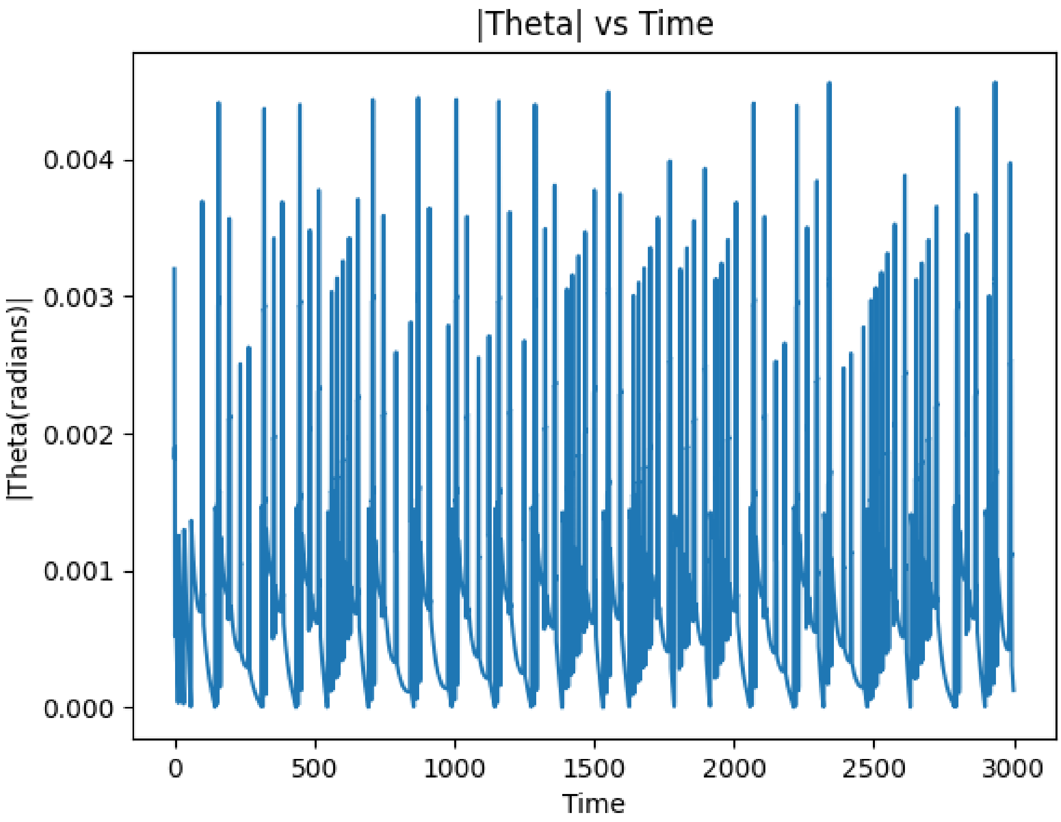
Name: Amani Jammoul
McGill ID: 260381641
COMP 417 - Introduction to Robotics & Intelligent Systems

Assignment 3 RL Controller

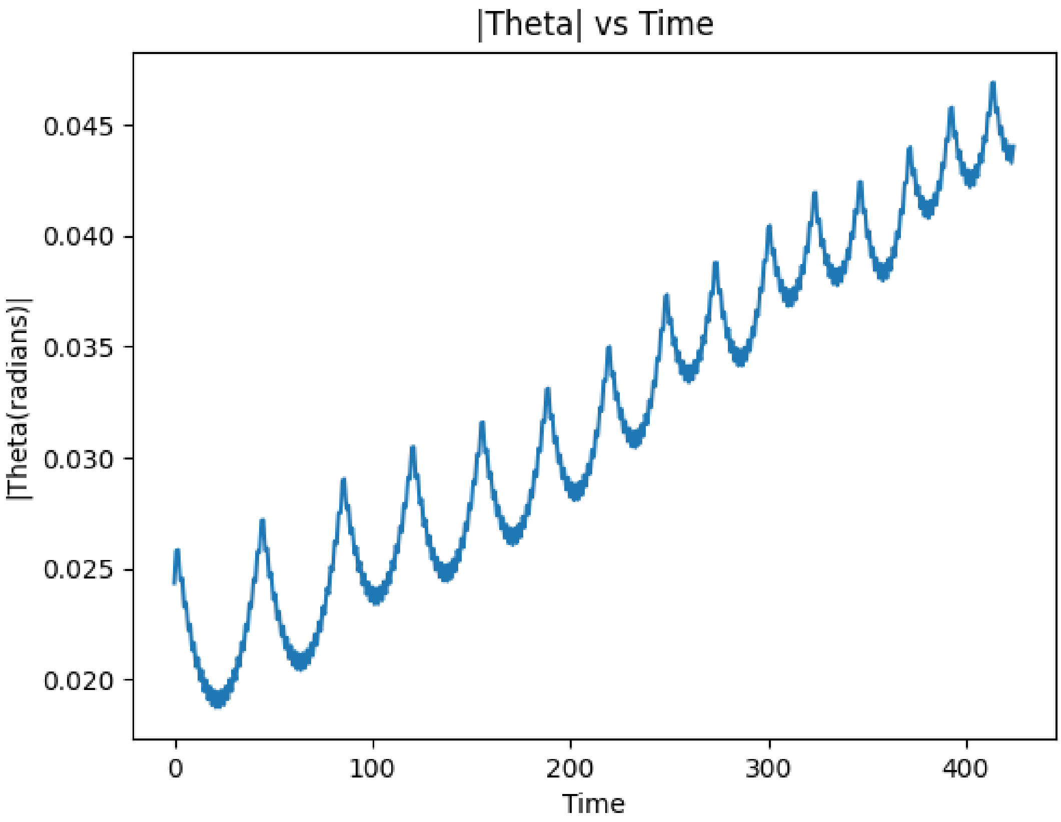
A. The following is the theta vs time plot for my simulation after 400 runs, where the pole angle is as stabilized as I can get it. During this simulation, gamma=0.95 and alpha=0.8. In addition, the inclusion of random actions stops after 200 rounds/episodes.



B. After applying this change, the agent performs worse than in part A (where the value was 0.8). After the same number of game rounds as in part A (400), the agent still failed to consistently keep the pole upright. Note: As in part A, random actions are no longer taken after 200 episodes. The role of the random action is to increase exploration. Since Q-Learning is a policy free RL algorithm, it uses random actions to explore the space and actions in order to learn and maximize total reward. All state values and Q values are initialized to zero at the start, so without the inclusion of some randomness, the agent would always choose the same action.

Random actions allow the agent to take new steps and explore their rewards, all while updating its Q table (which should converge to optimality). This is why, when we decreased the probability of choosing a random action (from 0.2 to 0.1), we reduced the exploration. And since I stopped choosing these values after 200 episodes (same as in part A), the agent did not do as much exploring, and therefore performed worse.

The following is the theta vs time plot for the same simulation time as in part A. Notice how it only lasts 400 timesteps, meaning it failed early (in part A, it lasted the entire time, 3000 timesteps). In addition, the pole angle fluctuates and does not remain close to 0.



C. This could be because the state currently only considers theta and theta_dot. It does not encompass the cart's position or velocity. A possible solution would be to include x and x_dot in the state, and apply q-learning with this. In this case, the reward would also consider how close the cart is to the center of the screen (if that is the goal). Another general improvement could be to discretize the space into more grid cells (i.e., instead of having a 40x40 grid for the states, the dimensions would be greater). This would lead to more precision, better action choices, and quicker learning.

D. State value matrix

```
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0

0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0
```

```
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
2.0129646609408005   2.0229488   1.36288   1.4100000000000001   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
18.922910574323144   18.354362177254714   5828.016640603653
1.4100000000000001   0.7050000000000001   0.7050000000000001   0.743   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
13331.140741825213   13424.539885069507   13423.92411182661
3939.5011308828343   0.0   0.0   0.743   0.781   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
13306.423724477801   59942.62558215666   59943.25420653638
13252.619463954126   18.248662434388237   4.463447792366544   0.0   0.0
0.7810000000000001   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
11064.127589021233   59941.95907418857   59930.27594895822
13966.471286871927   9513.240205209997   1.5620000000000003
0.7810000000000001   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0
```

```
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   11692.76263558904   11693.361829137428   20.7474291243858
20.14294967935747   11.627639978893999   12.7827717351374   6.6755320380892975
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   13.820068852059219   13.18140158066809   11.774731587195486
11.842580121827739   10.800744525811846   4.527327429091672
3.2252036562531807   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   7.03719293557481   3.7169321914625515   11.250818859882772
10.071437124882818   9.573774778449826   6.2274533430716374
2.5040137501524535   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   6.498220825563076   7.4444287641576175
8.845253646147478   8.215564624298514   4.667770111115434   3.0040004346615343
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   8.03261158663824   6.98789973398511
7.642996344983587   2.4621854812615376   3.2723223731882047   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   5.778829230598854   6.19217553311299
6.220650932488553   5.721006784622237   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   6.02097966256295
5.8624403759589265   5.324059597752102   4.691686266372841
0.47699999999999987   0.0   0.0   0.0   0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   5.283537384888154
5.073130008050249   4.384525151656656   1.5362019432220262   0.74376   0.0   0.0
0.0   0.0   0.0   0.0

0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   4.07264154151342
4.0474967439799965   3.629137029364241   2.0495993394470915   0.0   0.0   0.0   0.0
0.0   0.0
```

```
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  3.729245525346339
3.315242765491493  2.893449413979749  0.9573071999999999  0.0  0.0  0.0  0.0
0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
2.2039986943690084  2.693764990602462  2.3759834157715614
0.3249999999999999  0.0  0.0  0.0  0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1.7635059598174603  2.214954650337546  1.655429512997575  0.0  0.0  0.0  0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.8609941927004588  1.4700216806133666  1.4151645853218606
1.0911901945026121  0.0  0.0  0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1.071306395712898  1.0431783875736373  1.0372514245237134
0.39312000000000025  0.0  0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.6893623343651673  0.7655862023932065  0.5441467571391657
0.1730000000000001  0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.4982952815255959  0.44580403549931485  0.31742093567999996  0.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.30403855923545936  0.2908899261984049  0.20424482689843176
0.0969999999999999

0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.14488451583999976  0.1366310479224718  0.07655839999999987
```

```
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0
0.0   0.10383999999999982   0.0589999999999999   0.0
```

E. Reminder of my values in part A: gamma = 0.95, alpha = 0.8

When I adjusted the gamma to a value close to 0 (gamma=0.1), the pole simply fell to the ground, even after a long while. This indicates that the learning process is not efficient like this. Gamma is used as a discounting factor to quantify the importance of future rewards relative to the current reward. If it is too low, the agent will only consider immediate rewards, without placing weight on future rewards. Since immediate rewards are not enough to evaluate the value of a [state, action] pair, a lower gamma rate leads to an inefficient learning technique, and therefore an inefficient agent.

In order to get my results from part A, I had to increase the value of alpha. When alpha was close to 0 (0.001), the agent performed poorly. This is because alpha determines how much to accept the new q value. At a high rate, the controller takes large strides in updating q values, which leads to faster learning. When it is too low, the rate of learning decreases, and therefore it takes the agent a lot longer to get a proper q values table. However, after many episodes, these large updates become less useful and harmful at times. This is why I believe that, in order to improve this controller, alpha should be set to a high value at the start, and then it should decrease gradually. Because as time goes on, the agent grows its knowledge base and gets a more accurate understanding of the states and which actions to take.