

# HW0

Theo McArn

January 2026

## 1 Random Policy

**How do you think your random policy compares with your manual policy? What are some reasons for the difference in performance?**

This random policy performs much worse than I would expect the manual policy to perform. This is because, with the manual policy, the operator is granted privileged with external information about the agent's environment and can plan ahead to make intentional and efficient steps to the goal state. With the random policy however, the agent has no additional information about its environment, and only takes random actions. This difference in performance can be shown as well. From the plot in figure 1, it seems that it takes the randomly acting agent about 1000 steps to reach the goal state once. However, with the manual policy, I am able to reach the goal state in about 25 steps. Obviously much more efficient.

## 2 Better Policy

For the Better Policy, I still had the agent act randomly, however I shifted the probability distribution so that the agent is more likely to make "productive moves" moving upwards and to the right, towards the goal state. I changed the distribution so that there is a 90% chance of a productive move and only a 10% chance of an "unproductive move" (downwards or to the left). As you can see in figure 2, this policy results in an average final cumulative reward of about 22 compared to the cumulative reward of about 8 from the random policy.

## 3 Worse Policy

For the Worse Policy, I used a similar strategy as for the Better Policy, however the probability favors the "unproductive moves" more heavily, a 99% chance, compared to the "productive moves", a 1% probability. As you can see from figure 3, this policy results in an average final cumulative reward of about 5 compared to the cumulative reward of about 8 from the random policy.

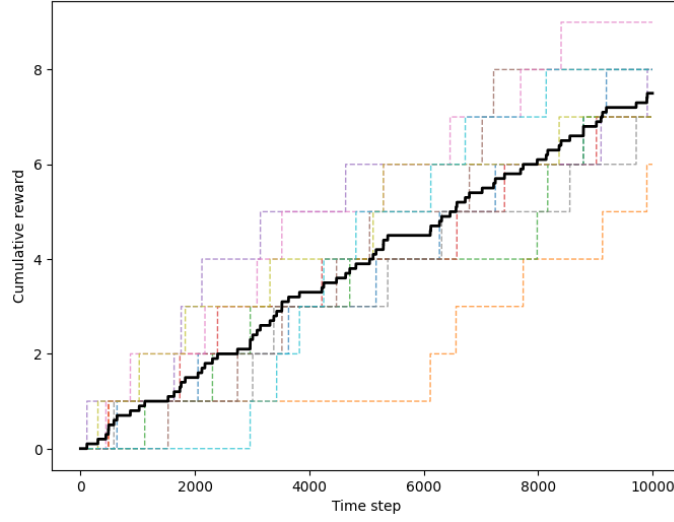


Figure 1: Cumulative Reward for RANDOM Policy

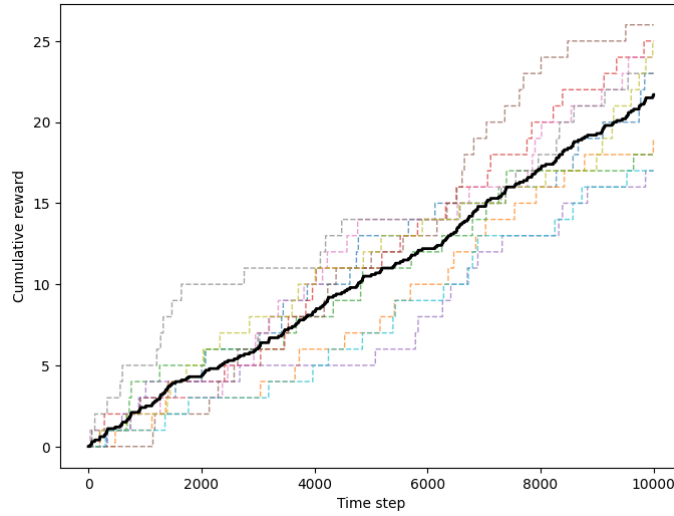


Figure 2: Cumulative Reward for BETTER Policy

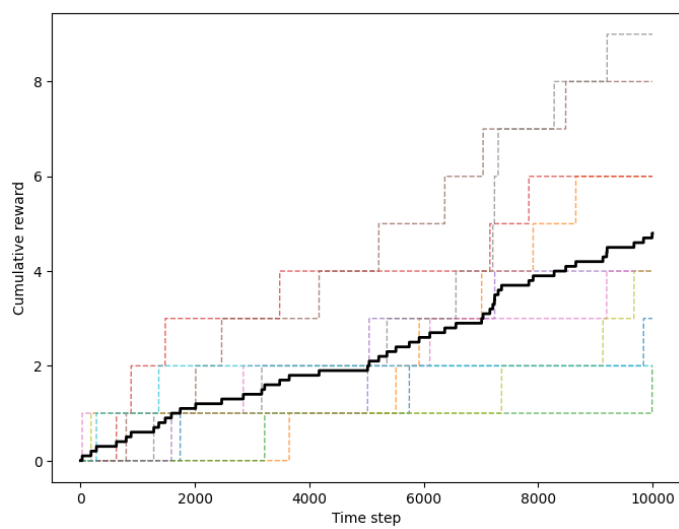


Figure 3: Cumulative Reward for WORSE Policy