Workshop 7

Recap: Classical Planning Problem

Not every problem belongs to classical planning problem

Deterministic action: S – a -> S’

- Every action only has a certain outcome, and you know what that outcome will be
- Counterexample: coin toss -> probabilistic actions
- Single-agent
- Static environment
- ……

Other action types

- **Probabilistic:** We could possibly end up in more than one state, and we know the probability distribution of these states (Example: Toss a fair coin)
- **Non-deterministic:** We know all possible outcome, but not the probability distribution
- **Stochastic:** limited info about possible outcomes

MDP problem

- Still use model-based approach to solve it

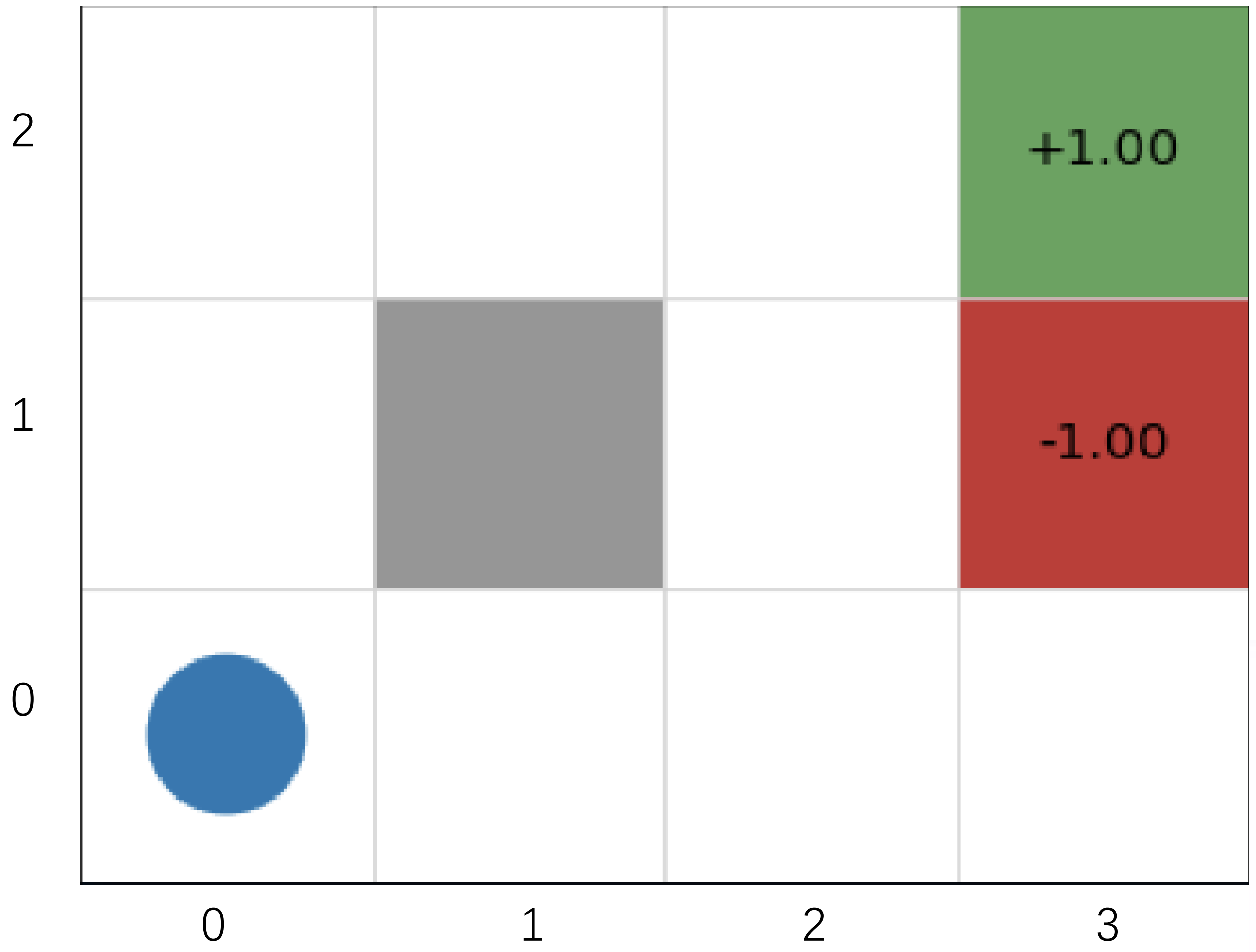
2 Models:

- **Goal-cost MDP model**: with a set of specific goal state, intend to achieve some goals, objective: minimize our cost to the goal
- **Discounted reward MDP model**: don’t have goal state, have terminal state instead, objective: maximize the reward

Solvers:

- Policy Iteration

Lecture Example



0.8 succ
0.1 slip left
0.1 slip right

Representations

S = {<x,y> | x belongs to (0,3), y belong to {0,2}} U {s_t} \ (1,1)}

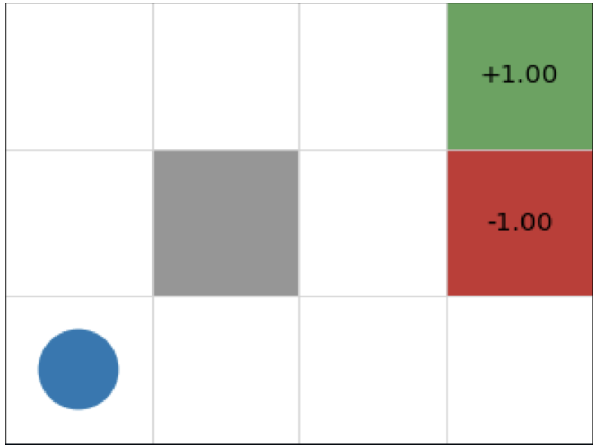
s0 = <0, 0>   S_T = {s_t}

Action function:

A(s_t) = {}

A(s) = {N,W,E,S}

except A((3,2)) = A((3,1)) = {exit}



Reward function:

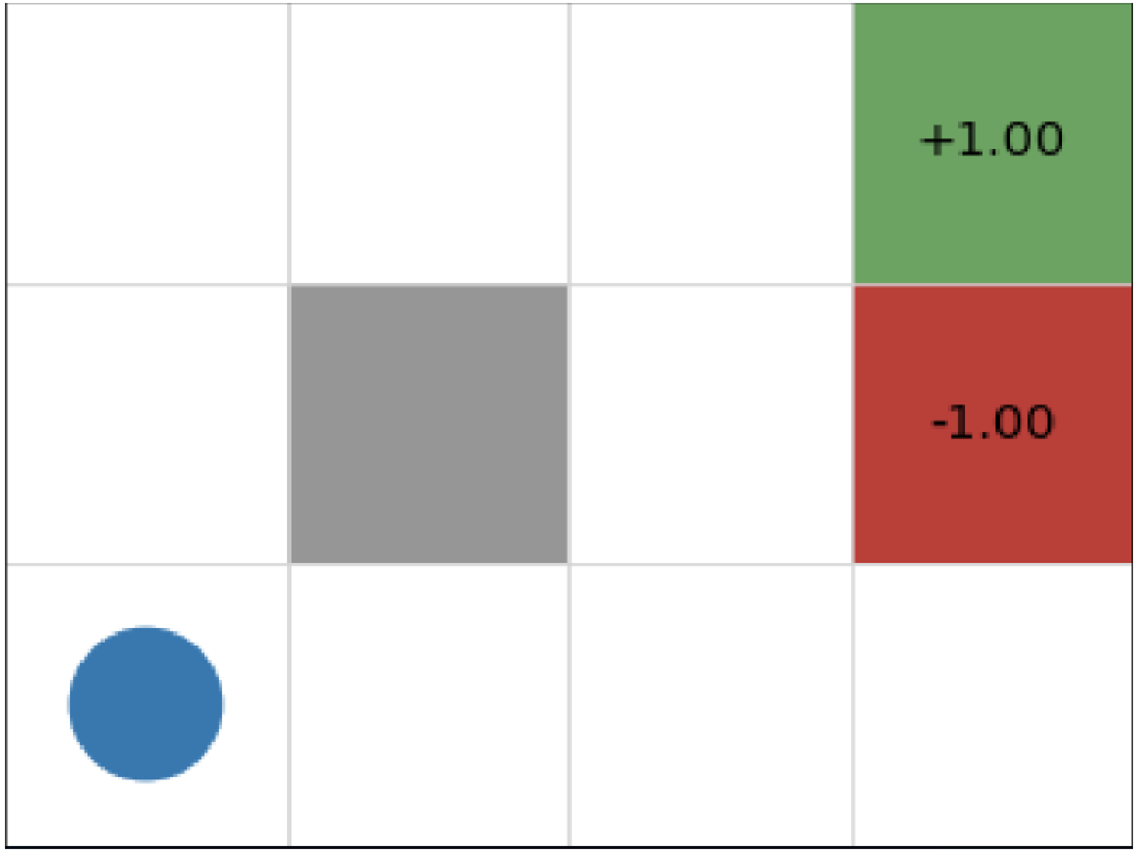
r(s, a) = 0 for any s, belong to S, a belongs to A

Except r((3,2), exit) = +1

And r((3,1), exit) = -1

Discount factor 0 < γ <1

Probability Distribution



Probability distribution for exit action

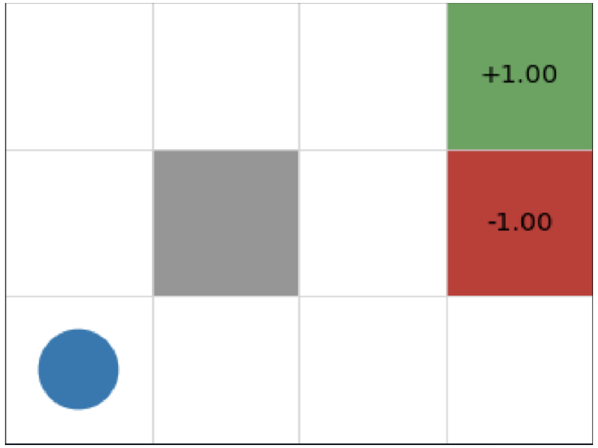
- P_exit(s_t | (3, 2)) = 1
- P_exit(s_t | (3, 1)) = 1
- P_exit(s' | any s except above 2 state) = 0

Probability Distribution for North action

P_N((x', y') | (x, y)) =

Common case

- Successful: If x',y' == x, min(2, y+1) then p = 0.8
- Slip Right: If x',y' == min(3, x+1), y then p = 0.1
- Slip Left: If x',y' == max(0, x-1), y then p = 0.1

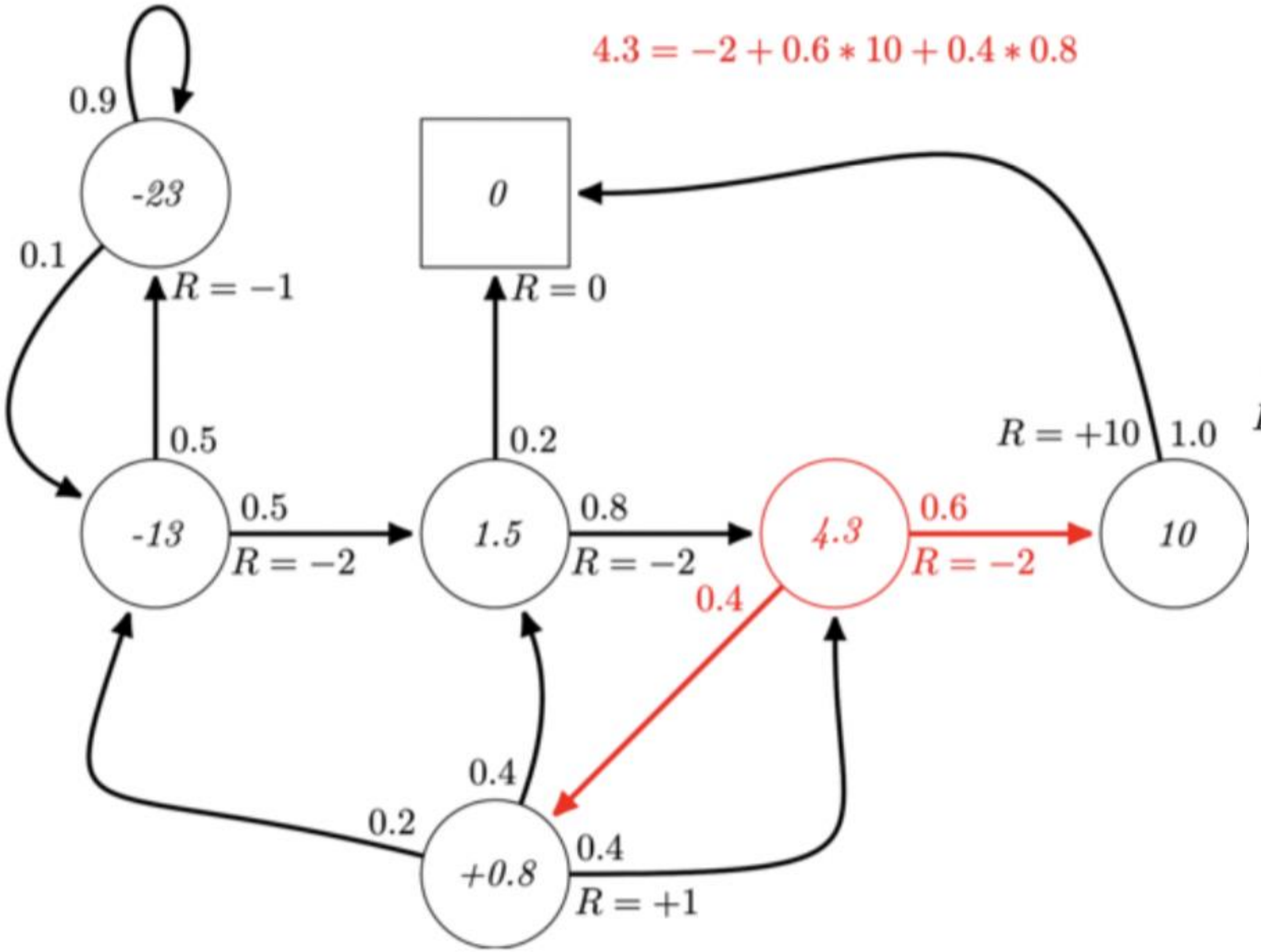


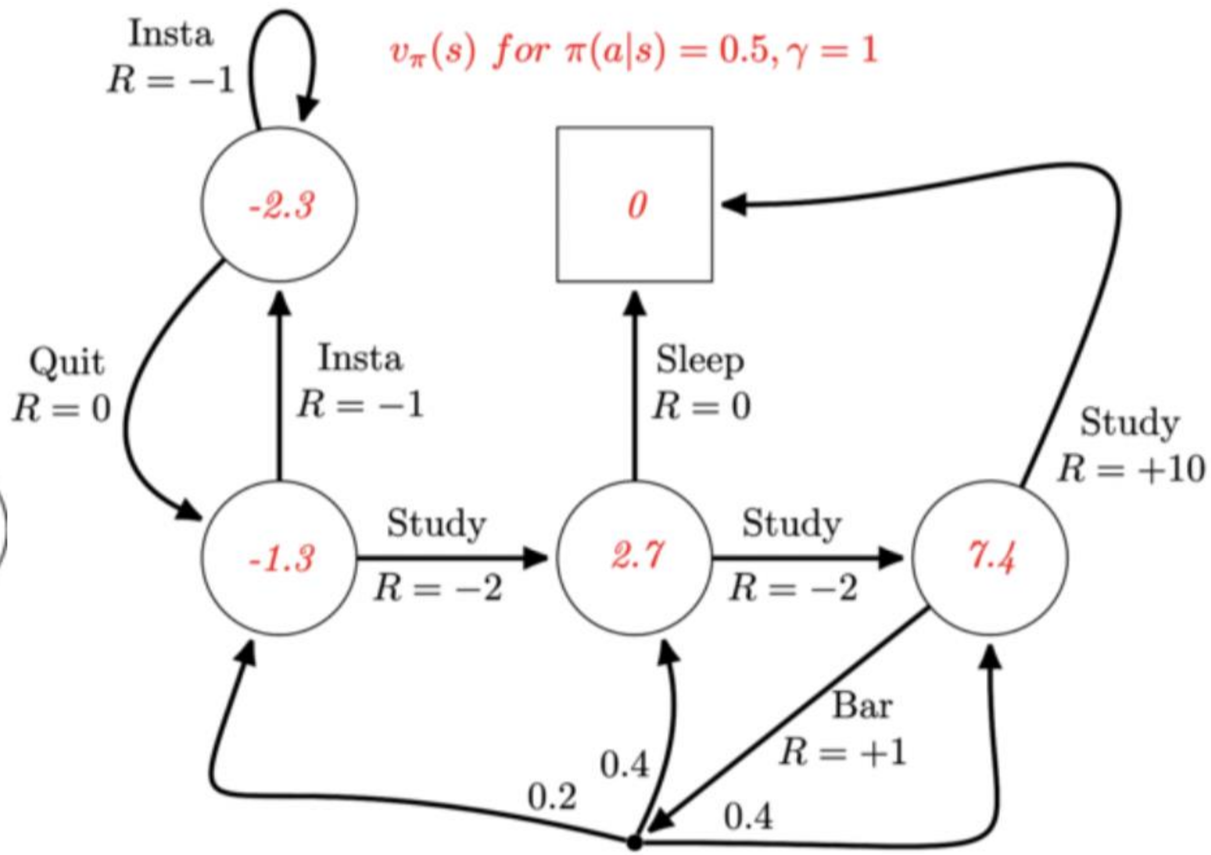
Special Case: Wall

- Do North and Successful: If x, y == x', y' == (1,0) then p = 0.8
- Do North but Slip Left: If x, y == x', y' == (2,1) then p = 0.1
- Do North but Slip Right: If x, y == x', y' == (0,1) then p = 0.1

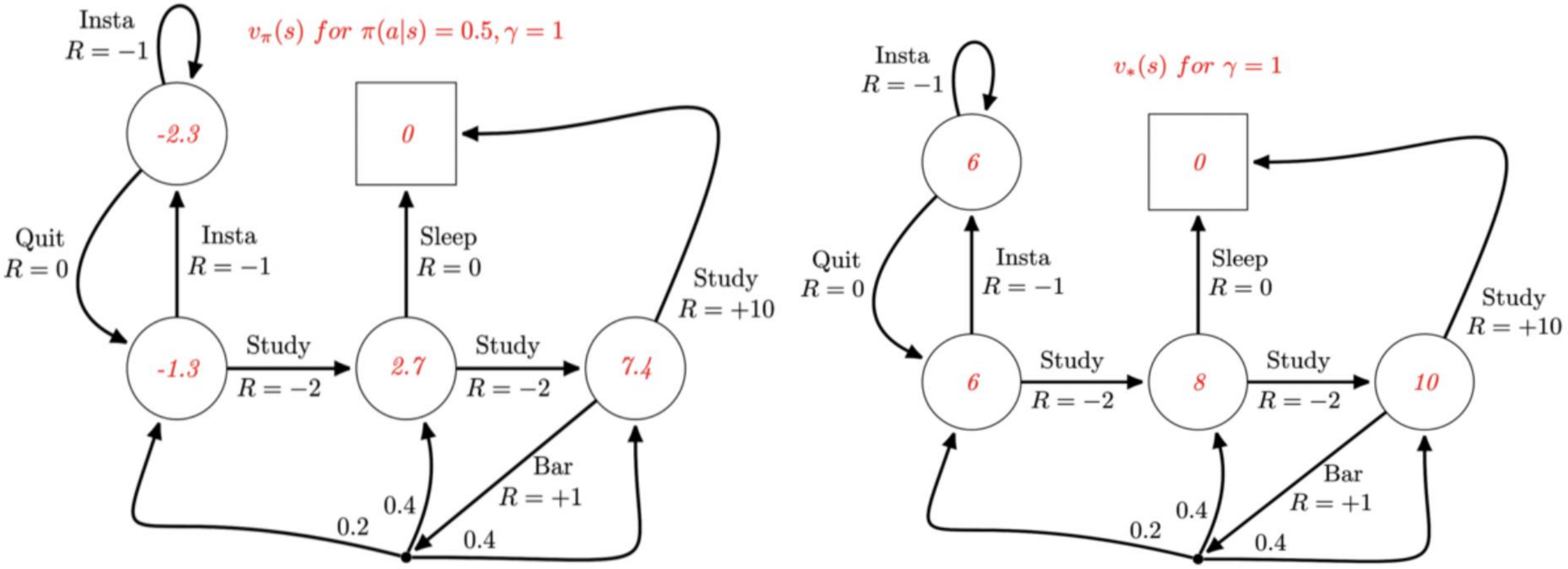
0.8 succ
0.1 slip left
0.1 slip right

Problem 2.A Compare the value functions for the Markov Reward Process and Markov Decision process shown below. Why are they different? Is there a different policy for the MDP which would result in the same values as shown in the MRP?

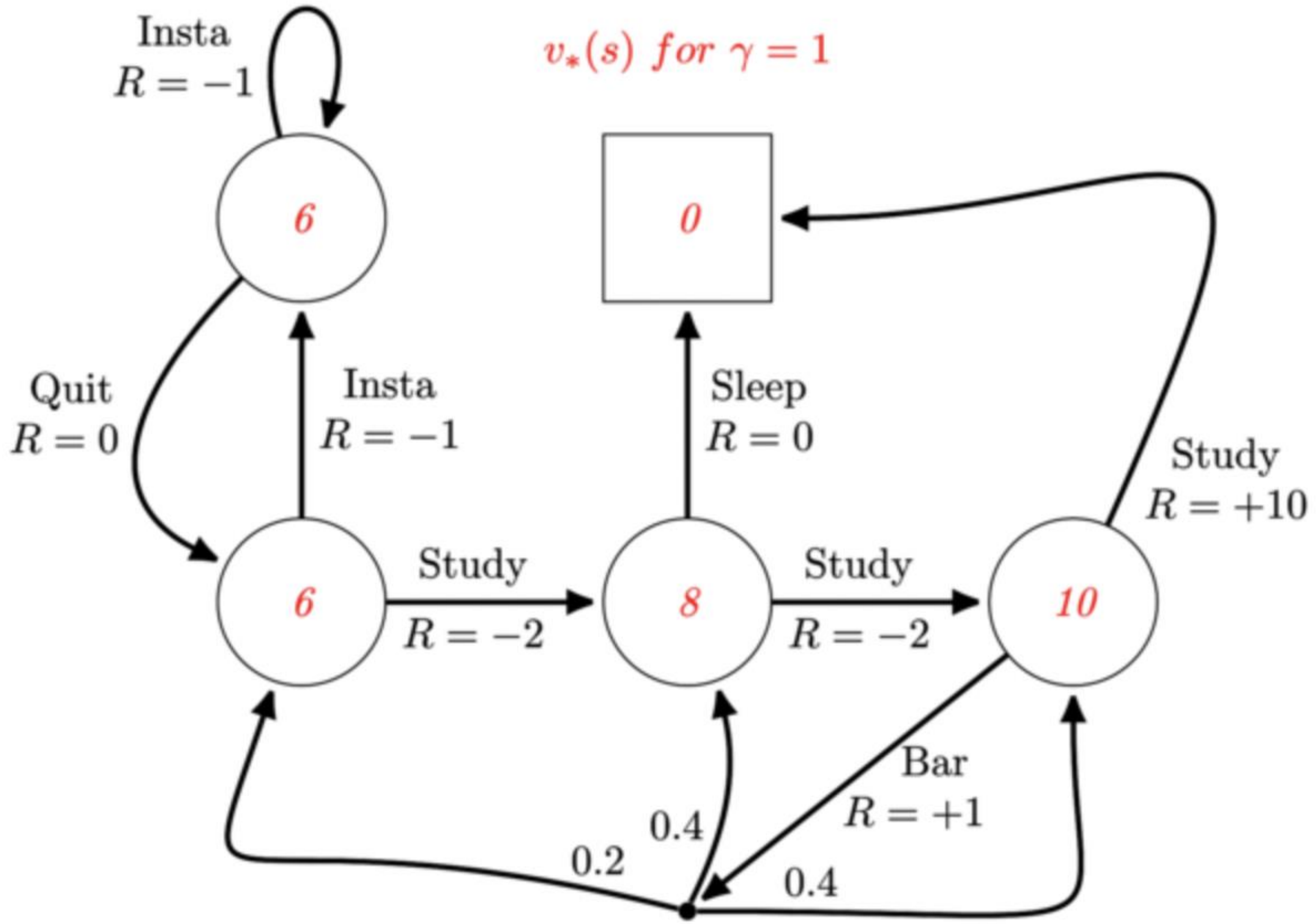




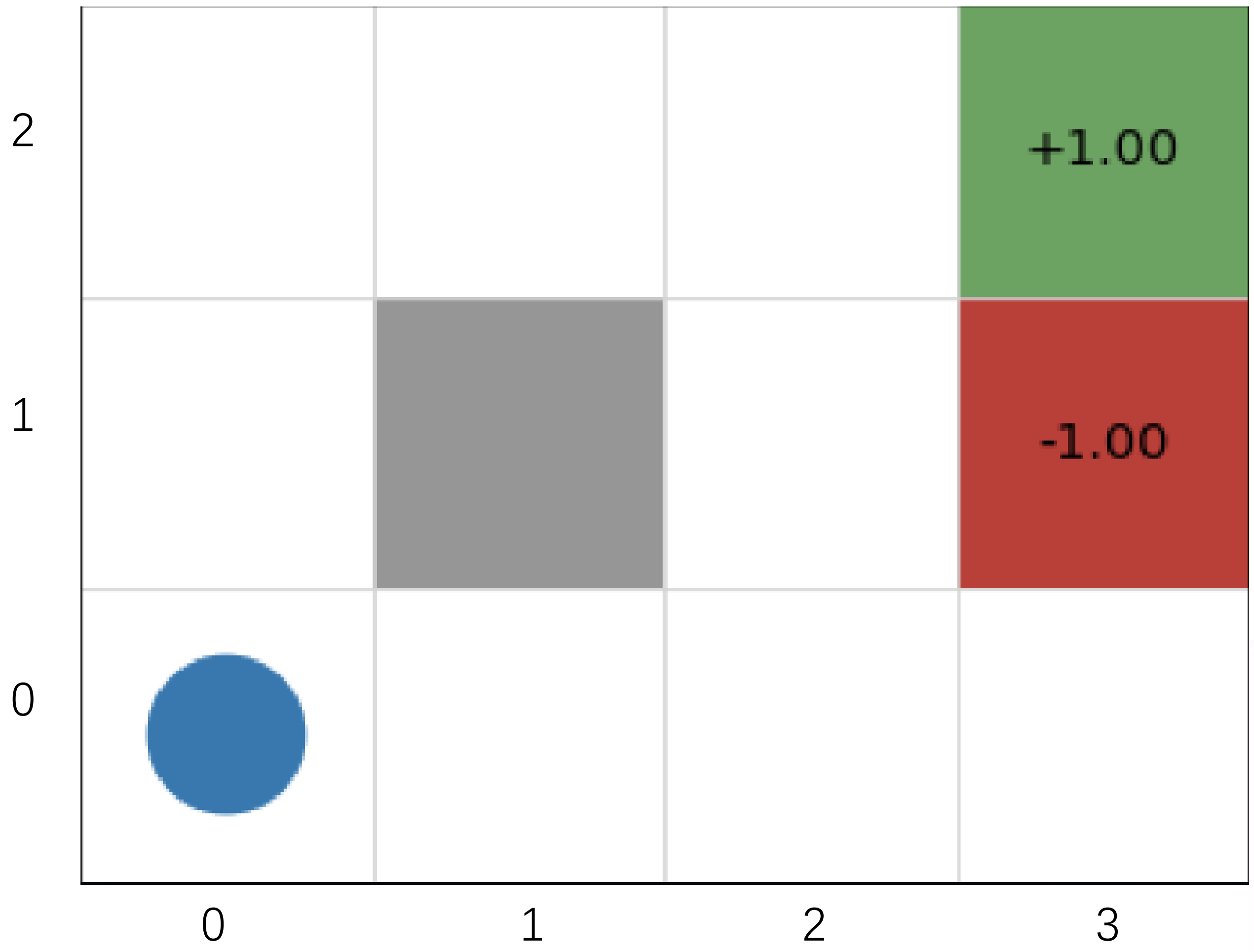
Problem 2.B Now compare the values to the optimal value function. Why are they different?



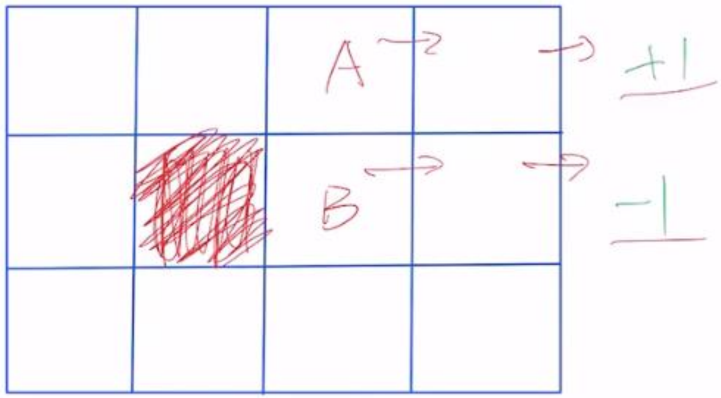
Problem 3 Given the optimal state value function above, what is the optimal action to take in the bottom left state? What about the rightmost state? How can you tell?

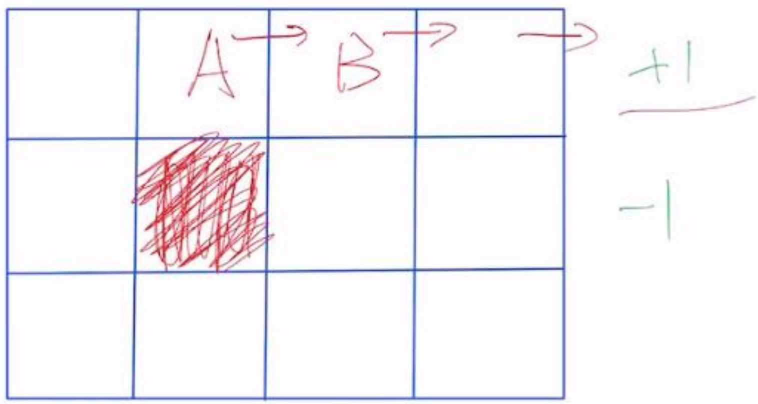


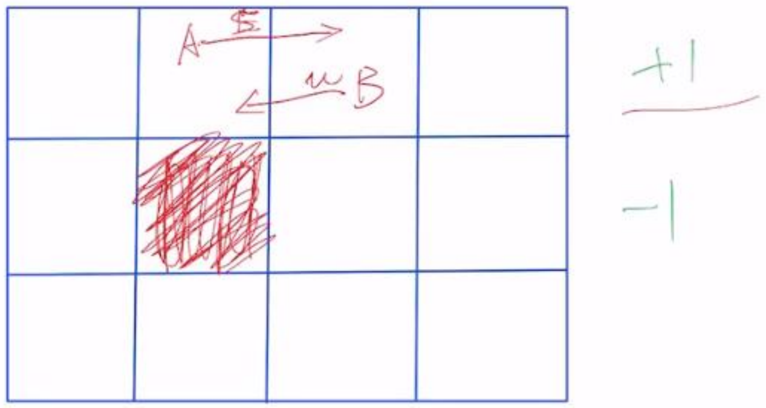
Lecture Example



0.8 succ
0.1 slip left
0.1 slip right

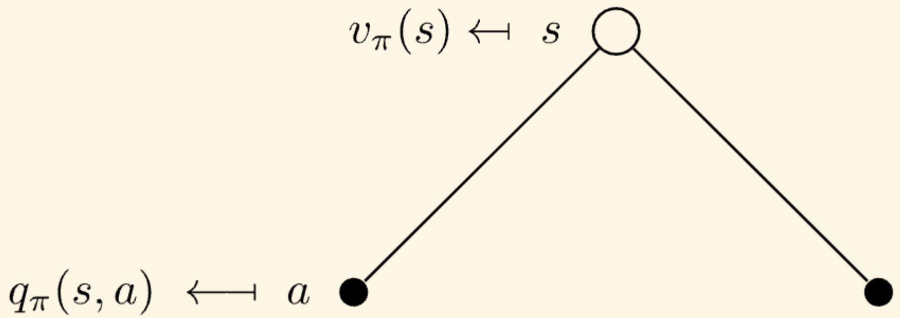

A
B
+1
-1


A
B
+1
-1


A
B
+1
-1

Formula for V(s)

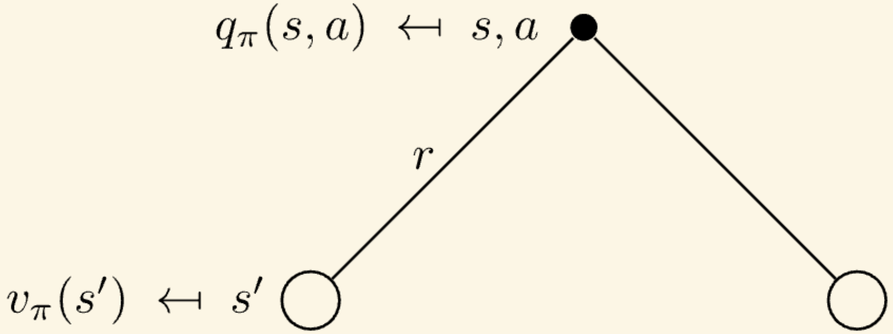
Bellman Expectation Equation for V^π (look ahead)



$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a \mid s)\, q_\pi(s,a)$$

Formula for Q

Bellman Expectation Equation for Q^π (look ahead)



$$q_\pi(s,a) = \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s')$$