by considering the long-term impact of our decisions
- What is your action ?

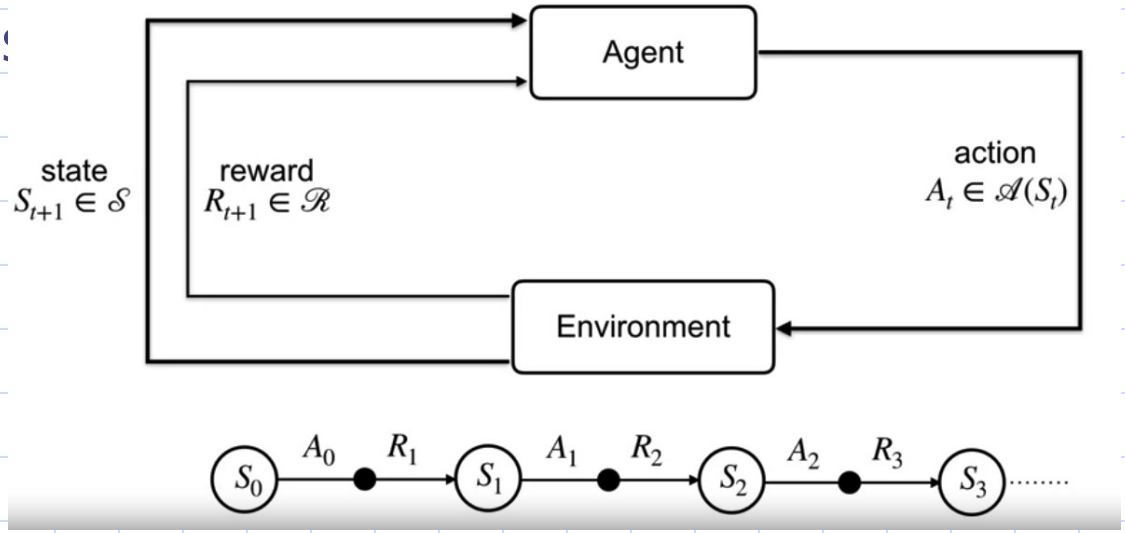
Markov Decision Processes

- How the situation changes as the rabbit takes



Markov Decision Processes

- Markov Decision Processes (MDPs) are mathematical frameworks used to model decision-making problems in which an agent interacts with an environment over a sequence of discrete time steps



Markov Decision Processes

- Markov Decision Processes terms:
  - **Agent**: A reinforcement learning agent is the entity that we are training to make correct decisions.
  - **Environment**: The environment is the surroundings with which the agent interacts.
  - **State**: The state defines the current situation of the agent
  - **Action**: The choice that the agent makes at the current time step.
  - **Policy**: A policy is the thought process behind picking an action.

Markov Decision Processes

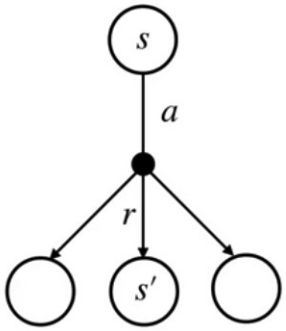
- Markov Decision Processes terms in a Robotic navigation task
  - **Agent**: the autonomous robot that moves through an environment to reach a destination.
  - **Environment**: In a grid world scenario where the robot navigates, the grid layout, obstacles, and target destination.
  - **State**: each cell can represent a state, indicating the position of the robot in the grid..

Markov Decision Processes

- Markov Decision Processes terms in a Robotic navigation task
  - **Action**: actions may include moving up, down, left, or right, or staying in the same position
  - **Policy**: a policy could specify that the robot should move towards the target destination while avoiding obstacles.

The dynamics of an MDP

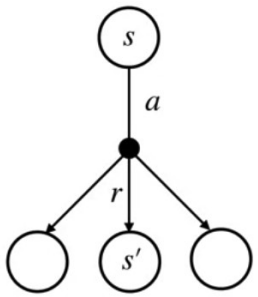
- When the agent takes an action in a state, there are many possible next states and rewards.
  - The transition dynamics function P
  - Given a state S
  - Action a,
  - P tells us the joint probability of next state S prime and reward are



$$p(s', r \mid s, a)$$

The dynamics of an MDP

- Since p is a probability distribution, it must be non-negative and it's sum over all possible next states and rewards must equal one.
- It means that the present state is sufficient and remembering earlier states would not improve predictions abou



$$p(s', r \mid s, a)$$

$$p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$$

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

Summary

- Understand Markov Decision Process
- Describe how the dynamic of MDP and defined

Q & A