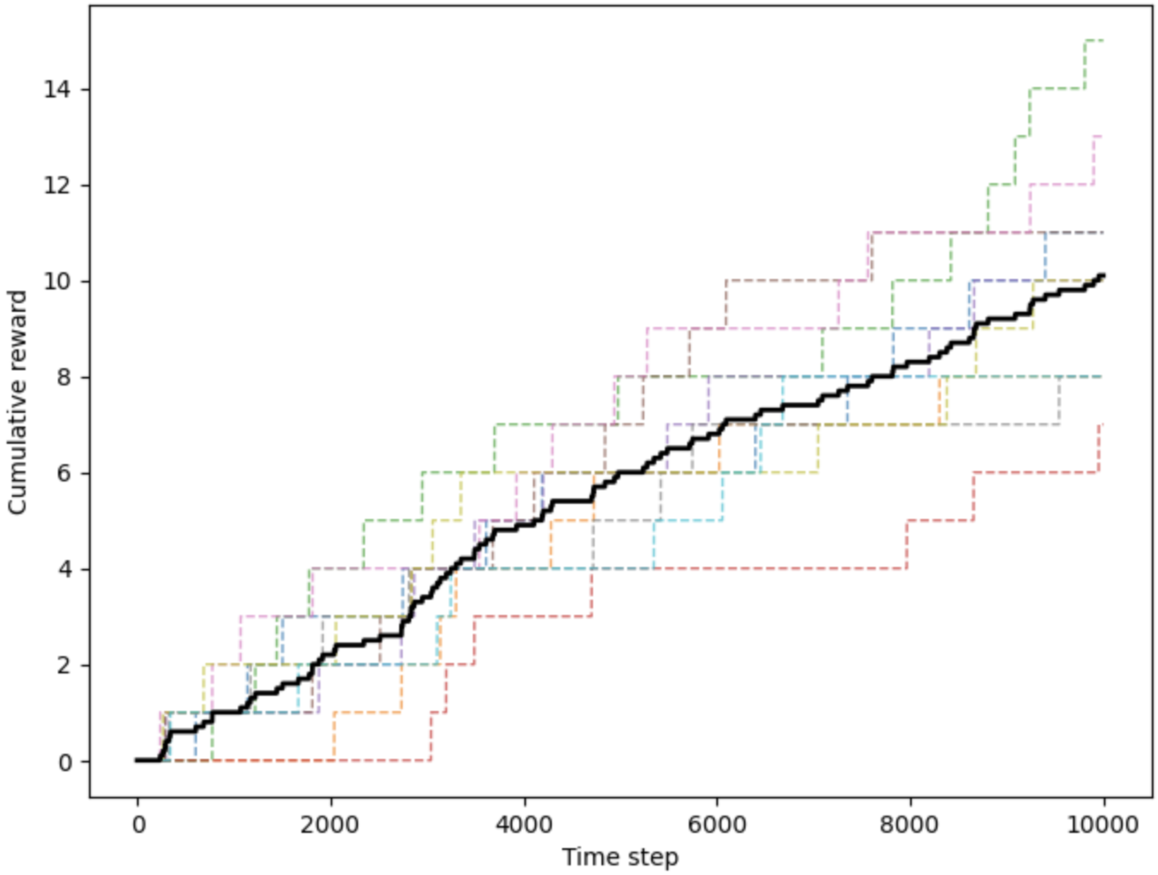
Luke Davidson

CS 5180

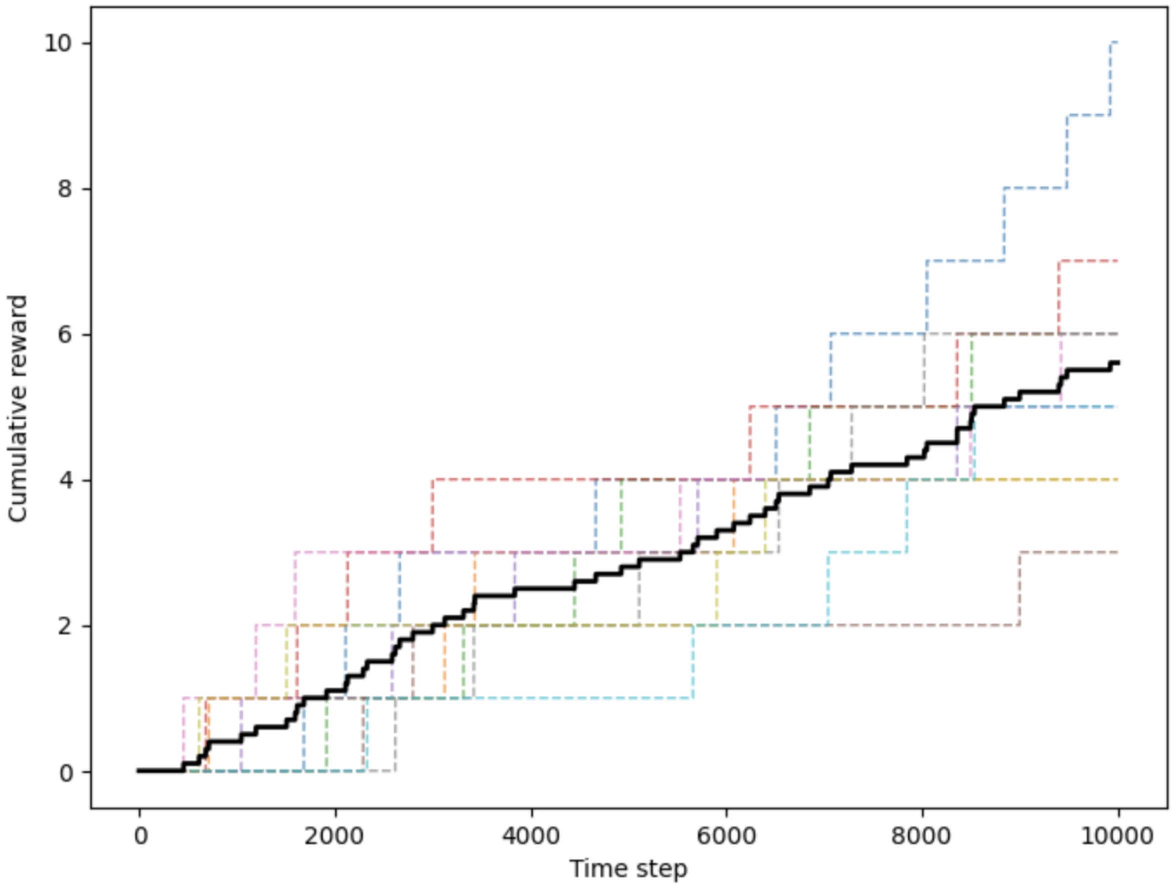
Ex0

3.) Plot:



Written: This policy will be worse than the manual policy because I have much more intuition on how to obtain the reward than a random generator does. I know that the reward is in the top right, and I generally know where the “doors” are and how the environment lies around them. Using this knowledge, I would be able to navigate up and to the right more efficiently and often than a completely random policy.

4. Worse plot:

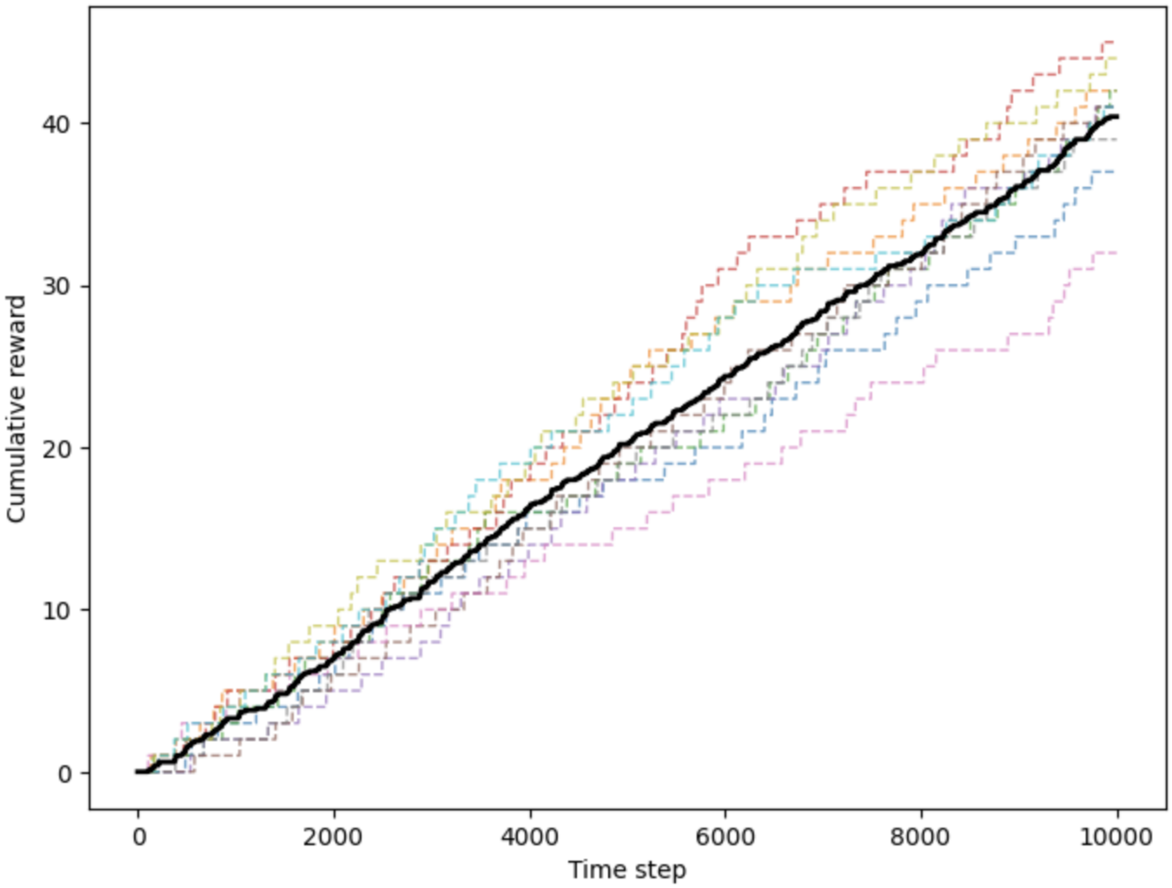


Explanation: The policy used in this test is one that ensures that the same action is not selected twice in a row. For example, the actions may come out as

UP, LEFT, RIGHT, DOWN, UP, DOWN, RIGHT, UP, LEFT, RIGHT,...

This policy is expected to be worse because of the way the four rooms are laid out. The reward is in the top right, which ideally would be achieved by a series of UP's and RIGHT's. Similarly, since there is a lot of open space, a good strategy would be to continue going in the same direction until you hit a wall to figure out the surroundings. Since this policy (basically) eliminates those two strategies, it will find the reward much less often.

Better plot:



Explanation: The policy used in this test is one that favors the selection of UP and RIGHT more so than DOWN and LEFT. 70% of the time, the policy will choose an action of UP or RIGHT (or 35% chance each), and 30% of the time the policy will choose an action of DOWN or LEFT (15% chance each). Similar to the explanation in the worse case, the reward is in the top right, so a collection of UP's and RIGHT's will likely lead to better success in each room, and ultimately the reward. I had to do some tuning of the 70/30 hyperparameter since I didn't want it to be too heavily favored towards UP/RIGHT. If it was too heavily favored towards UP/RIGHT, the agent would get stuck in corners of rooms for longer periods of time than desired. If too small, it would be very similar to the random policy. As you can see in the above plot, a 70% chance of selecting UP or RIGHT led to much higher rates of success.