MAIR mock exam 2016 part “Markov models for multi-agent learning”

Name: .......................................................... Number: .................

Calculator allowed. Devices with an internet connection are not allowed. Closed book. Answers must be written in boxes. Answers must be justified by a calculation. Scratch paper is provided on the back.

1. A robot is trapped in a maze. Initially it has to choose one of two directions. If it goes to the right, then it will wander around in the maze for three minutes and will then return to its initial position. If it goes to the left, then with probability 1/3 it will depart the maze after two minutes of manoeuvering, and with probability 2/3 it will return to its initial position after five minutes of travelling. Assuming that the robot is at all times equally likely to go to the left or the right, what is the expected number of minutes that it will be trapped in the maze?

2. Consider the following Markov reward process. Suppose a discount factor of $\gamma = 0.5$.  

   \[
   \begin{aligned}
   s &: s(r = 6, p = 0.6); a(r = -4, p = 0.4), \\
   a &: a(r = 2, p = 0.2); c(r = -3, p = 0.8), \\
   b &: s(r = 3, p = 0.5); c(r = -1, p = 0.5), \\
   c &: a(r = -1, p = 0.6); b(r = 1, p = 0.2); c(r = -1, p = 0.2).
   \end{aligned}
   \]  

   Node $s$ is the start node.

   (a) Give a state transition diagram, including probabilities and rewards.

   (b) Give the probability transition matrix, $P$, and the immediate reward matrix, $R$.  

(c) Express the vector of optimal values \( v^* = (s,a,\ldots,c) \) as a solution of a system of linear equations. It is not necessary to simplify this system of linear equations.

On decision nodes (provided there are any) the probabilities must represent the policy that is to be evaluated.

(d) Perform two steps of value evaluation. Use a table to show the results.
Scratch paper.
Scratch paper.