# Exercise 1
# Introduction Markov Decision Processes

## 1 Markov Decision Processes

Remember that a Markov Decision Process (MDP) is a tuple $< S, A, P, R, \gamma >$ where:

- $S$ is the set of *environment states*
- $A$ is the set of possible *actions*
- $P$ are the *state transition probabilities*
- $R$ is the *reward function* and
- $\gamma$ is the *discount factor*.

**Task:** Formalize the following problems into an MDP:

1. Road Crossing: Our agent is standing on one side of a large road and has to cross it. Each lane of the road can have a car on it, or not. Each timestep, the lane occupancy changes randomly. The agent can either cross a single lane in each time step, or halt between lanes on a traffic island. Give $< S, A, P, R, \gamma >$ for this problem with a single lane.
2. Chess against a random opponent. Explain what $< S, A, P, R, \gamma >$ need to be conceptually and give approximate cardinalities (i.e., how many elements there are) for $S$ and $A$.

## 2 Maze Runner

Imagine you design a robot whose task is to find its way through a maze. You decide to give it a reward of +1 for escaping the maze and a reward of zero at all other times. Additionally, you decide that a discount factor of 1.0 should suffice. The task seems to break down into episodes - the successive runs through the maze - so you decide to treat it as an episodic task, where the goal is to maximize *expected total reward*. After running the learning agent for a while, you find that the agent is taking a very long time to solve the maze, i.e., he is dawdling.

**Question:** What is going wrong? How can we modify the above MDP to alleviate this problem?

## 3 Markov Property

Imagine your goal is to teach an agent to act inside a very dynamic, real-world environment. It is important that the agent learns to react to moving so to, for example, avoid colliding with them. Because money is sparse, as it always is, the agent senses the environment exclusively through a consumer type camera.

**Question:** Does modelling the environment state $s_t$ as the camera image sensed at timestep $t$ fulfill the *Markov property*? If not, why? How could you make it so? (*Remember:* "The future is independent of the past given the present.")