Q1 and Q2

The code file is titled “Ex0_Q1 and Q2_Environment and Manual Policy”

Q3. In the Manual-Policy because we already know where the goal point is, the point which causes the most immediate reward, we could have the optimal policy in each state. The optimal policy is following the shortest pass between the start point and the goal point. In this case, we could have the optimal value (cumulative rewards). However, in the random policy regardless of the goal point we will just select actions randomly which will cause much less cumulative rewards and value.

Knowing the goal point location, in the manual policy, we could also select actions which cause the robot never to reach the goal point location and as a result it will have zero cumulative rewards or less cumulative rewards compared with random policy.

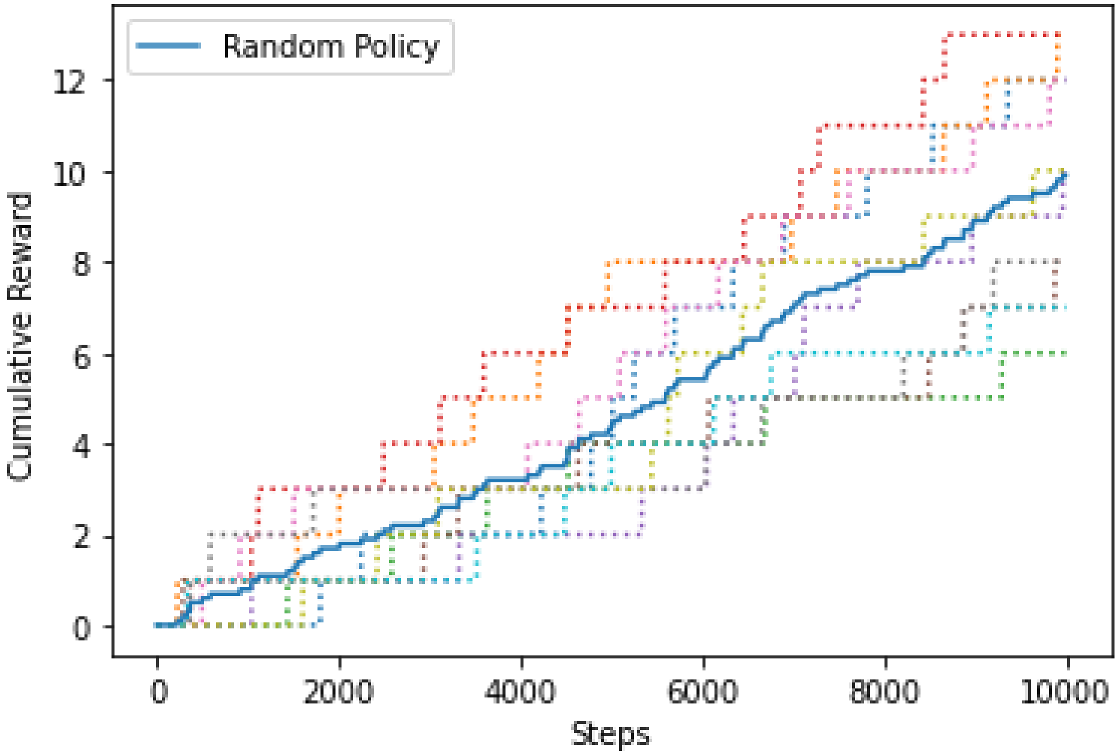

Figure 1: The cumulative rewards of Random policy

In figure 1 we could see the cumulative rewards of Random policy.

Q4.

In worse policy which is the worse policy because it will never gain reward, we only select LEFT action. In this case, because of stochastic action the robot sometimes moves up and down, but it always is transferring between (0,0), (0,1), (0,2), (0,3) and (0,4) States.

As a Better Policy, we try to follow the path to the goal point but not perfectly, so it is not optimal policy. For that, initially, we try to go up till reaching states with y=8. Then moving right till having states with x=10. At the end we will go up again to reach the goal point. It is worth saying because of stochastic action we could not guarantee that the robot would reach the goal point in each trial. For example, for initial phase the robot will continue going up till hitting the wall, most of the time robot will continue hitting the wall but because of stochastic policy there would be a step in which the robot will move right and keep going up till reach states with y=8 and going right till reaching the goal point.

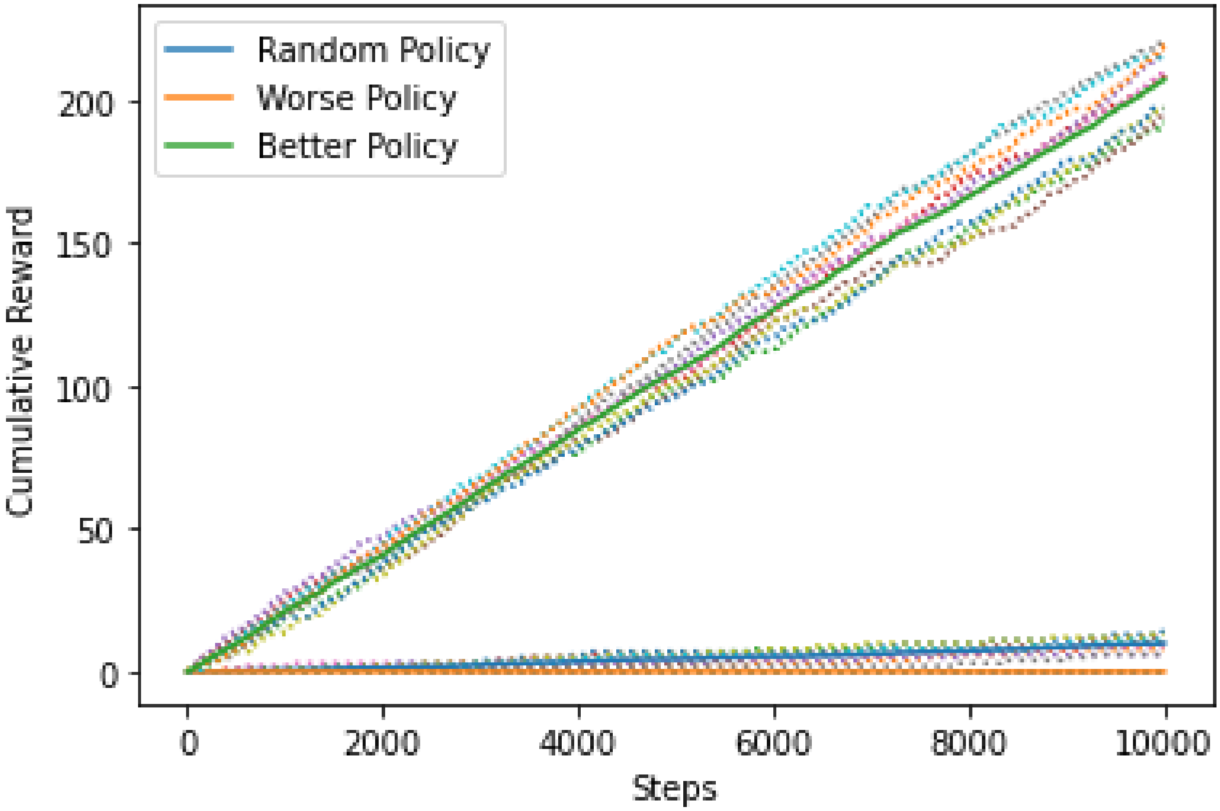

Figure 2 shows the cumulative rewards of Random policy, Worse Policy and Better Policy.