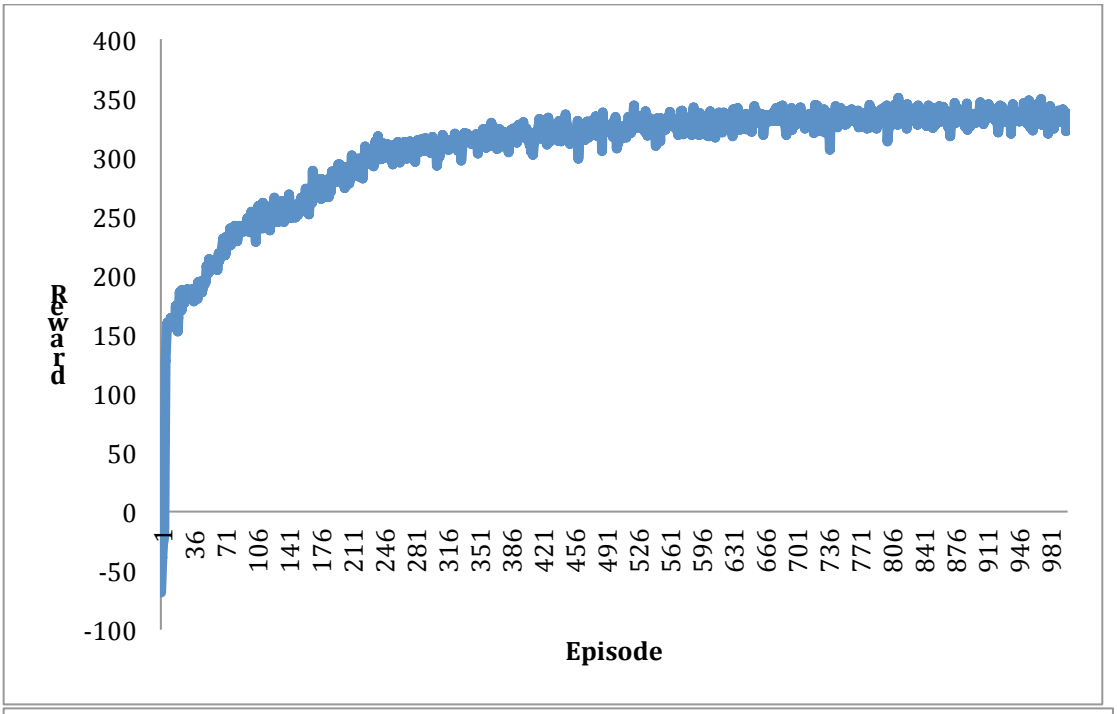
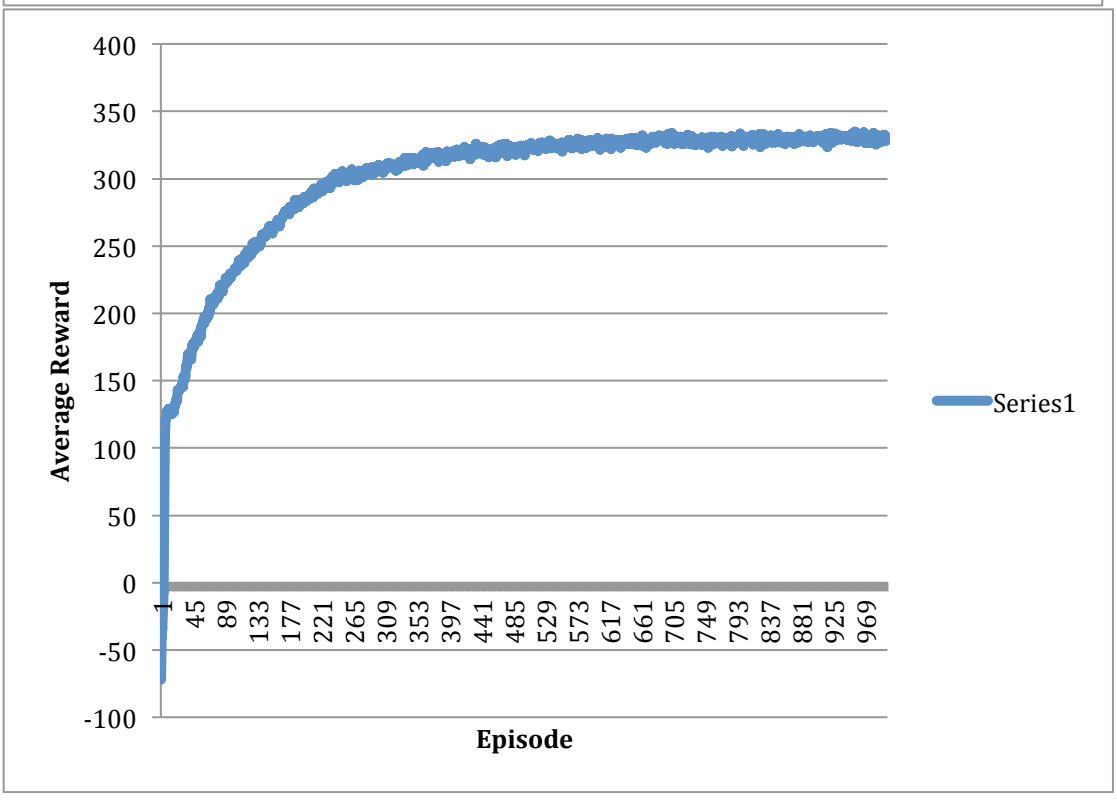
1. A) The reward value is consistently 0.0. In this situation the robot never goes past the slippery parts of the map. This is because there's the possibility for a negative reward if the robot goes past them and there is no randomness to make the robot go past them.
   B) After increasing the epsilon value to 0.1 the robot is able to obtain higher reward values. The first few reward values were below 0.0, however, by the 1000^{th} episode the robot yielded a reward value of 234.0. This is a result of the introduced randomness, which takes the robot past the slippery points of the map.
   C) After increasing the epsilon value to 0.5 the robot gets lower reward values. The reward values were almost always lower than 0.0. This is the result of too much randomness in the system.
2. A) In this simulation the results are consistently negative. The agent doesn't know about the thief and constantly bumps into it. Since the thief is moving the Q-value cannot be updated properly.
   B) When the agent knows about the thief the results are much better. The first few episodes yield a negative reward value. However, the reward value grows quickly and is consistently in the upper 200.0 range.
   C) I performed this investigation through trial and error. Firstly, I experimented with the learning rate. I initially set it to 0.5. With this test the reward values were in the low 200 – mid 200 range. I decided to go lower and set the learning rate to 0.45. After this change I noticed the reward numbers started to go a bit higher. To be thorough I tested the learning rate at higher values 0.65 and 0.8. Both tests yielded reward values much lower than those of the 0.45 test. So I decided to go even lower than 0.45. After testing a few more numbers I noticed the robot yields the best results when the learning rate is 0.07. After finding the optimal learning rate I decided to test the epsilon value. I went through the same testing procedure as I did with the learning rate and found the best epsilon value is also low, 0.01. The learning rate in a Q-Learning Agent determines how newly acquired information is processed compared to older information. When the learning rate is low old information is strongly considered when making choices so the robot learns from its past actions more when the learning rate is low. The epsilon value determines the randomness of the robot's actions. If this value is low the robot can make more educated choices in later episodes.

3.


400
350
300
250
200
150
100
50
0
-50
-100
Reward
1
36
71
106
141
176
211
246
281
316
351
386
421
456
491
526
561
596
631
666
701
736
771
806
841
876
911
946
981
Episode


400
350
300
250
200
150
100
50
0
-50
-100
Average Reward
1
45
89
133
177
221
265
309
353
397
441
485
529
573
617
661
705
749
793
837
881
925
969
Episode
Series1