2. Exercise Sheet

Exercise 1 Evolutionary Stable State

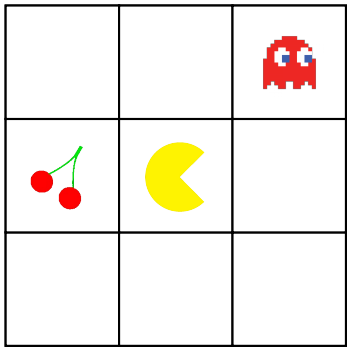
In which of the following payoff matrices is C an Evolutionary Stable Strategy. Why?

$$A = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 6,6 & 0,5 \\ 5,0 & 1,1 \end{bmatrix} \end{matrix} \quad B = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 5,5 & 2,5 \\ 5,2 & 1,1 \end{bmatrix} \end{matrix} \quad C = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 1,1 & 2,1 \\ 1,2 & 1,1 \end{bmatrix} \end{matrix} \quad D = \begin{matrix} \\ C \\ D \end{matrix} \begin{matrix} \begin{matrix} C & D \end{matrix} \\ \begin{bmatrix} 1,1 & 0,1 \\ 1,0 & 1,1 \end{bmatrix} \end{matrix}$$

Exercise 2 Agent-Environment Interface

Consider the Pac Man game in a small environment as in the picture. An agent can win the game, if all items are collected. The game consists of the elements:

1) **Agent:** Pac Man
2) **Environment:** ghosts, items, walls
3) **Actions:** left, right, up, down (neutral)



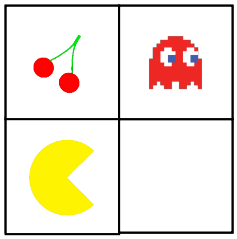
a) What is an appropriate reward for an agent, which is trying to learn how to win the game? Give an example.

b) What could be an appropriate state observation? Give an example. (Hint: a state observation is an encoding of the current state of the game, which your agent can act according to.)

c) Provide a policy, which maps an action to each state of the state observation you provided in b).

Exercise 3 Markov Decision Process

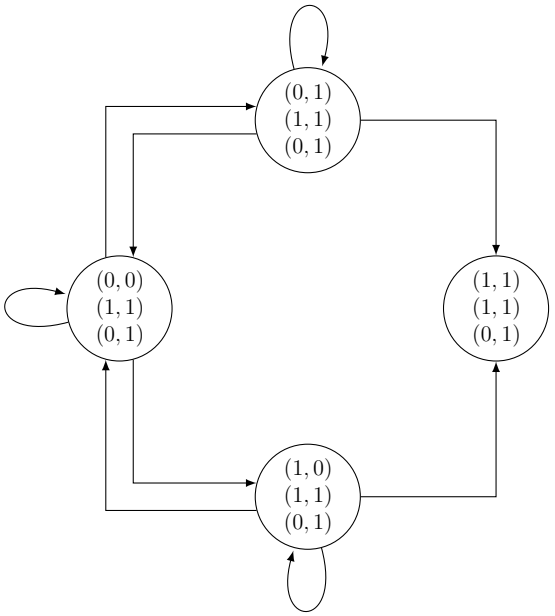
a) Complete the table with probabilities and expected rewards for the finite MDP of the Pac Man example using the state representation (Pac Man's pos., ghost's pos., item's pos.) and a map of size 2x2. Assume that the ghost is moving in each direction with the same probability. The game ends if you collide with the ghost.



s	s'	a	$p(s' \mid s, a)$	$r(s, a, s')$
(0,0); (1,1); (0,1)	(1,0); (0,1); (0,1)	Right		
(0,0); (1,0); (0,1)	(0,1); (1,1); (0,1)	Up		
(0,0); (1,0); (0,1)	(0,1); (0,0); (0,1)	Up		
(0,0); (1,0); (0,1)	(0,1); (1,0); (0,1)	Up		

b) Complete the transition graph to the right assuming that the ghost is not moving. Why are some transitions missing? (Don't forget to add the action nodes to the edges)

c) How would the graph change if the ghost would be able to move?



Exercise 4 (Discounted) Return

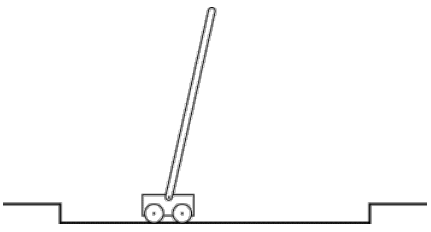
The objective here is to apply forces to a cart moving along a track to keep a pole hinged to the cart from falling over. A failure is said to occur if the pole falls past a given angle from vertical or if the cart runs off the track. The pole is reset to vertical after each failure.

a) This task could be treated as episodic, where the natural episodes are the repeated attempts to balance the pole. If the reward had been +1 for every time step on which failure did not occur, what would be the meaning of the return at each time?

b) Alternatively, we could treat pole-balancing as a continuing task, using discounting. In this case the reward would be -1 on each failure and zero at all other times. What would be the meaning of the return at each time?

c) Discuss Pros and Cons of both reward functions. Is there a difference in the expected behaviour of agents trained with different reward functions?

d) Try to come up with an even better reward function. Explain what its benefits are in comparison to both other reward functions.



Exercise 5 Iterative Policy Evaluation

Consider the following simplified Pac-Man game:

- The game continues as long as Pac-Man did not collect the single cherry.
- Pac-Man can move in the four cardinal directions: up, down, left, right
- State transitions and rewards:
  - reaching an empty cell yields a reward of 0
  - a colliding with a ghost yields a reward of -999
  - collecting the item yields a reward of +100 and ends the game
  - any other action that takes the agent off the grid, leaves the state unchanged and gives a reward of -1
- This is an undiscounted task: $\gamma = 1$
- For simplicity the ghost is not moving.

a) Complete the values for the value function for k = 1 and complete the given backup diagram for a Pac Man following the equi-probable random policy (all actions equally likely), for all s $\pi(a|s) = 1/4$

b) What would the optimal policy look like? (Either repeat the value function calculation two more times or use a computer program to calculate the necessary state values.)