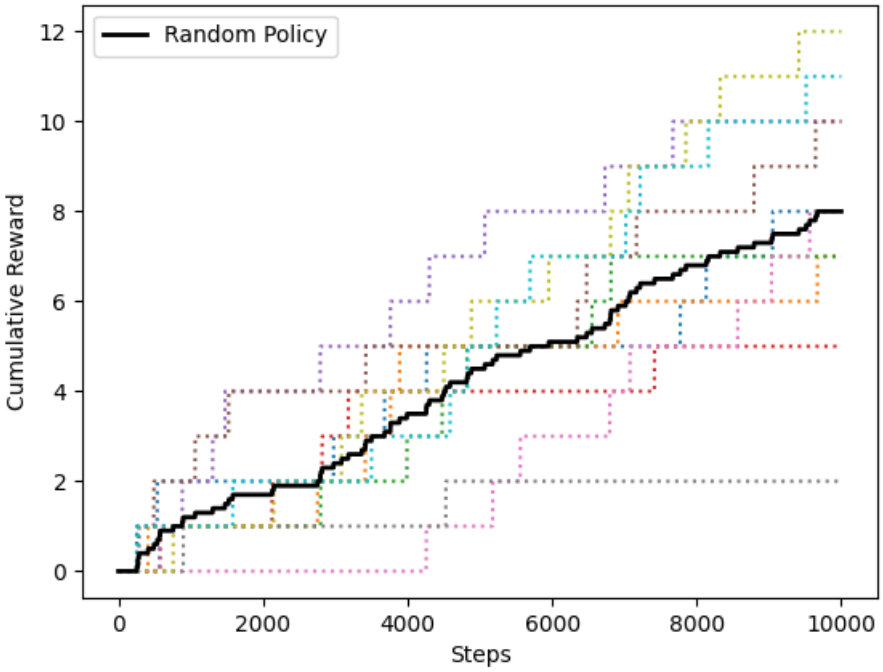
Q3:

Plot:

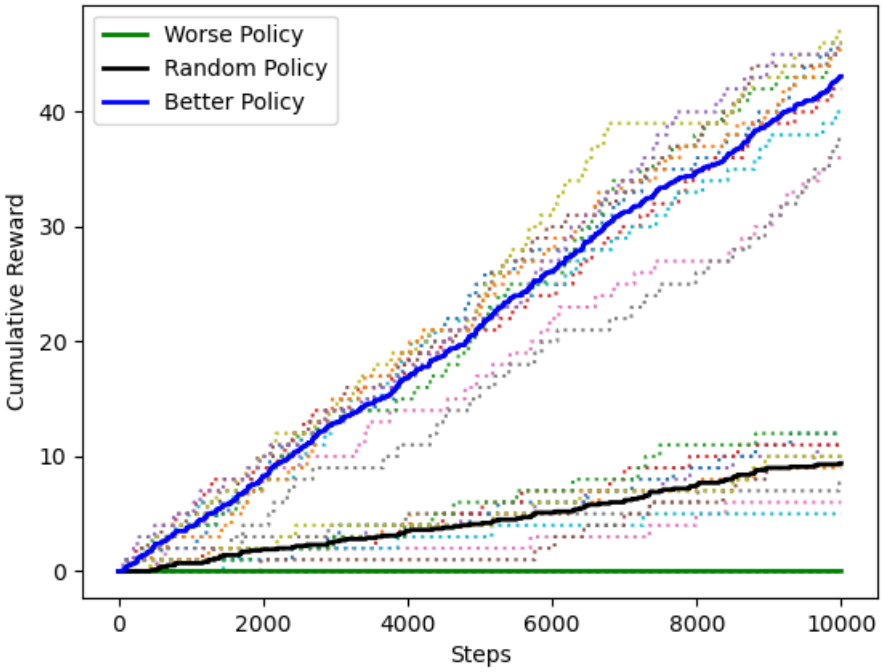


Written: How do you think this compares with your manual policy? What are some reasons for the difference in performance?

I think that the random policy is not as effective as the manual policy. In the case of the random policy, actions are completely arbitrary, making it challenging for the agent to establish a clear objective. Consequently, the agent struggles to discern between good and bad actions. Conversely, with the manual policy, the agent possesses a clear understanding of the target location, allowing it to make subjective selections of favorable actions that align with reaching the specified target location.

Q4:

Plot:



Written: Describe the strategy each policy uses, and why that leads to generally worse/better performance.

Manual Policy: This strategy involves selecting actions based on user inputs, with the user supervising the action selection process.

Worse Policy: This strategy entails exclusively selection of the LEFT action.

Random Policy: The strategy is to let the agent autonomously select actions randomly without any supervision.

Better Policy: This strategy involves commanding specific actions for the agent to execute in response to particular states, according to the target location. In such instances, the agent is under control and has a sole action for movement.

The superior performance of the **better policy** can be attributed to the reduction in the randomness of actions. In these states, the agent's actions are predetermined and represent the optimal choices. This results in a decrease in action variety but an increase in the certainty of selecting the best actions. The **random policy** is the same as the random search, so it exhibits a high degree of action uncertainty. Hence, the performance is not very good. When compared to the aforementioned two policies, the **worse policy** exclusively contains one LEFT action. Since the target location is at the upper right corner, so there is no way for this policy to approaching to the target.