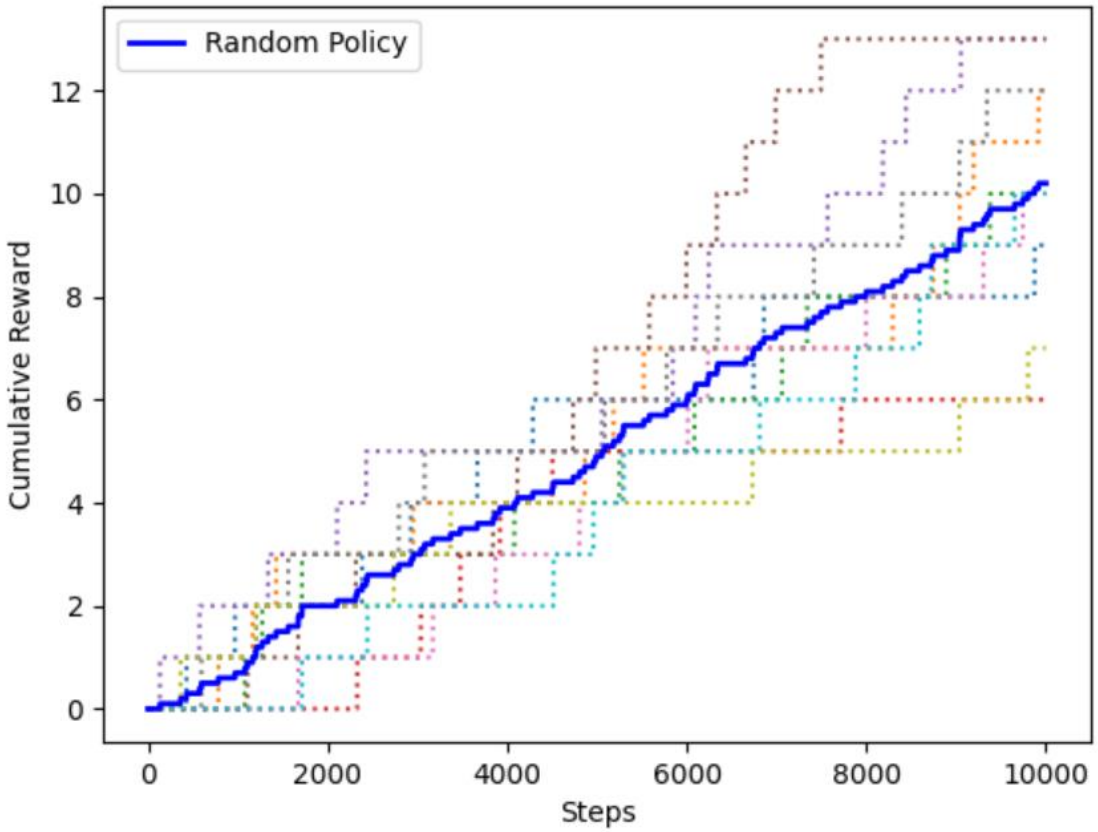
Q3:

Plot:



Written: How do you think this compares with your manual policy? (You do not have to run your manual policy for 104 steps!) What are some reasons for the difference in performance?

1. *I think random policy is not as accurate as manual policy.*
2. *When I control the agent manually, I am controlling the agent purposefully towards the target location and I can confirm the location of the agent, while random policy is just aimless exploration.*

Q4：

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

1. **Manual Policy:**
   Operate agent movement by getting commands from user input

2. **Random Policy:**
   Operate agent movement by random commands

3. **Better Policy:**
   Make the agent move in a specific direction in some specific position.
   For example,
   when the agent reaches the route leading to the four gap parts of the wall in the map [example: (1, i), i=0, 1, 2,...8], make the agent tends to move in the direction of the gap;
   when agent reaches the square area formed by the middle wall of the map [the square area formed by (2, 2) (4, 2) (4, 4) (2, 4)], make agent tends to move to the left, below;
   when agent reaches the rectangular area formed by the top right of the map [the rectangular area formed by (6, 5) (10, 5) (10, 10) (6, 10)], make agent tends to move to the right, above.

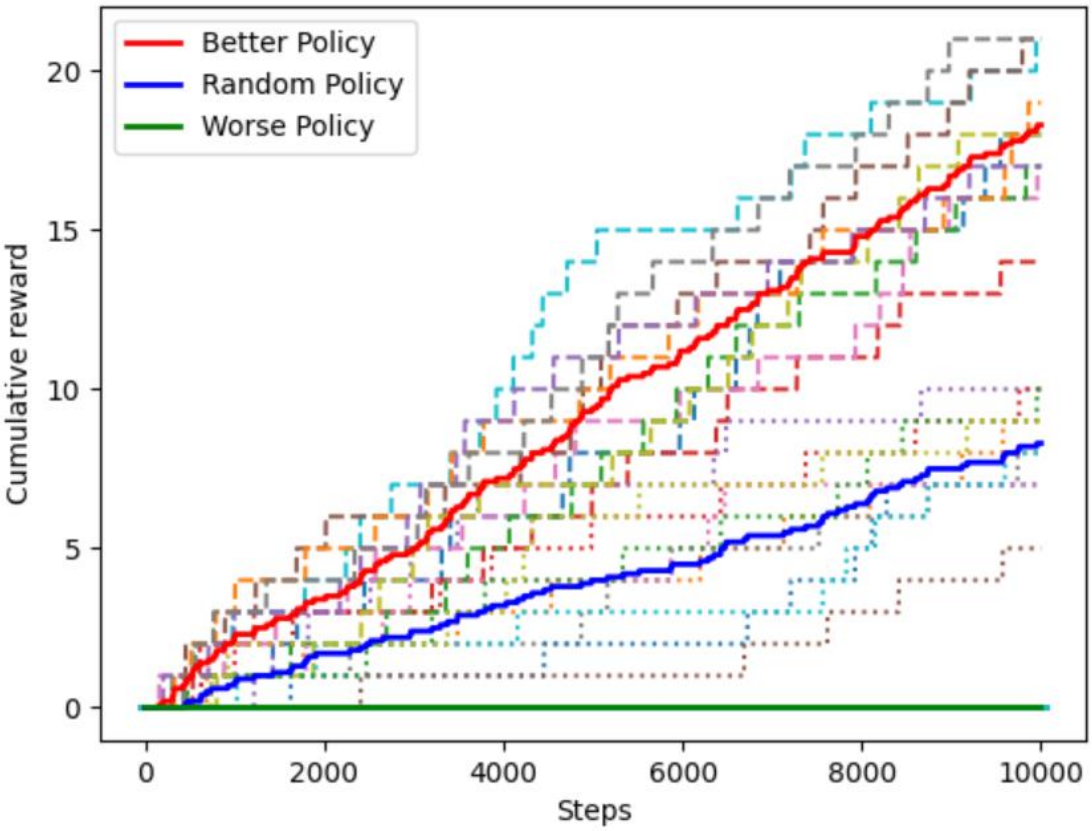
   This increases the likelihood that the agent will move towards the target location, making it easier to reach the target location.

4. **Worse Policy:**

   On top of the random policy, remove the agent move up command.

   Thus, the agent never reaches the target position in the top right corner of the map.

Plot:


Better Policy
Random Policy
Worse Policy
Cumulative reward
Steps
0
5
10
15
20
0
2000
4000
6000
8000
10000