

Tyler Edmiston  
Eddie Azinge

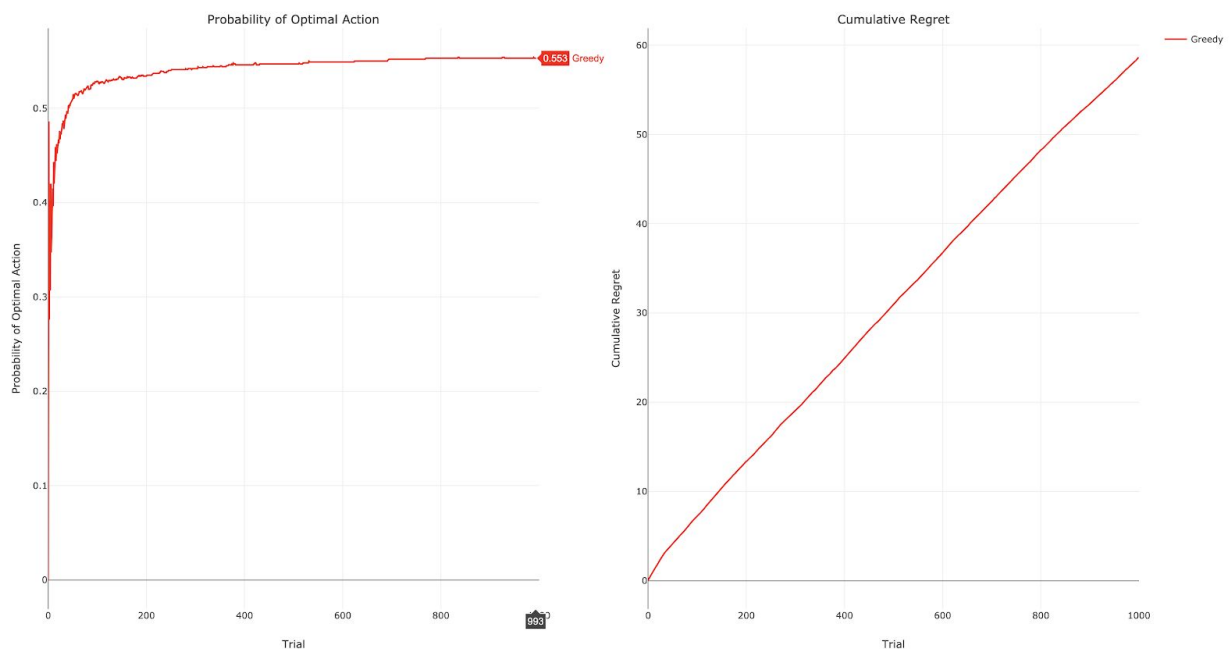
## Homework 4 Report

### Problem 1: Greedy Agent

Expected:

The Greedy agent is expected to find a good solution that returns correct over 50% of the time, but not always the optimal. Cumulative regret is expected to be stable at a linear rate.

### Sim Results



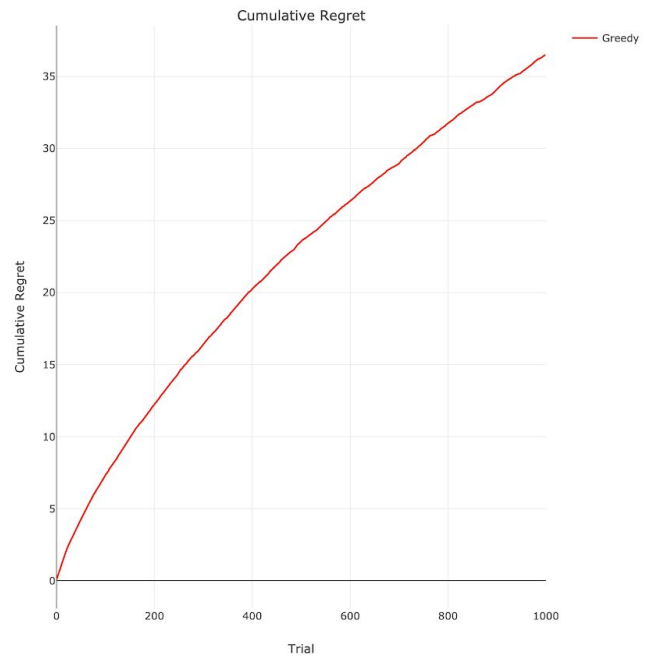
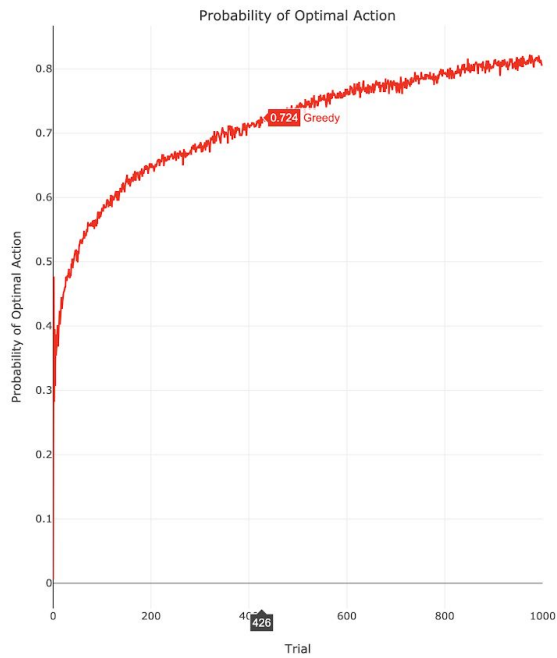
## Problem 2

### Expected

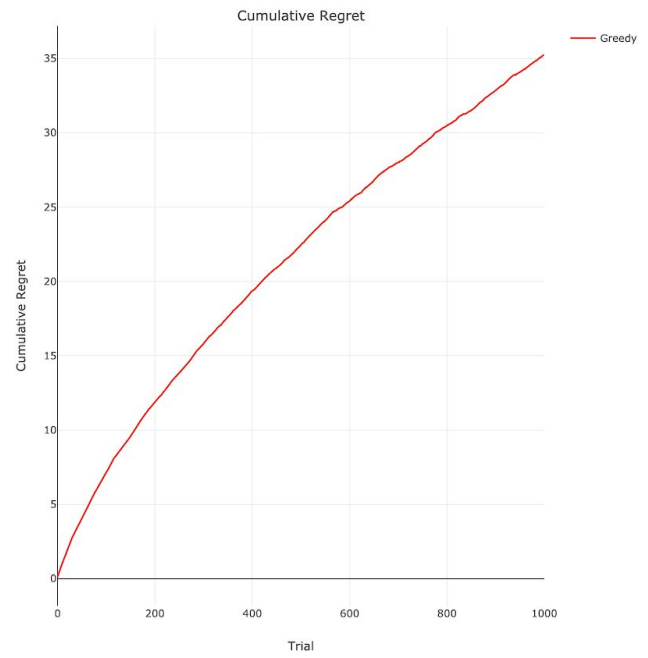
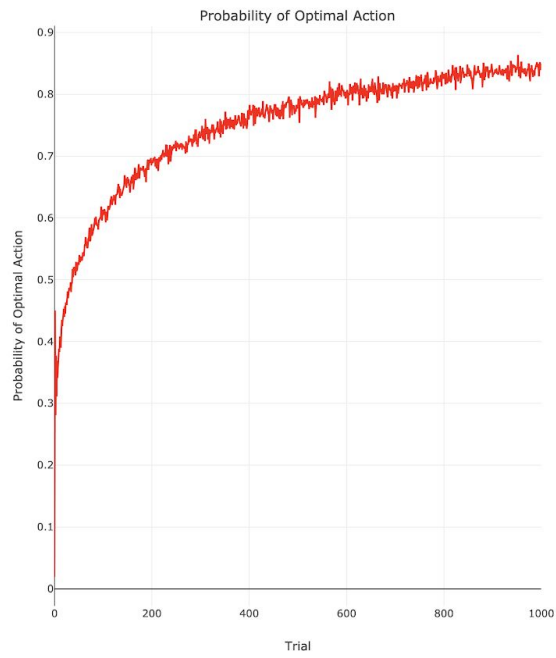
Epsilon Greedy is expected to converge the probability of finding the optimal answer as  $1 - \epsilon$ . Cumulative regret should be logarithmic as time goes on.

### Sim Results

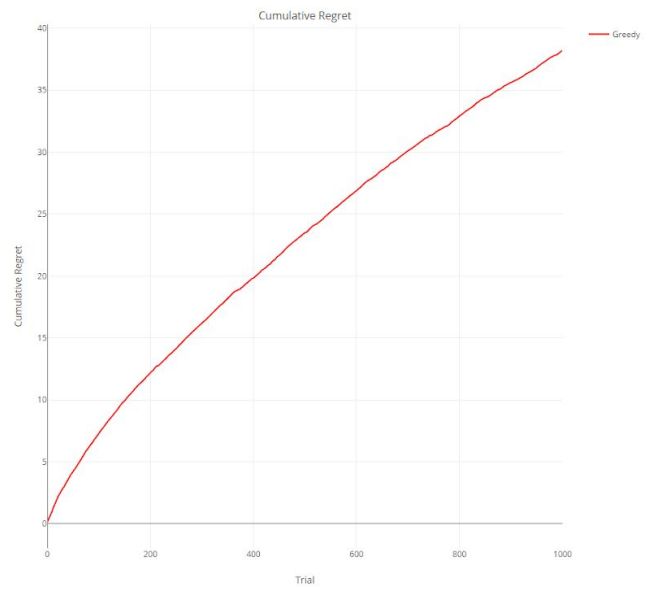
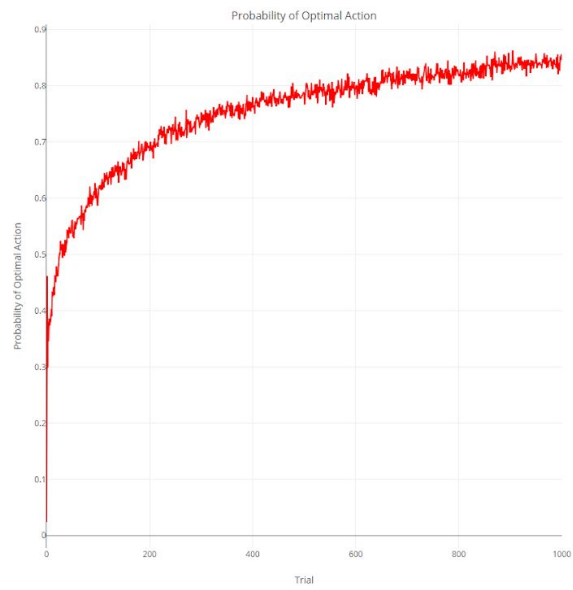
$\epsilon = 0.05$



E=0.10:



E = 0.15



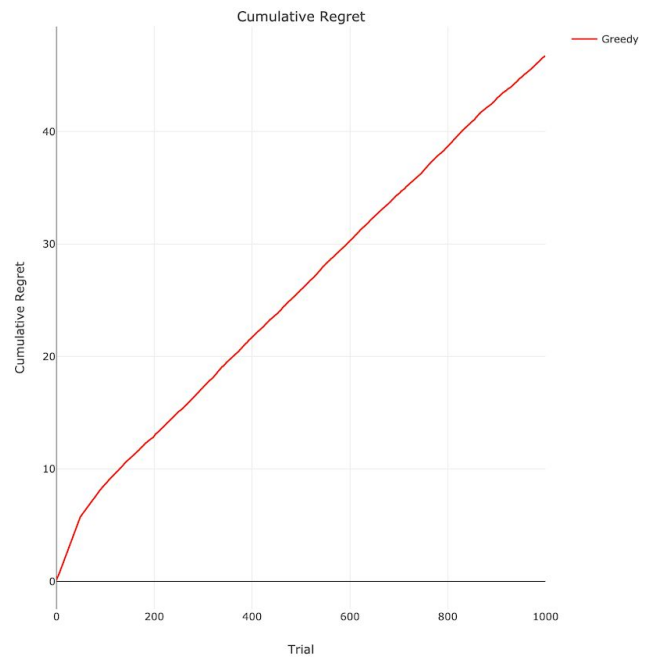
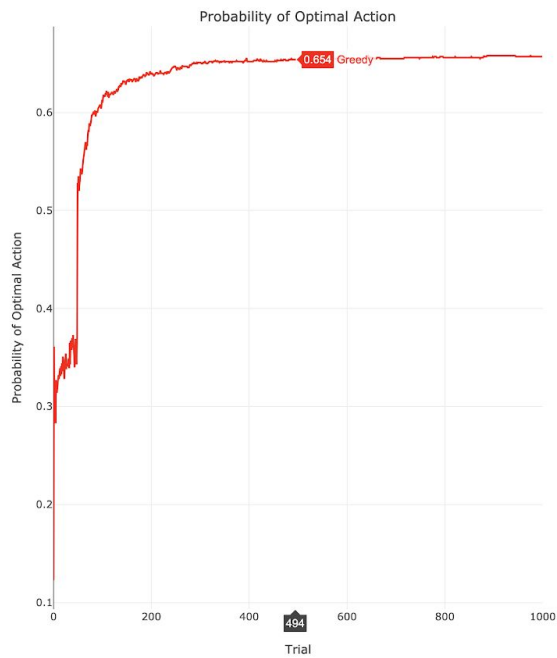
### Problem 3

#### Expected

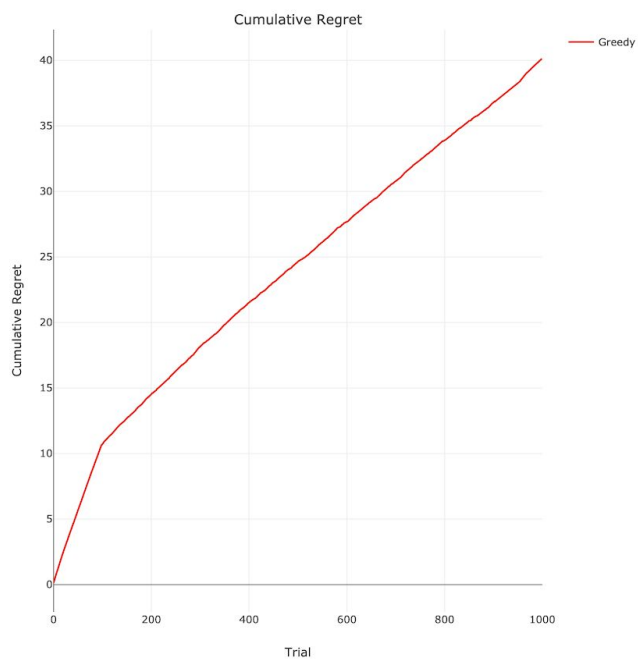
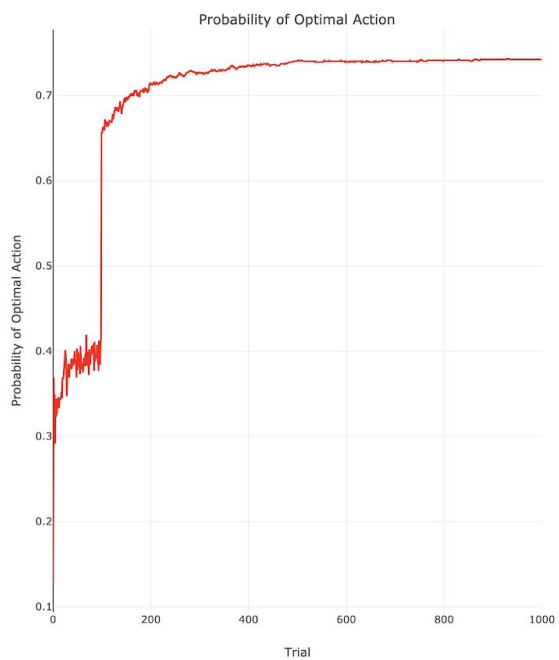
Knowing we have a finite time horizon  $T$ , we determined our values of epsilon based off of the results we got from the epsilon greedy approach. The values of epsilon should be fairly similar for epsilon first and for epsilon greedy.

#### Sim Results

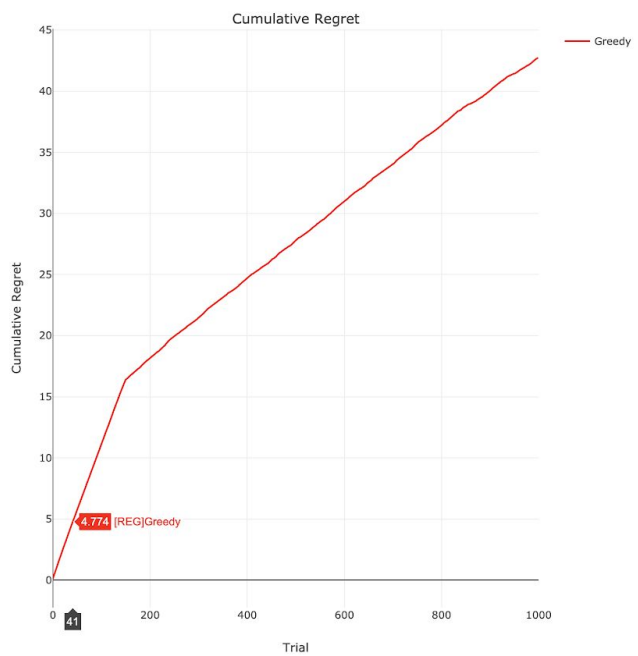
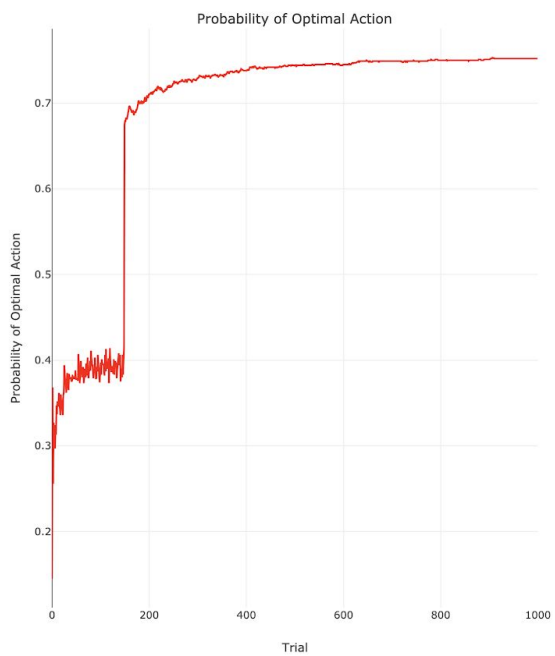
$\epsilon=0.05$



E=0.10



E=0.15



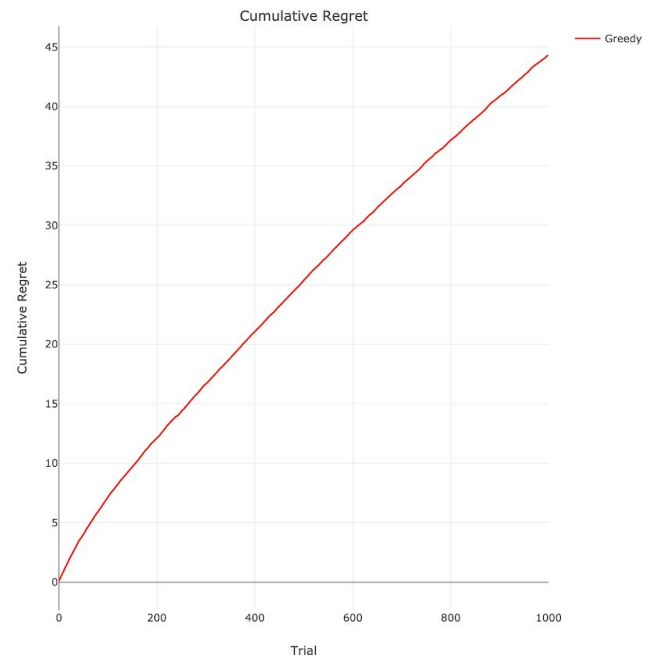
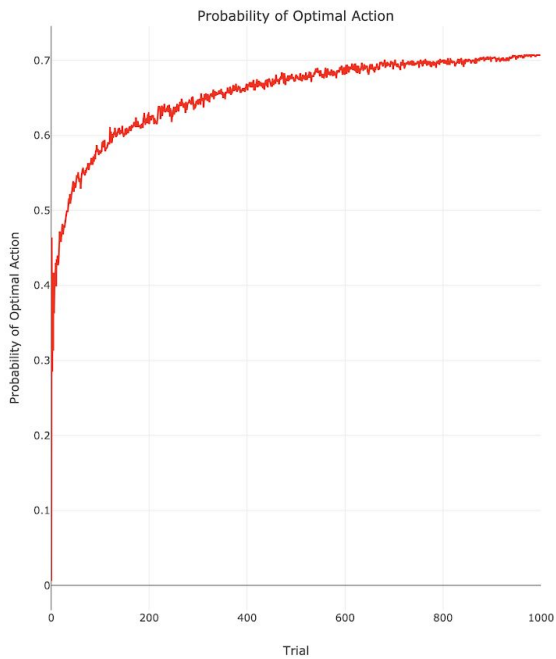
## Problem 4

Expected

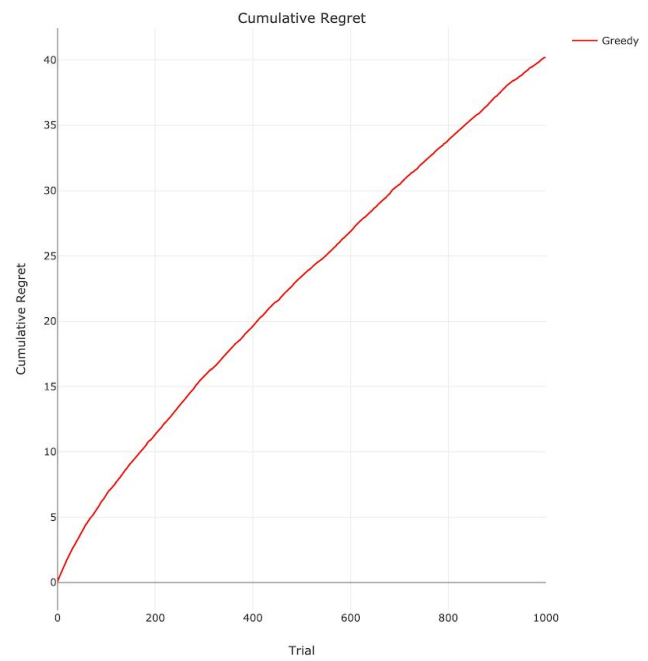
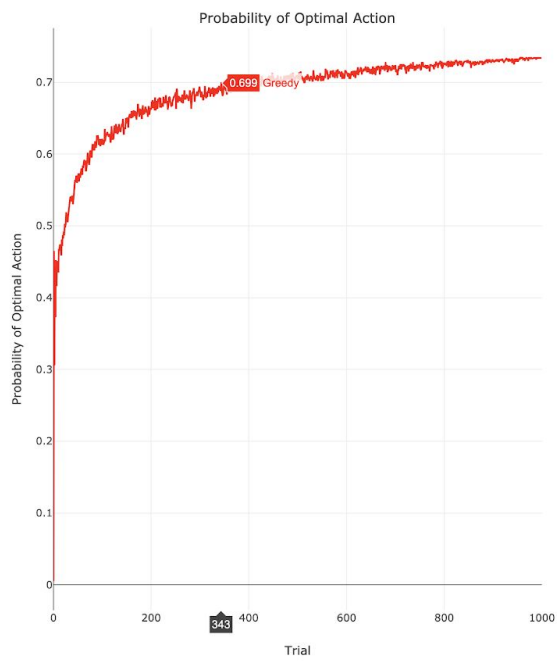
Sim Results

linear decrease:  $(e \cdot (1 - (\text{current trial} / \text{total trials})))$

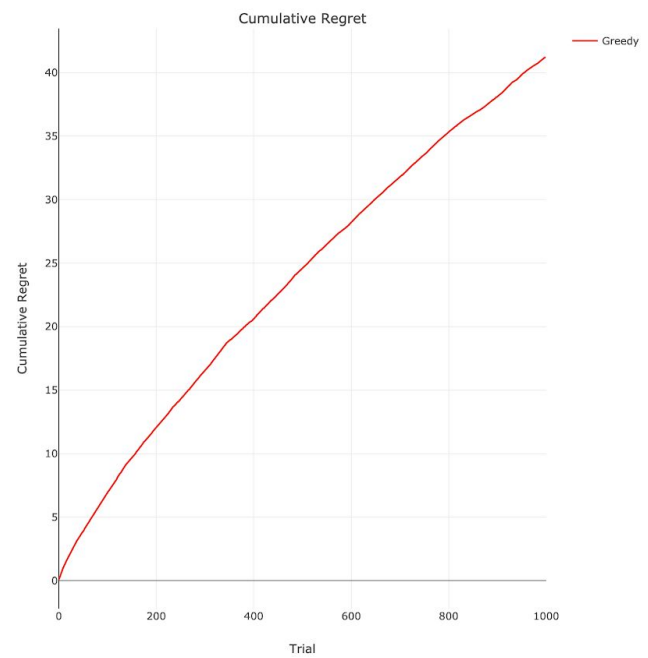
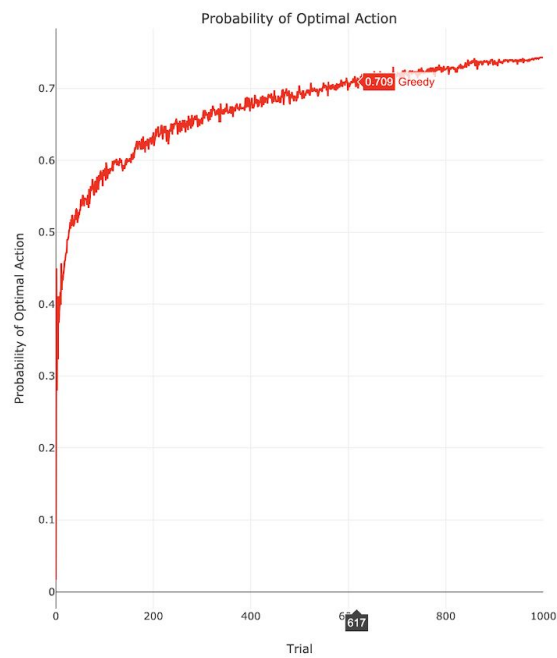
$E = 0.05$



$E = 0.10$



$E = 0.15$

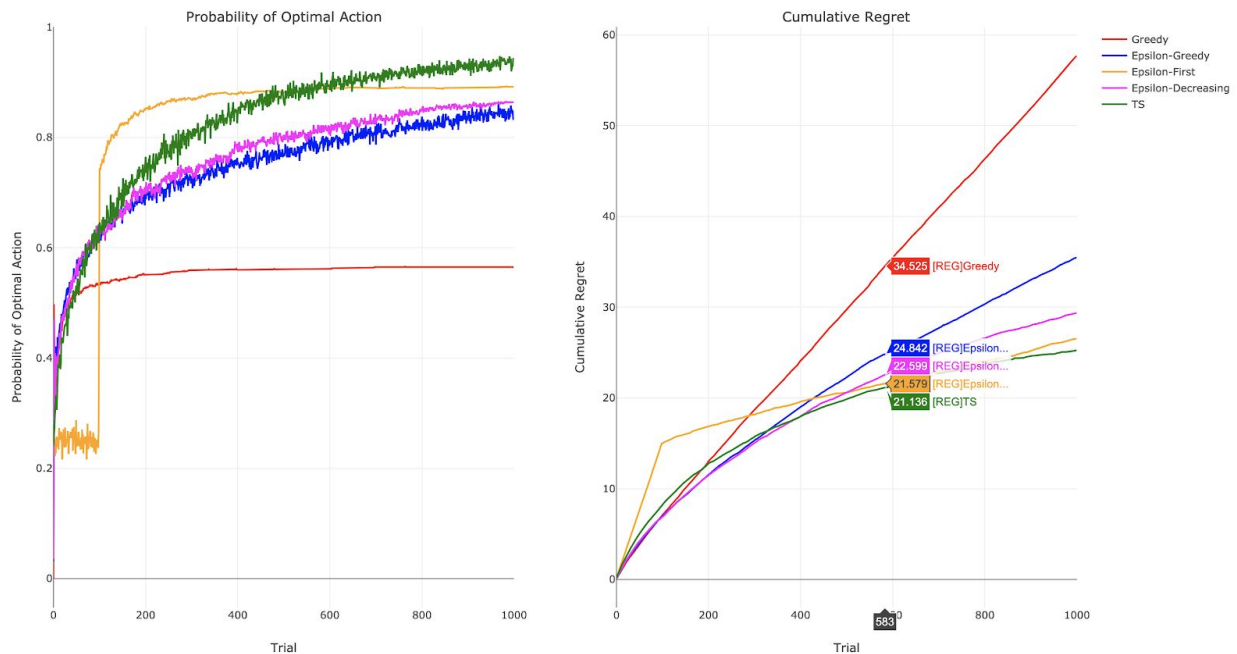


## Problem 5

### Expectations

We used the `np.random.beta` function to run Thompson Sampling

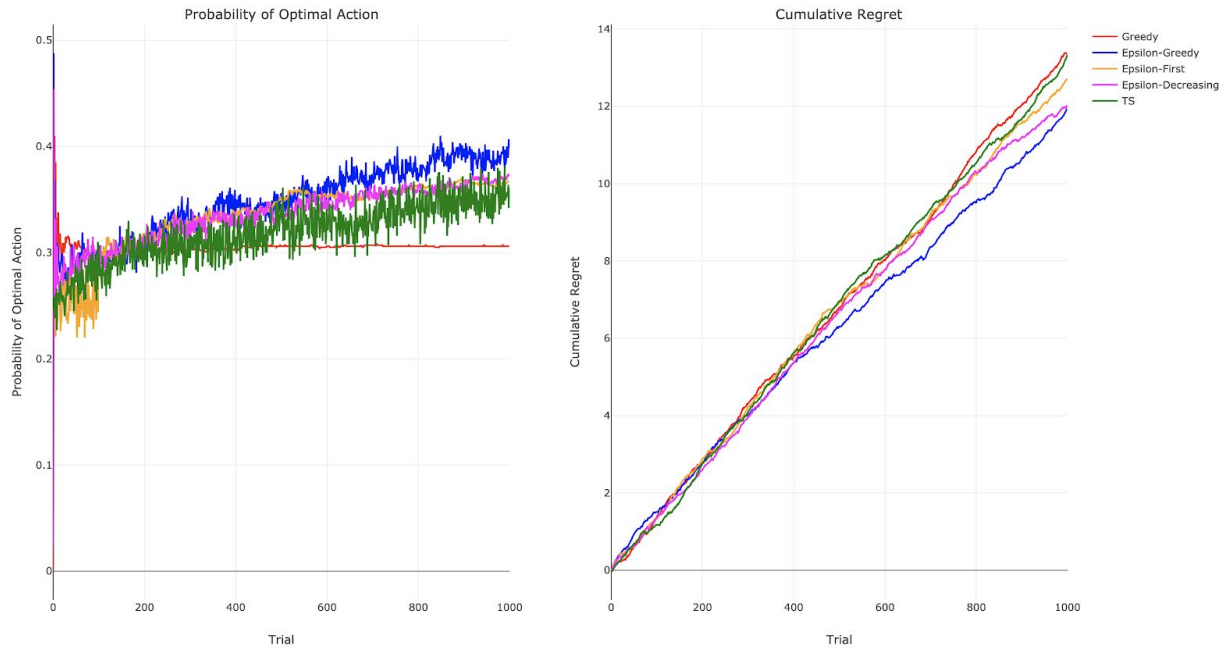
### Sim Results



### Compare

```
# Reward Signal Probabilities
# P(R_t = 1 | do(A_t))
P_R = np.array(
    |   # A = 0      1      2      3
    |   [0.49, 0.52, 0.51, 0.50]
)
K = len(P_R)
```





## Reflect

All of the approaches are much closer to each other having the rewards be so close. Having the reward signals so close naturally decreases overall regret. Having them so close, though, makes it harder to truly identify the best choice for the problem. If we did not know or did not have a finite time  $T$ , Thompson Sampling would end up having the least cumulative regret. As it continues, it will do a better job identifying the optimal choice compared to the other options.