Markov Decision Processes

Question 1. Let $\gamma = 0.8,\ T = 5$ and the following rewards are received: $R_1 = 4,\ R_2 = 10,\ R_3 = -8,\ R_4 = 6,\ R_5 = 5$. What are $G_0, ..., G_5$?

Answer: $G_0 = 12,\ G_1 = 10,\ G_2 = 0,\ G_3 = 10,\ G_4 = 5,\ G_5 = 0$

Question 2. You are playing a turn-based fighting game, and face an enemy. You have 2 health points, while your enemy has 1. You make one of two moves in a turn, attacking or healing:

- By attacking, you kill your enemy with a 20% chance.
- By healing, you restore one health point with a 60% chance (you can't have more than 2 health points).

After your move, your enemy attacks you, making you lose 1 health point with a 50% chance. You get a reward of 1 for each restored health point, a reward of 5 for killing your enemy, and -5 for dying. The game ends if either you or your enemy dies. Construct an MDP that represents this game, and draw its transition graph!

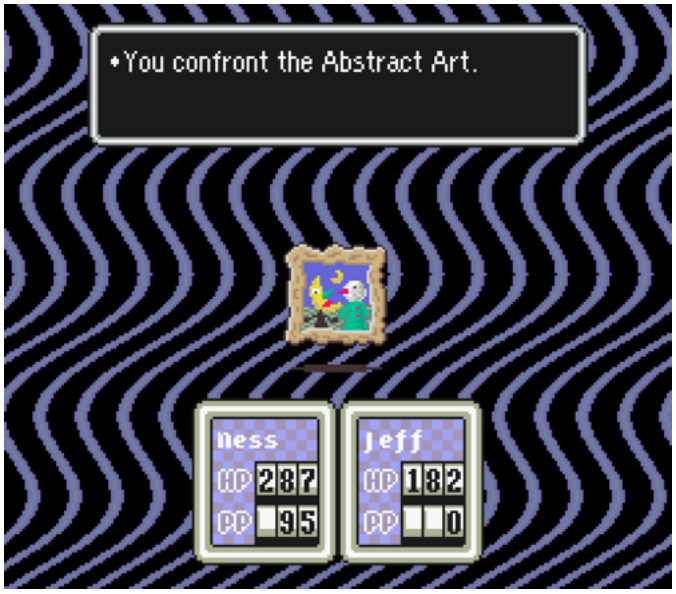
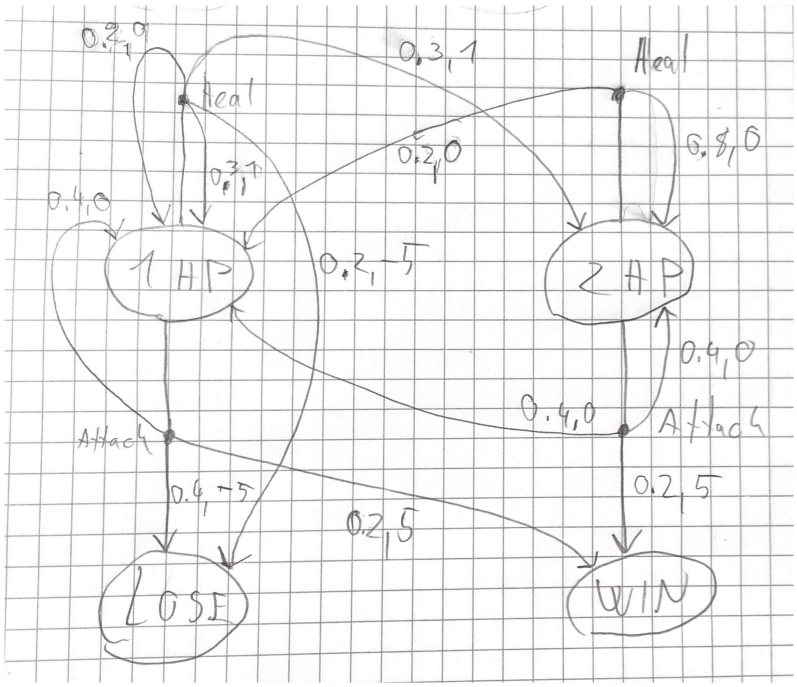


FIGURE 1. A turn-based fighting game.

Answer:



Question 3. Consider the continuing MDP shown in Figure 2. The only decision to be made is that in the top state, where two actions are available, left and right. The numbers show the rewards that are received deterministically after each action. There are exactly two deterministic policies, π_{left} and π_{right}. What policy is optimal, if
a) $\gamma = 0$,
b) $\gamma = 0.9$,
c) $\gamma = 0.5$?
d) Suppose that at a given time step t, you are in the top state. What is G_t for π_{left} and π_{right} in terms of γ?

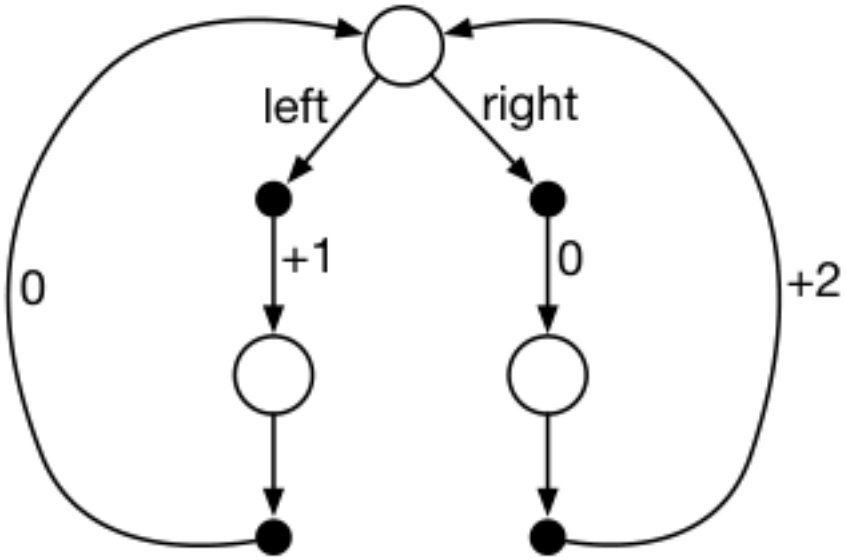


FIGURE 2. MDP

Answer: At any time step t, if we are in the top state, the discounted return will be the following:
If $\pi = \pi_{left}$:

$$G_t = 1 + 0\gamma + \gamma^2 + 0\gamma^3 + ... = \sum_{k=0}^{\infty} \gamma^{2k} = \frac{1}{1-\gamma^2}$$

If $\pi = \pi_{right}$:

$$G_t = 0 + 2\gamma + 0\gamma^2 + 2\gamma^3 + ... = \sum_{k=0}^{\infty} 2\gamma^{2k+1} = \frac{2\gamma}{1-\gamma^2}$$

At $\gamma = 0.5$ both policies are optimal. If $\gamma < 0.5$, π_{left} is optimal, and if $\gamma > 0.5$, π_{right} is optimal.