Shaun Mathew

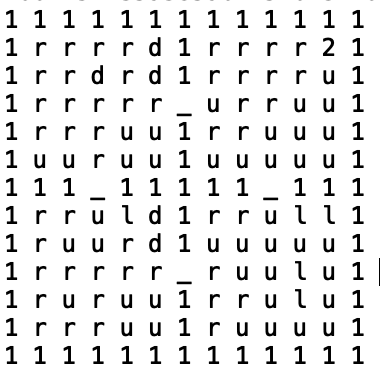
**Comparison of all Algorithms**

Note: I have included the graphs for every alg except MC at the end since there are a lot of them.

**On-Policy First Visit MC Control Results:**

Took about 150,000 episodes to converge to the optimum

Optimal Policy:

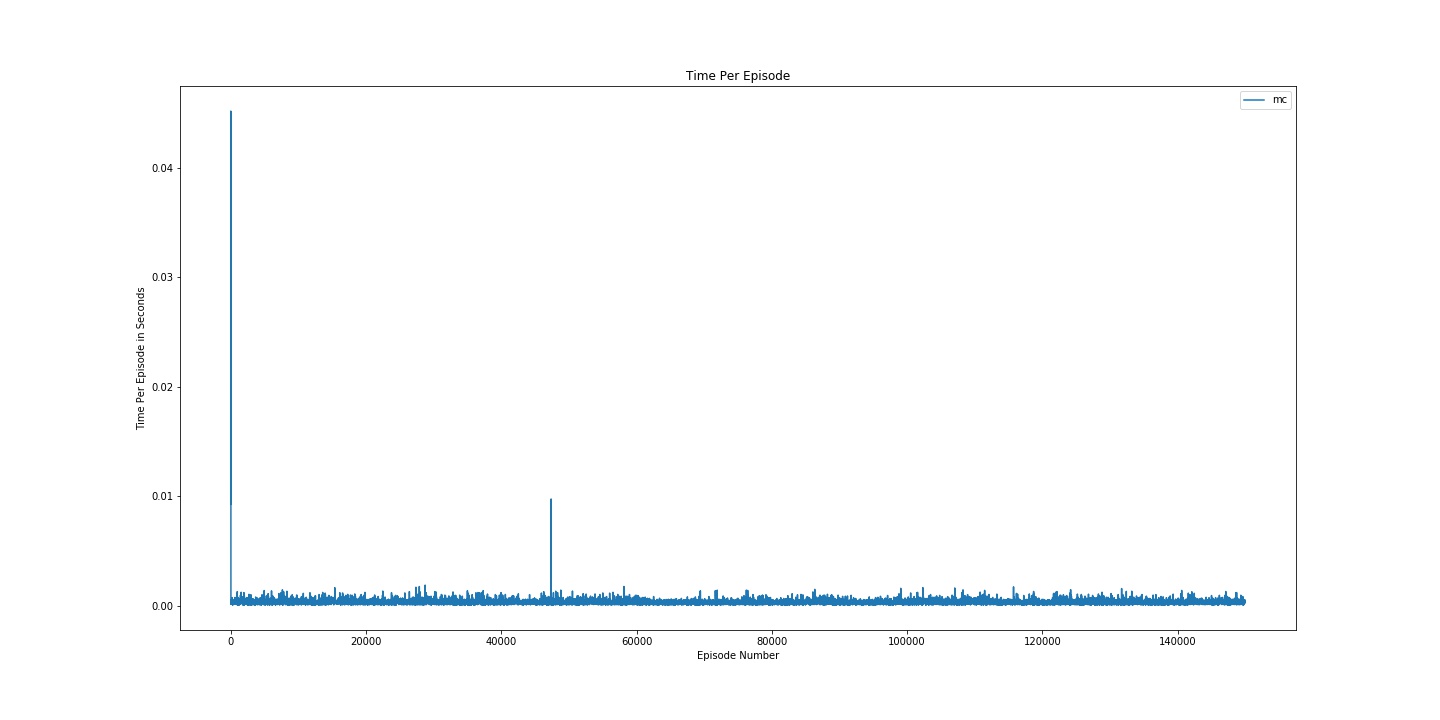


2 is goal state, 1 is wall and \_ is door

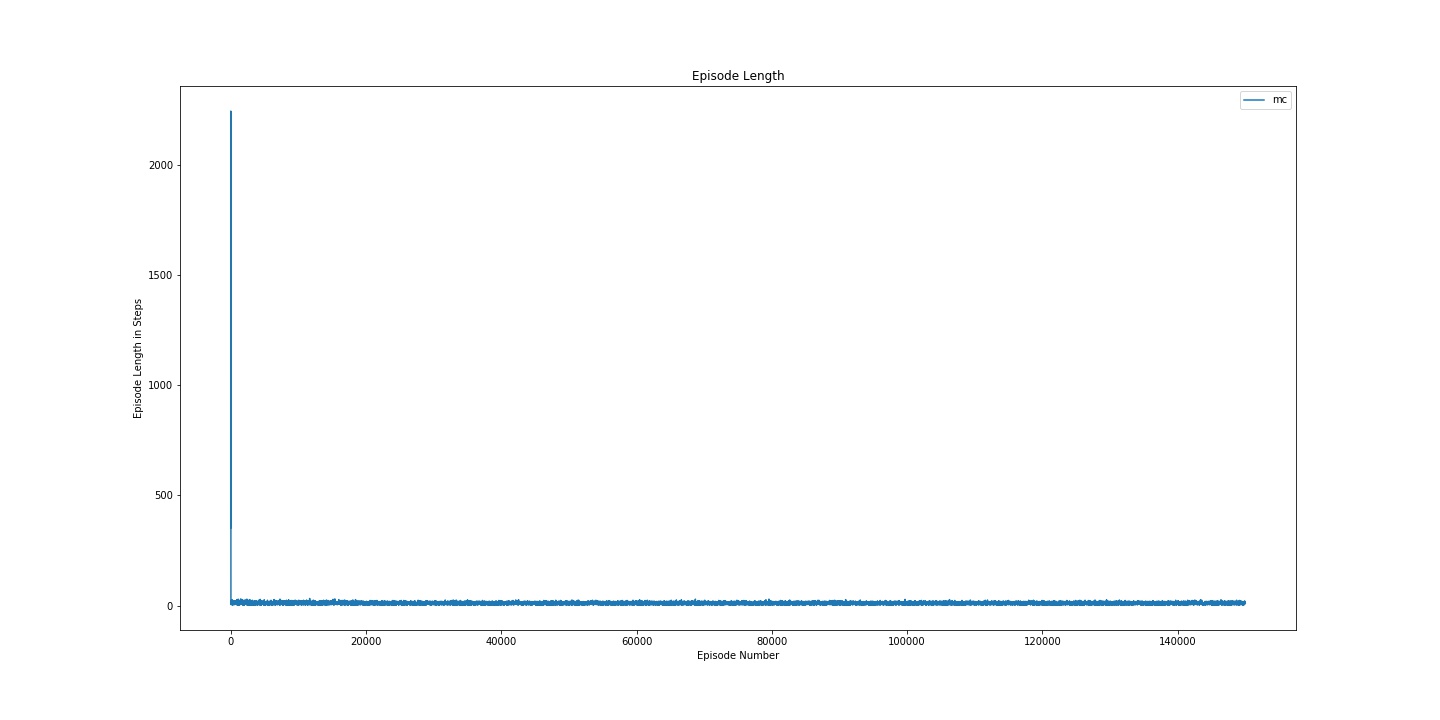
**Total Computation Time:**

﻿57.35938596725464s

**Time Per Episode**



**Number of Steps Per Episode**



The On-Policy Epsilon Soft First Visit Monte Carlo Algorithm took more episodes to converge than the other algorithms. It starts with a high number of steps per episodes but quickly plateaus. Episode execution times are much longer in comparison to the temporal difference methods.

**Fixed Epsilon**

The algorithms were run for 40000 episodes to ensure convergence for all algs and alphas. These are the results averaged over 3 runs.

**Total Computation Time**

*Alpha 0.05*

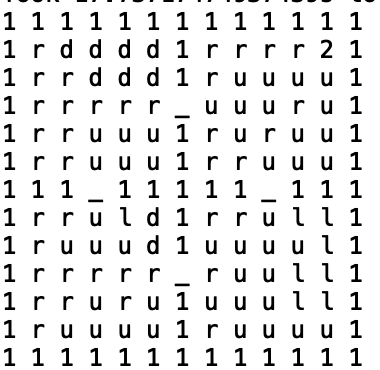
Sarsa: ﻿ ﻿﻿12.199689706166586s

﻿Q: ﻿ 12.230715354283651s

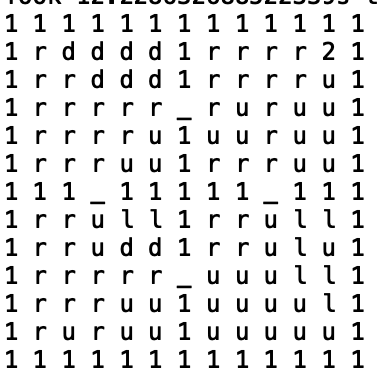
﻿Expected Sarsa: ﻿﻿14.82646099726359s

Double Q: ﻿ ﻿﻿17.488329728444416s

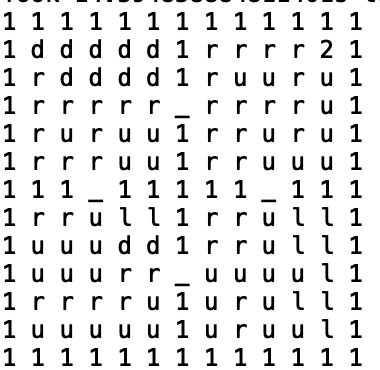
Sarsa Policy



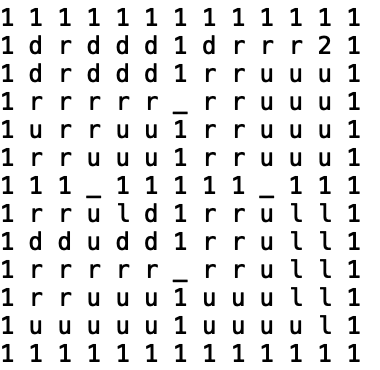
Q Policy



Expected Sarsa Policy



Double Q Policy



*Alpha 0.1*

Sarsa: ﻿ ﻿﻿﻿11.239166021347046s

﻿Q: ﻿﻿11.420618057250977s

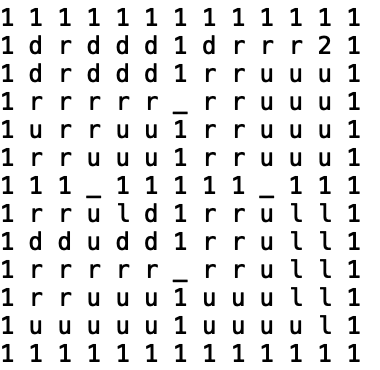
﻿Expected Sarsa: ﻿﻿14.035706202189127s

Double Q: ﻿ ﻿﻿16.795975923538208s

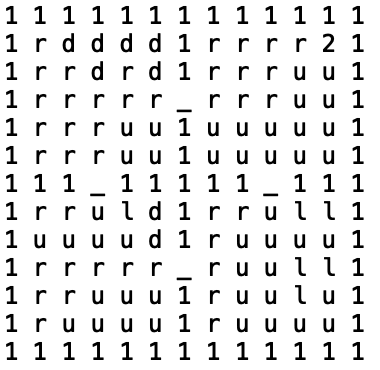
Sarsa Policy



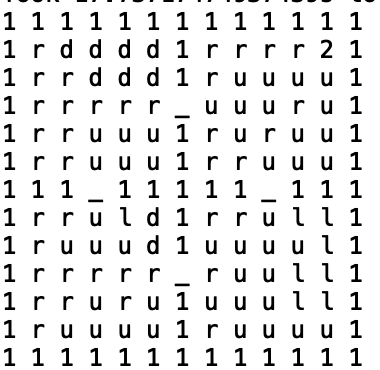
Q Policy



Expected Sarsa Policy



Double Q Policy



*Alpha 0.2*

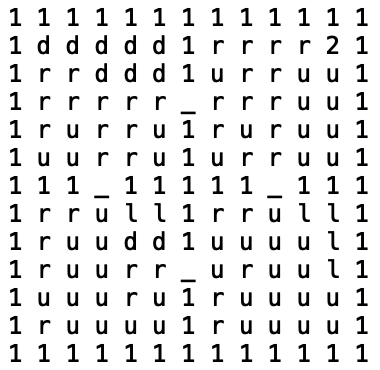
Sarsa: ﻿ ﻿﻿ ﻿﻿11.743133942286173s

﻿Q: ﻿ ﻿12.124346017837524s

﻿Expected Sarsa: ﻿ ﻿﻿14.819084326426188s

Double Q: ﻿ ﻿﻿﻿﻿17.263063033421833s

Sarsa Policy



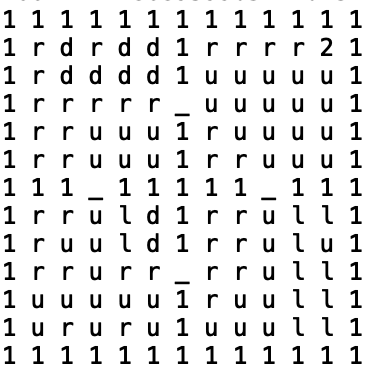
Q Policy



Expected Sarsa Policy



Double Q Policy



Expected Sarsa and Double Q have the longest total running times in all cases. Sarsa consistently has the shortest running times. Running Times for different alpha show no discernible pattern.

Looking at the average times per episode (Figures 1 – 15), we see that is very noisy and so we attempt to gain some more insight by computing a running average.

Using a moving average (Figures 16 – 30), we can clearly see that a similar trend holds. Sarsa is the quickest, followed by q learning, leaving expected sarsa and double q learning as the slowest.

**Number of Steps per Episode Running Average**

We saw that the running average was a good choice to combat noise and hence, we have opted to use that.

From Figures (31 – 45), Number of Episodes taken decreases with time, showing that the algorithms are converging. They converge to about the same number of steps per episode. Sarsa and Q tend to take fewer steps initially than Expected Sarsa and Deep Q Learning. Alpha values seem to have an effect on the initial number of steps taken by the algorithm. For an alpha of 0.1, Sarsa and Q take the most initial steps when compared to other alpha values. Higher alphas for Expected Sarsa reduce the number of initial steps taken by the algorithm.

**Variable Epsilon**

Fewer episodes were needed to achieve convergence when a variable epsilon was used. The function used e^(-t/700) if (t+1) >= 1000 else 0.25. This ensures a good chance of exploration early on while adopting a greedier search as the number of episodes increases. The

algorithms were run for 20000 episodes to ensure convergence for all algs and alphas.

**Total Computation Time**

*Alpha 0.05*

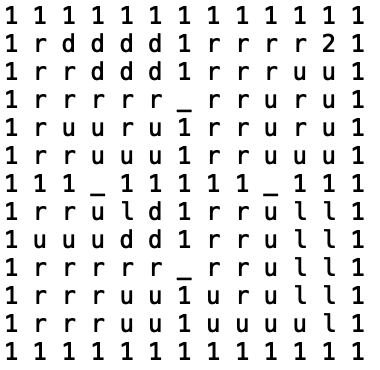
Sarsa: ﻿ ﻿6.702401638031006s

﻿Q: ﻿ ﻿﻿6.869043429692586s

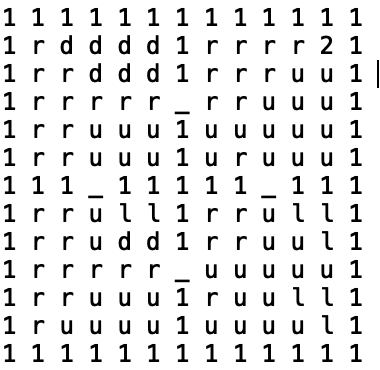
﻿Expected Sarsa: ﻿﻿ ﻿ ﻿8.725509723027548

Double Q: ﻿ ﻿﻿ ﻿12.172817627588907s

Sarsa Policy



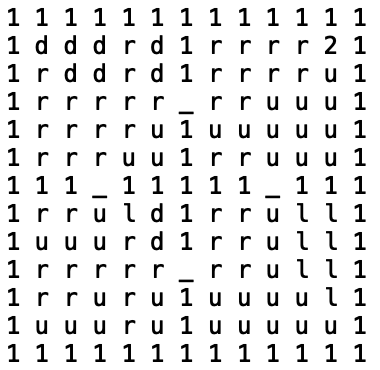
Q Policy



Expected Sarsa Policy



Double Q Policy



*Alpha 0.1*

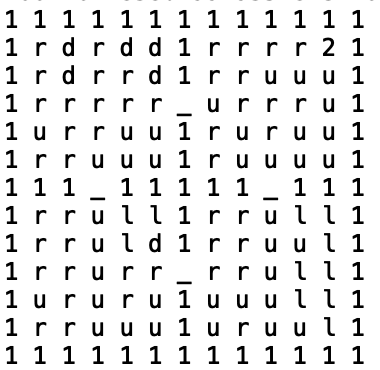
Sarsa: ﻿ ﻿﻿﻿ ﻿6.526455720265706s

﻿Q: ﻿ ﻿6.811911980311076s

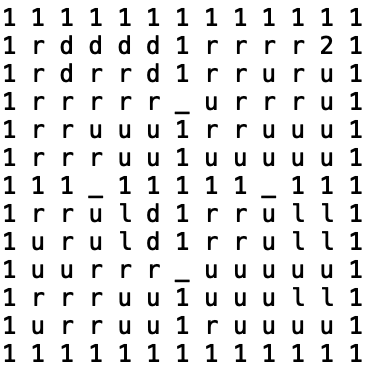
﻿Expected Sarsa: ﻿﻿ ﻿ ﻿8.085472583770752s

Double Q: ﻿ ﻿﻿ ﻿ ﻿10.394423961639404s

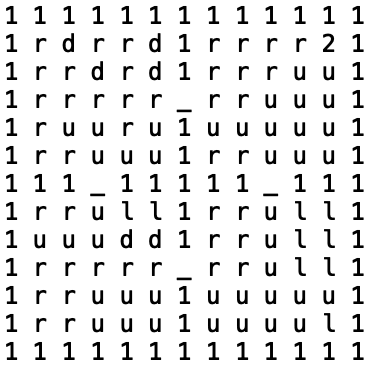
Sarsa Policy



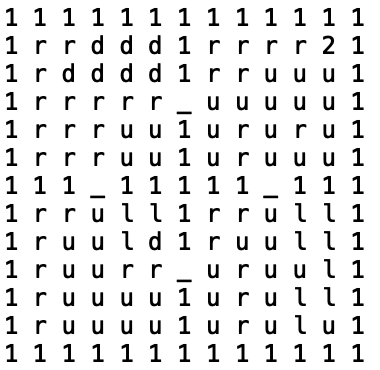
Q Policy



Expected Sarsa Policy



Double Q Policy



*Alpha 0.2*

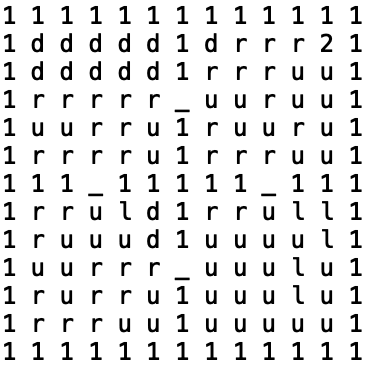
Sarsa: ﻿ ﻿﻿ ﻿﻿ ﻿ ﻿6.25777530670166s

﻿Q: ﻿ ﻿ ﻿ ﻿6.394557396570842s

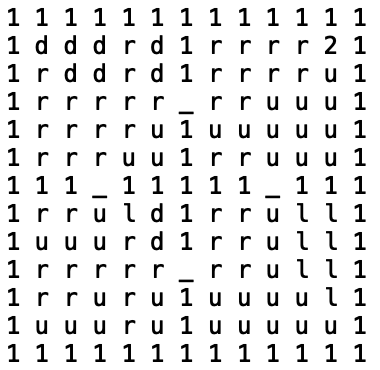
﻿Expected Sarsa: ﻿ ﻿﻿ ﻿ ﻿8.023312330245972s

Double Q: ﻿ ﻿﻿﻿﻿ ﻿10.52975312868754s

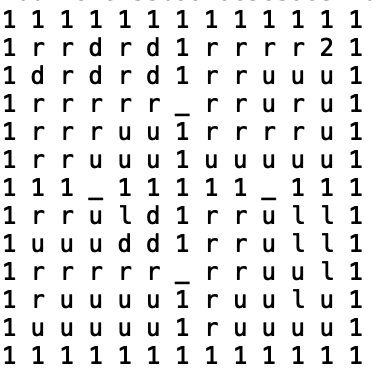
Sarsa Policy



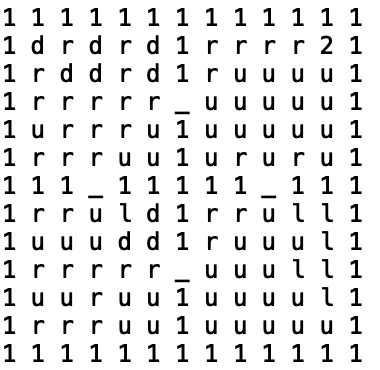
Q Policy



Expected Sarsa Policy



Double Q Policy



We see a similar result here. Sarsa is the quickest with Q learning right after. Expected Sarsa and Double Q Learning still take the longest amount of time.

**Computation Time Per Episode**

We see a similar trend to the fixed epsilon. Sarsa and Q Learning are the quickest whereas the other two are slower in comparison. We also see that the time per episode is lower initially in comparison to a fixed epsilon.

**Number of Steps per Episode**

The number of steps per episode is much higher initially in comparison to a fixed epsilon. This is to be expected because of the higher exploration parameter at the start. Overall, there is a similar trend in that they converge to a similar number of steps.

**Fixed Epsilon**

**Computation Time Per Episode**

*Alpha 0.05*

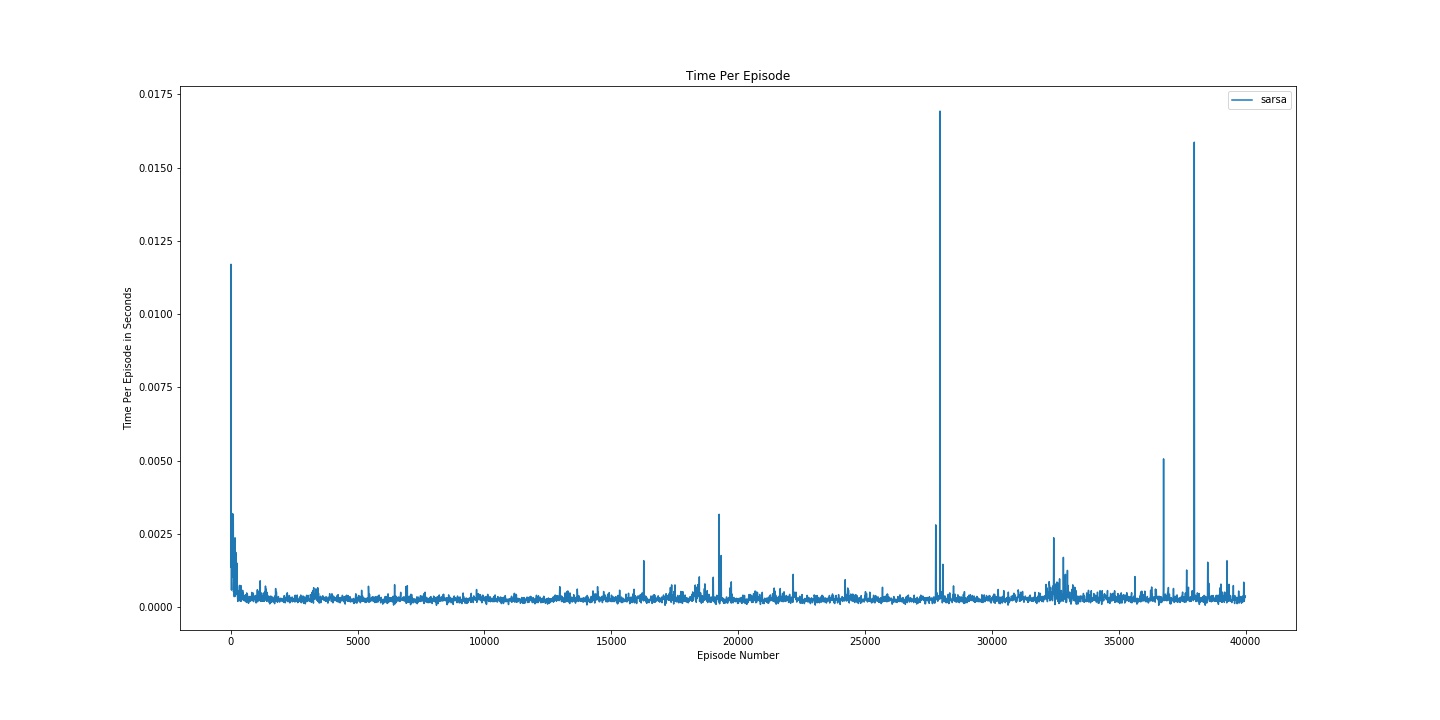


Figure : Sarsa Time Per Episode (a = 0.05) Fixed Eps



Figure : Q Time Per Episode (a = 0.05) Fixed Eps

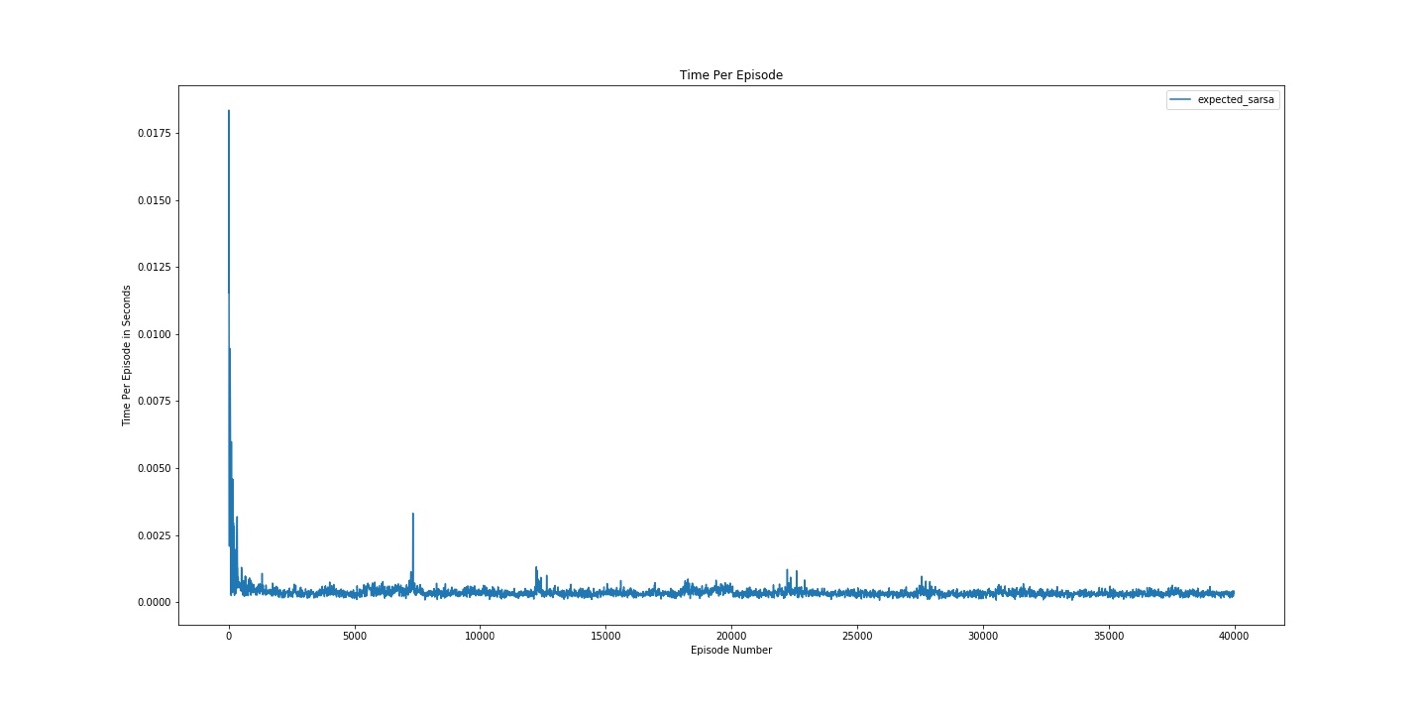


Figure : Expected Sarsa Time Per Episode (a = 0.05) Fixed Eps

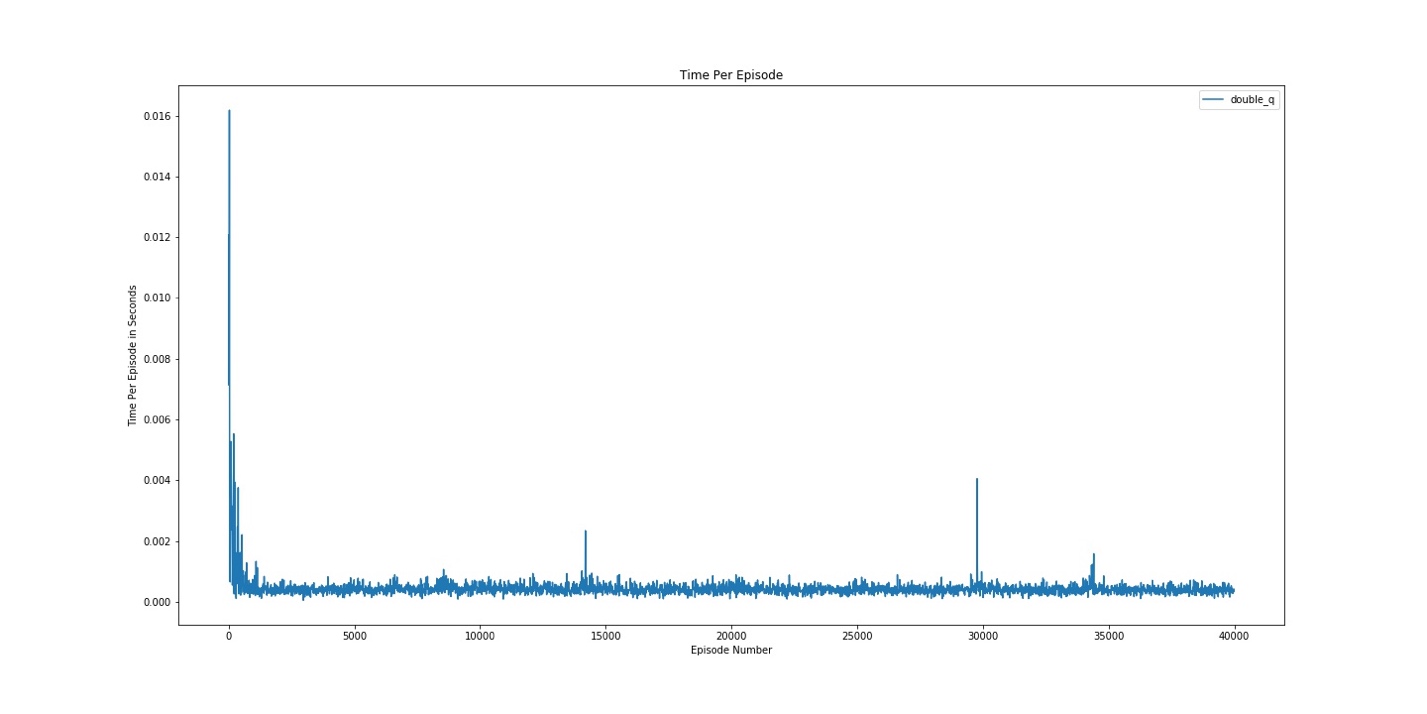


Figure : Double Q Time Per Episode (a = 0.05) Fixed Eps

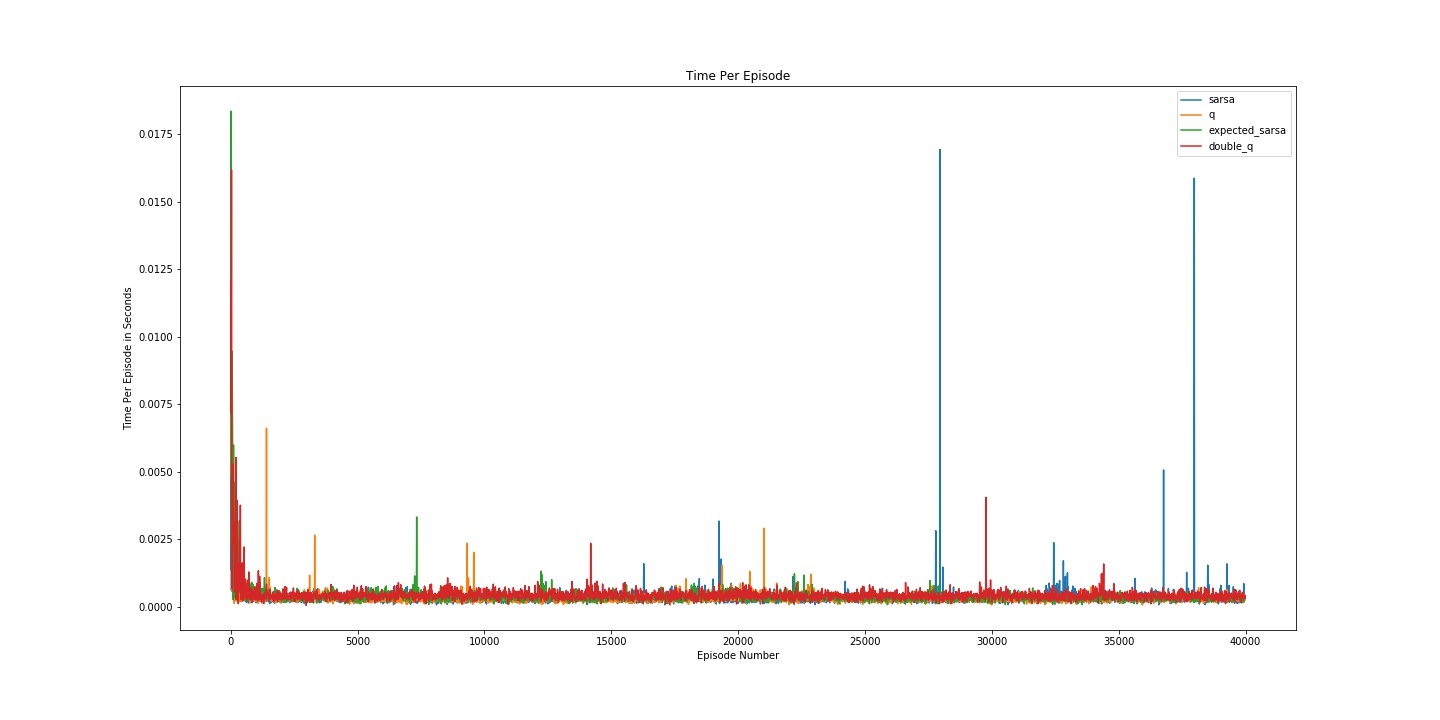


Figure : Combined Time Per Episode (a = 0.05) Fixed Eps

*Alpha 0.1*

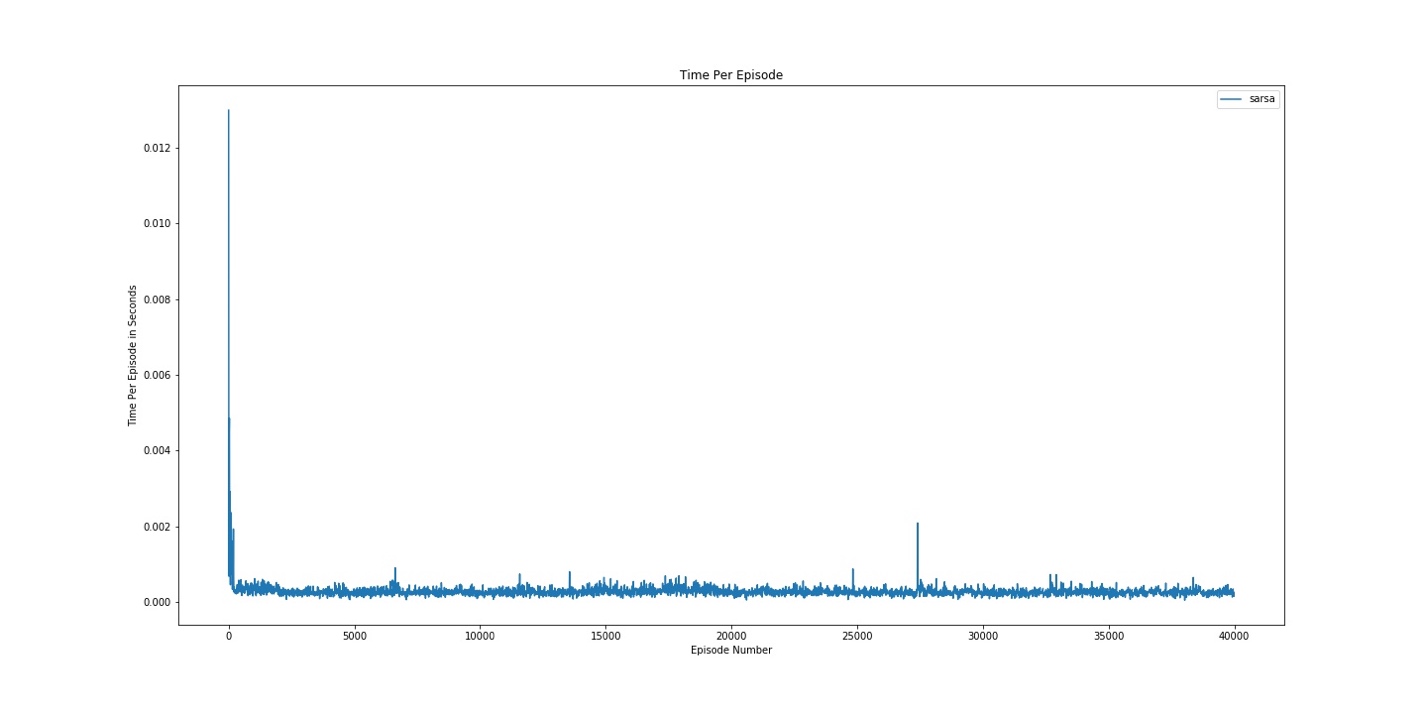


Figure 6: Sarsa Time Per Episode (a = 0.1) Fixed Eps



Figure 7: Q Time Per Episode (a = 0.1) Fixed Eps

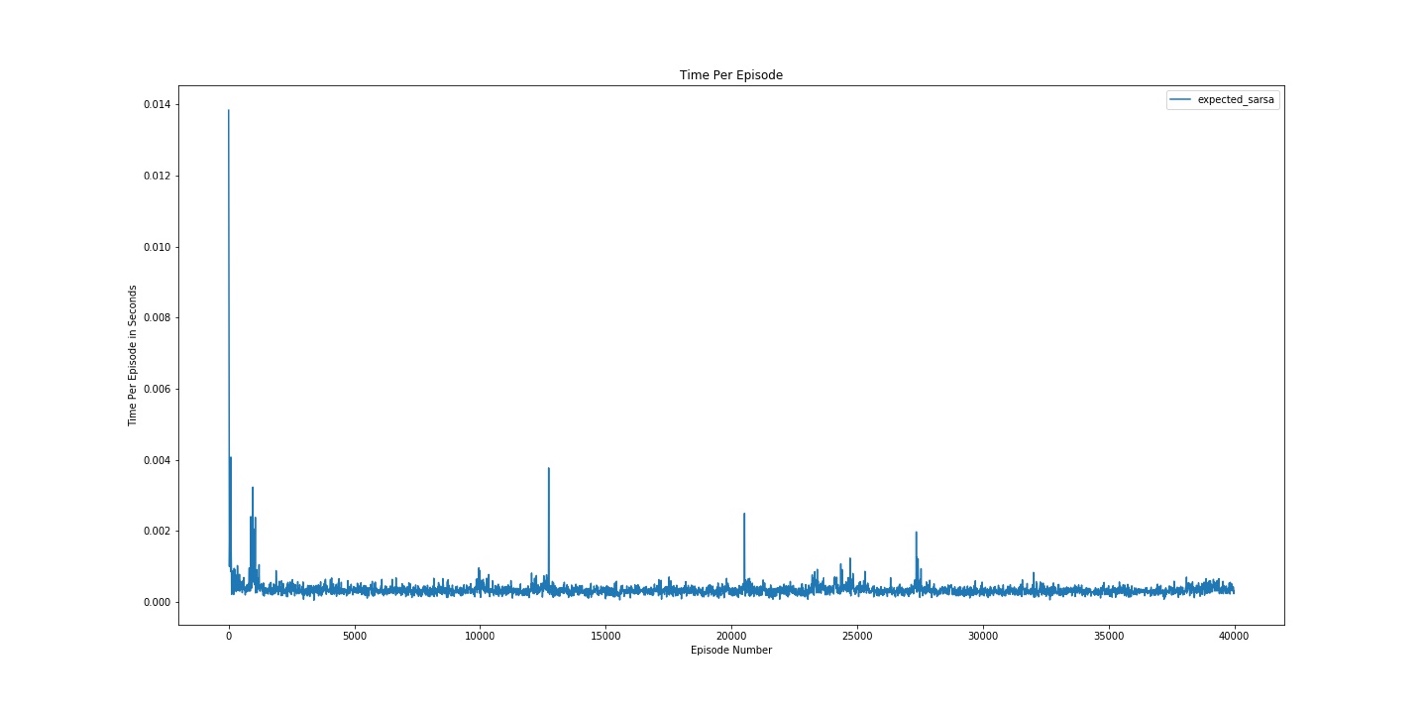


Figure 8: Expected Sarsa Time Per Episode (a = 0.1) Fixed Eps

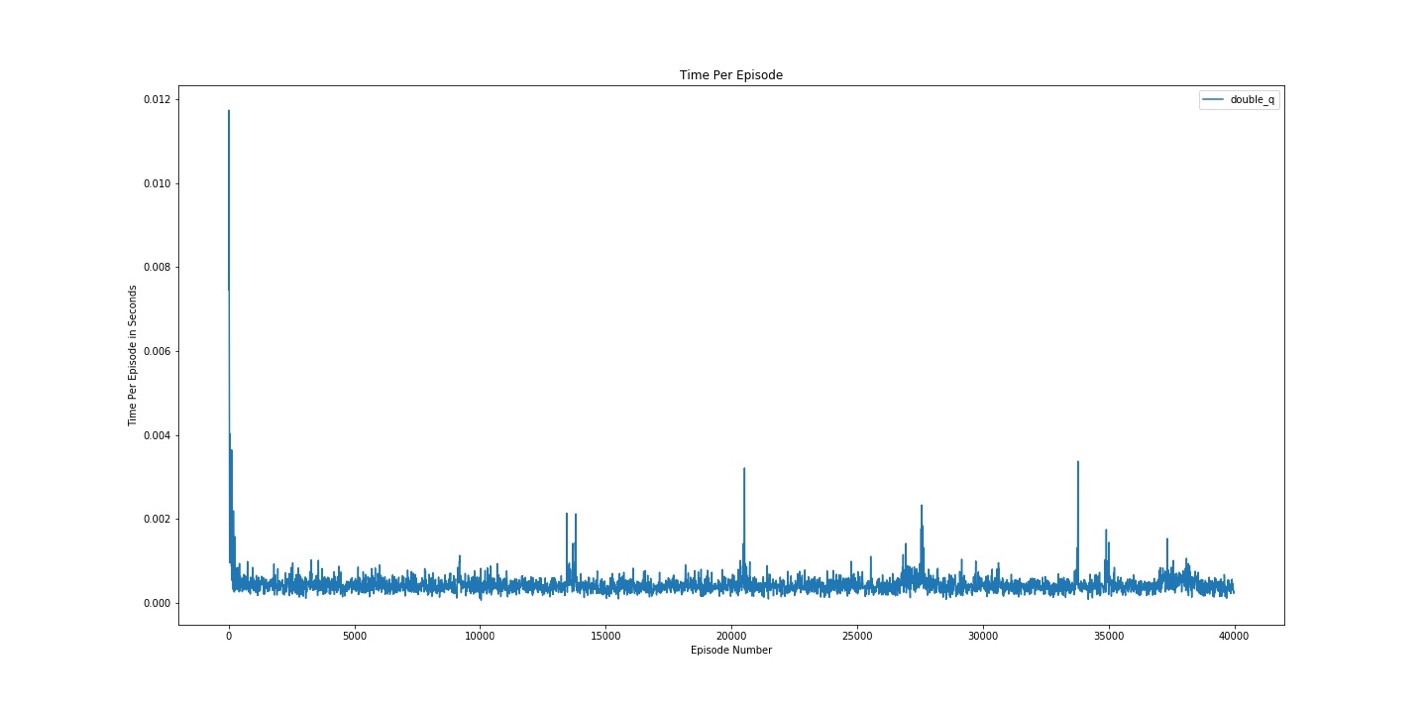


Figure 9: Double Q Time Per Episode (a = 0.1) Fixed Eps

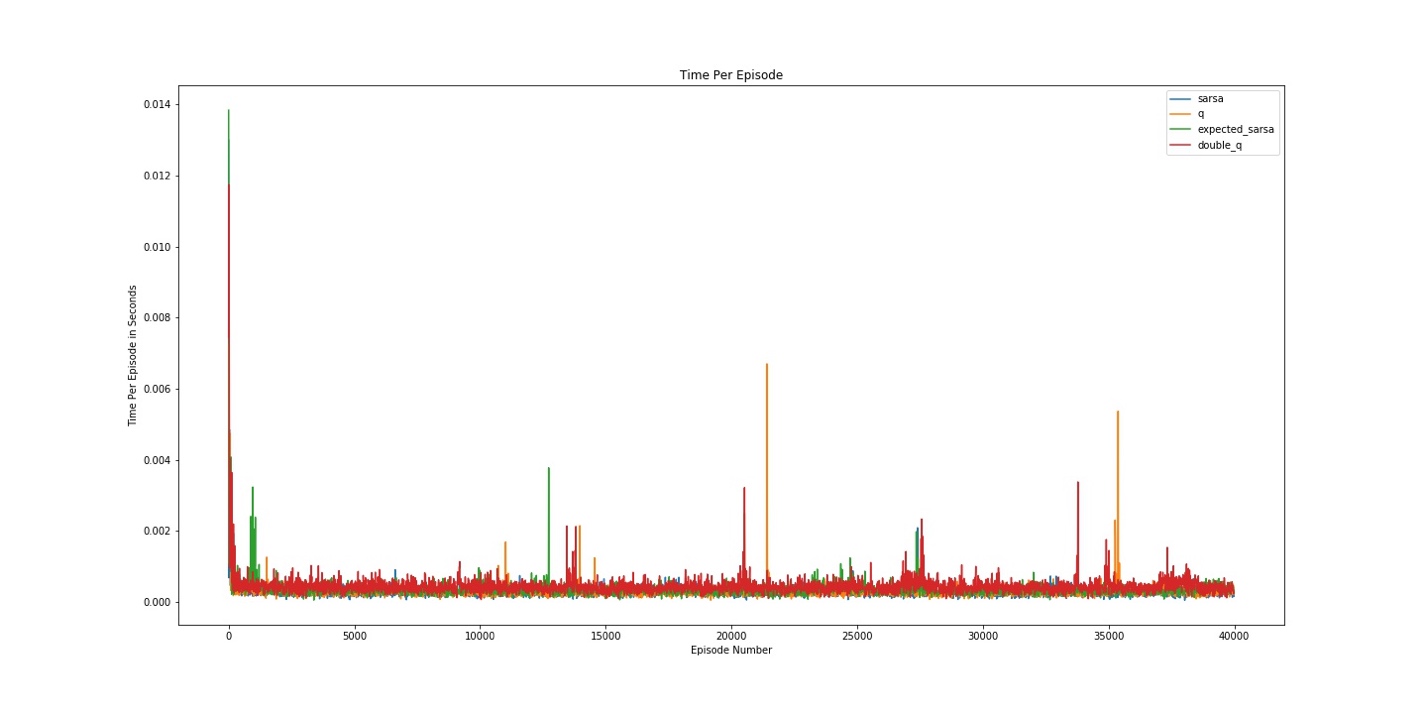


Figure 10: Combined Time Per Episode (a = 0.1) Fixed Eps

*Alpha 0.2*

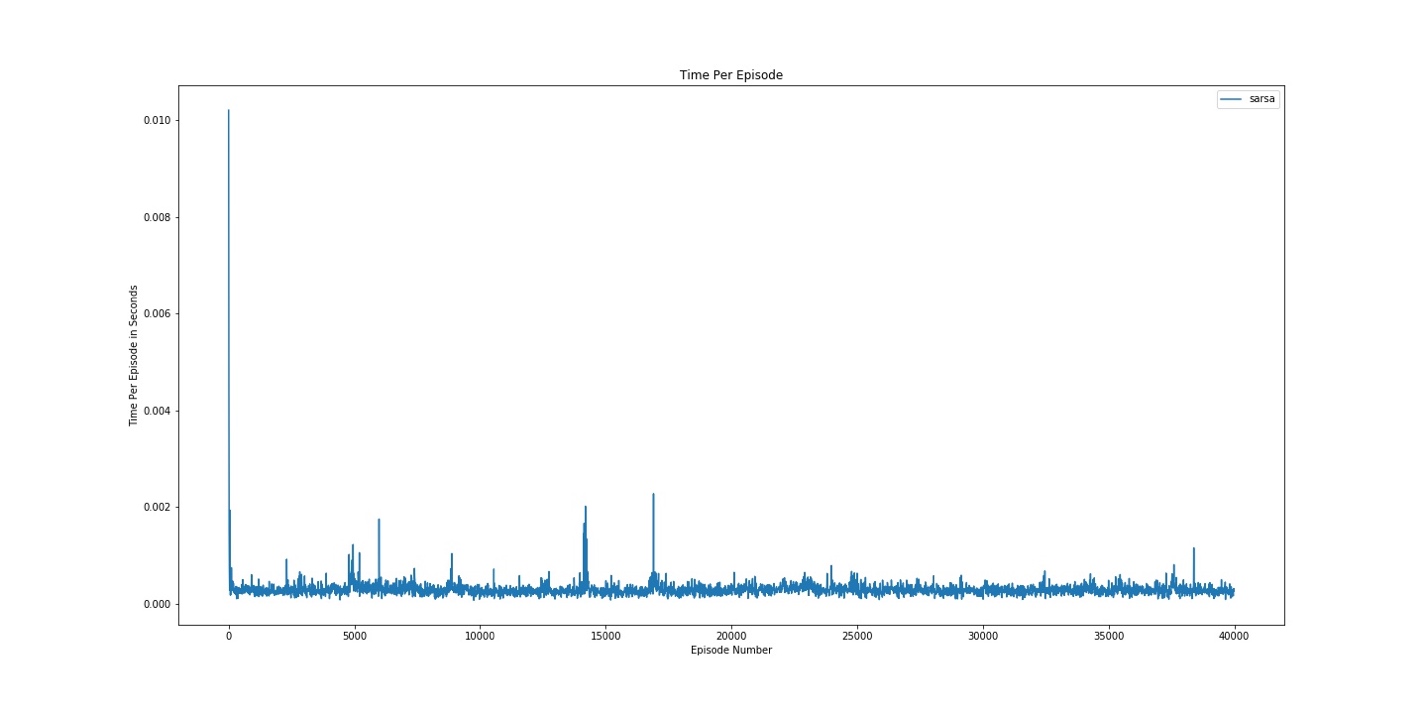


Figure 1: Sarsa Time Per Episode (a = 0.2) Fixed Eps

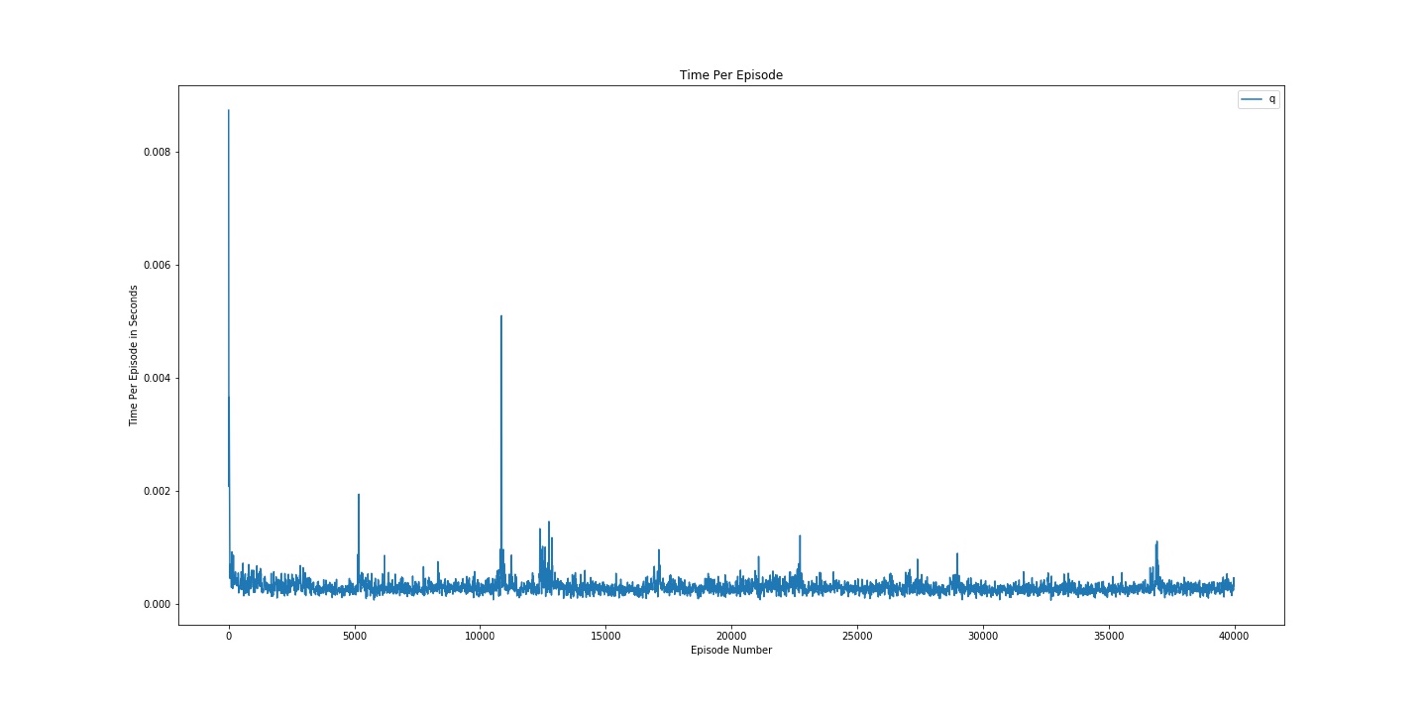


Figure 1: Q Time Per Episode (a = 0.2) Fixed Eps

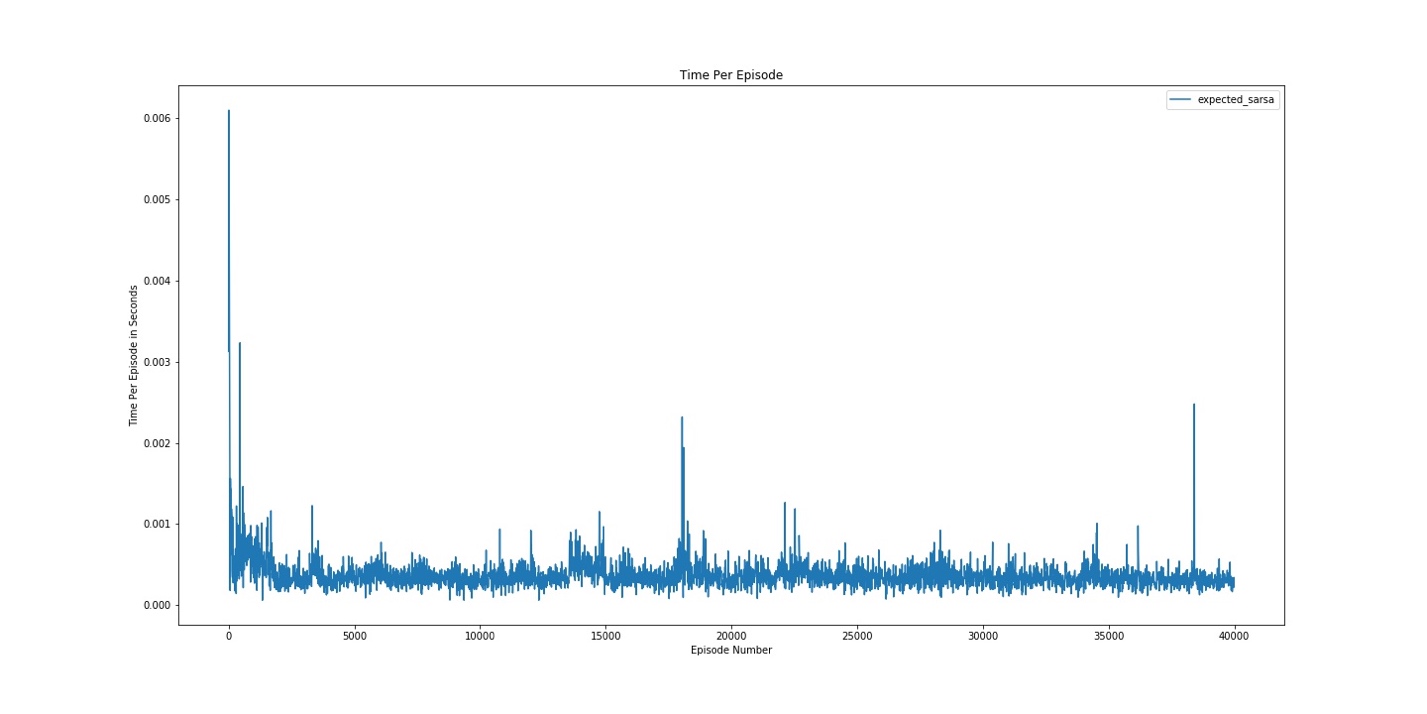


Figure 1: Expected Sarsa Time Per Episode (a = 0.2) Fixed Eps

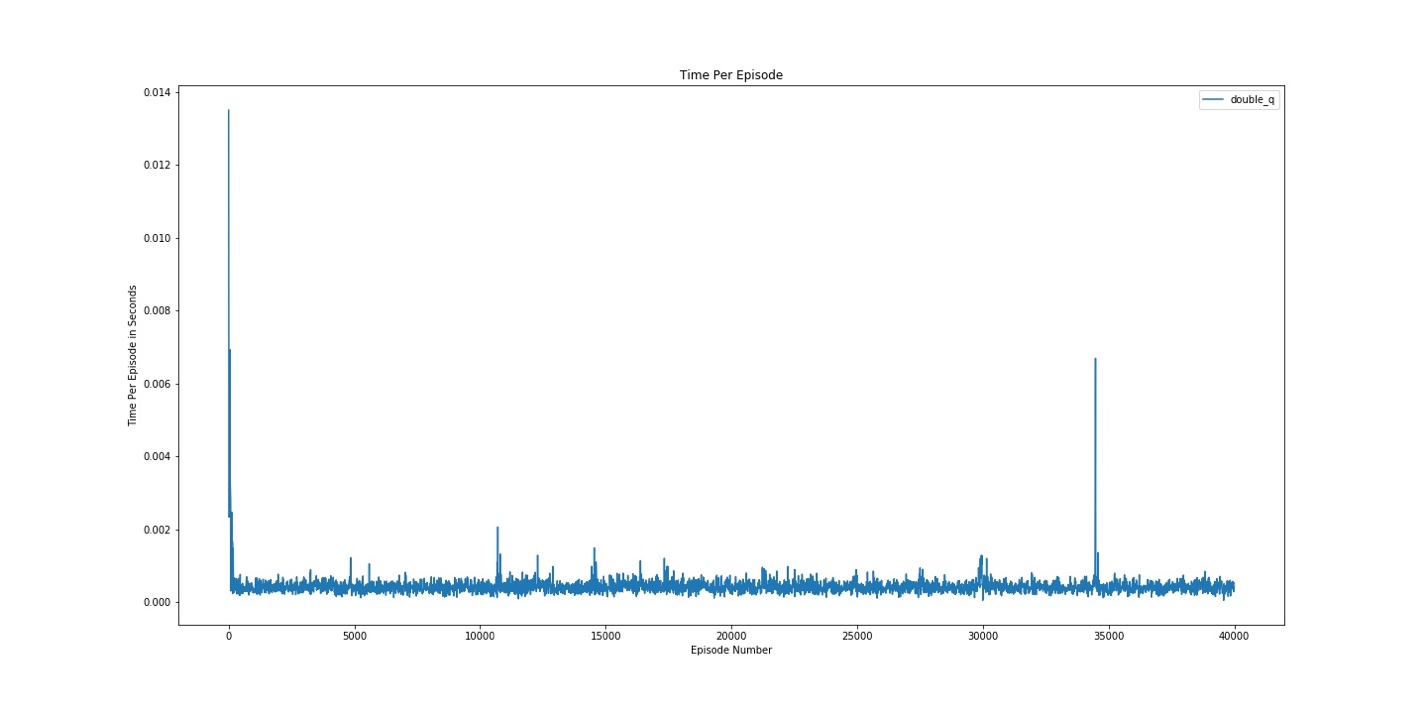


Figure 1: Double Q Time Per Episode (a = 0.2) Fixed Eps

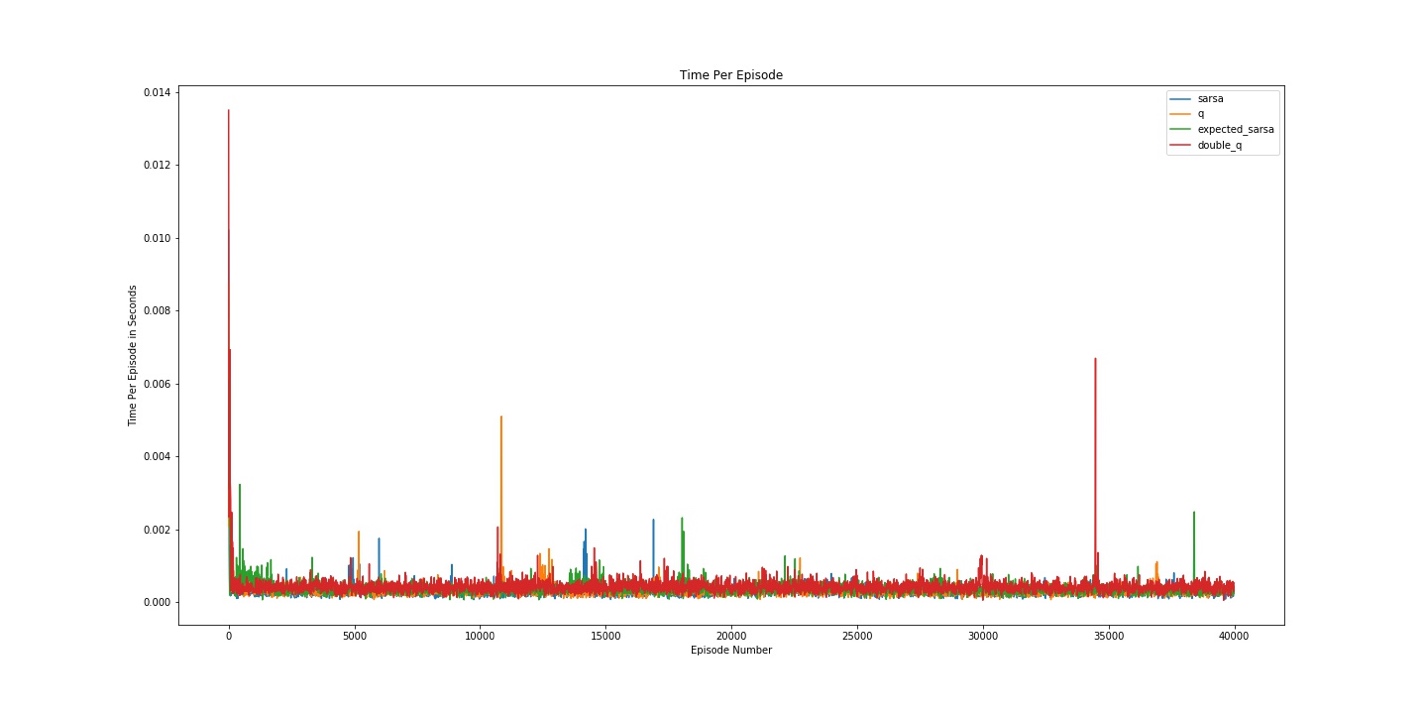


Figure 1: Combined Time Per Episode (a = 0.2) Fixed Eps

**Computation Time Per Episode Running Average**

*Alpha 0.05*

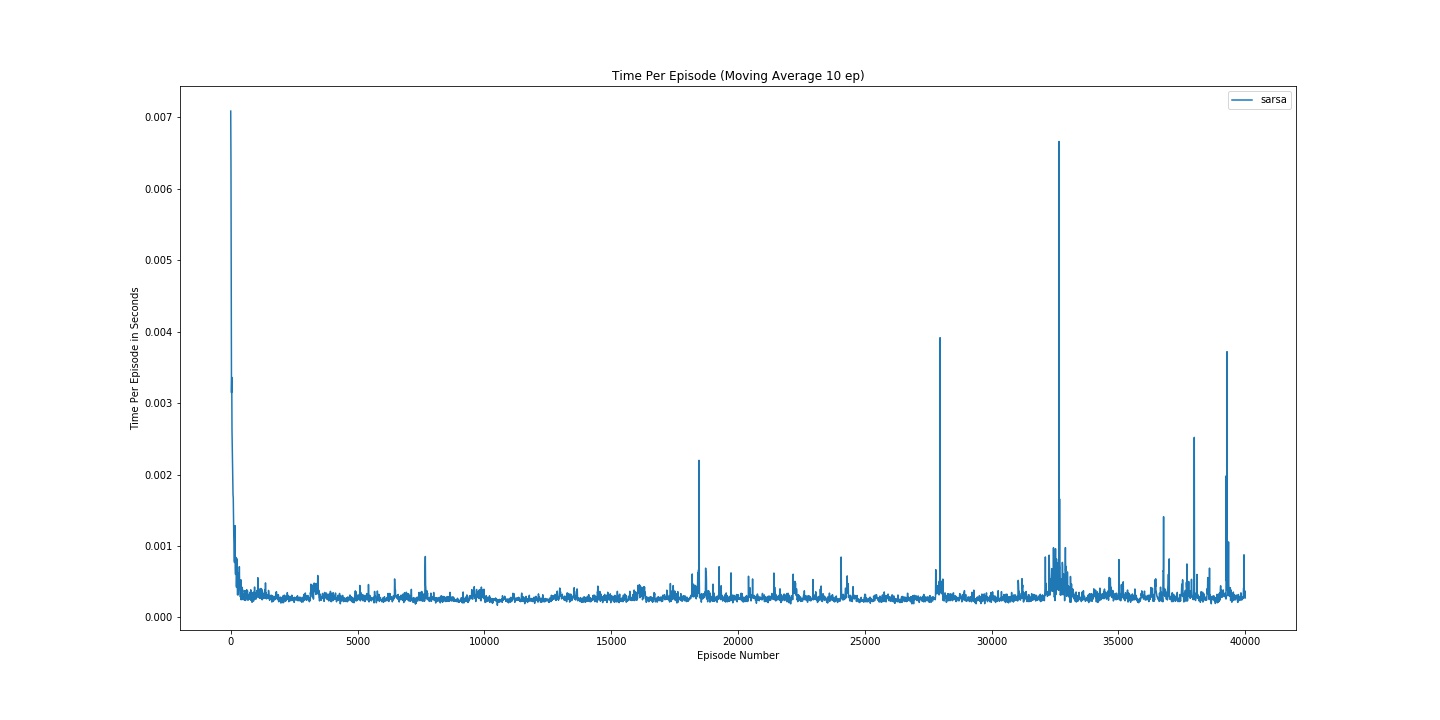


Figure 6: Sarsa Time Per Episode (a = 0.05) Fixed Eps

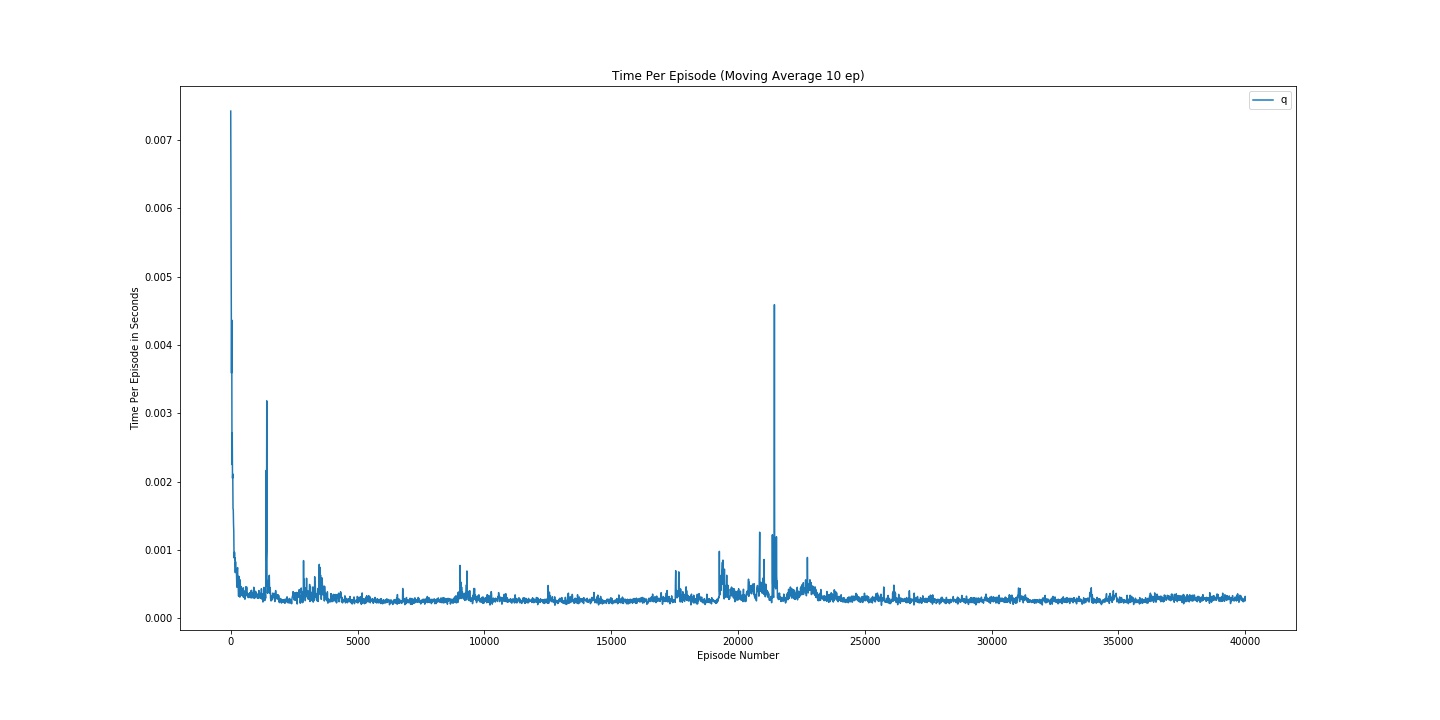


Figure 17: Q Time Per Episode (a = 0.05) Fixed Eps

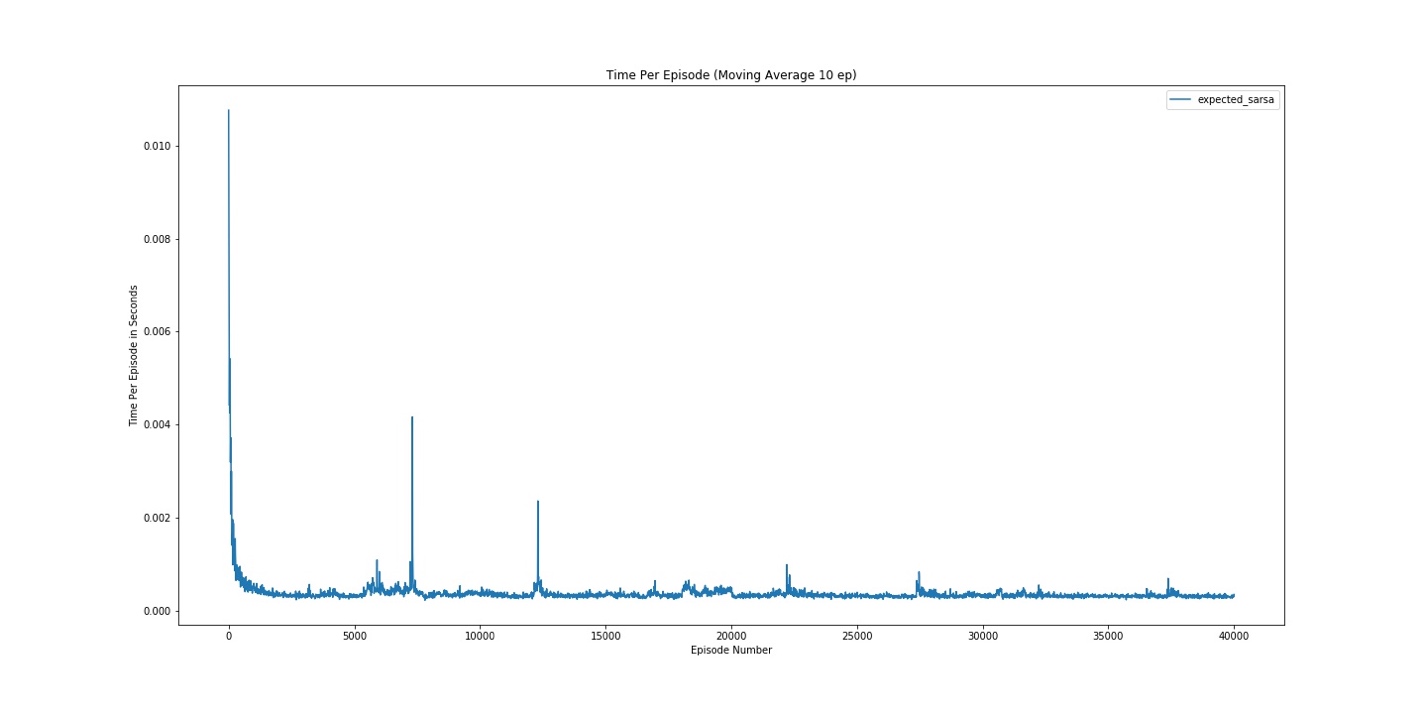


Figure 18: Expected Sarsa Time Per Episode (a = 0.05) Fixed Eps

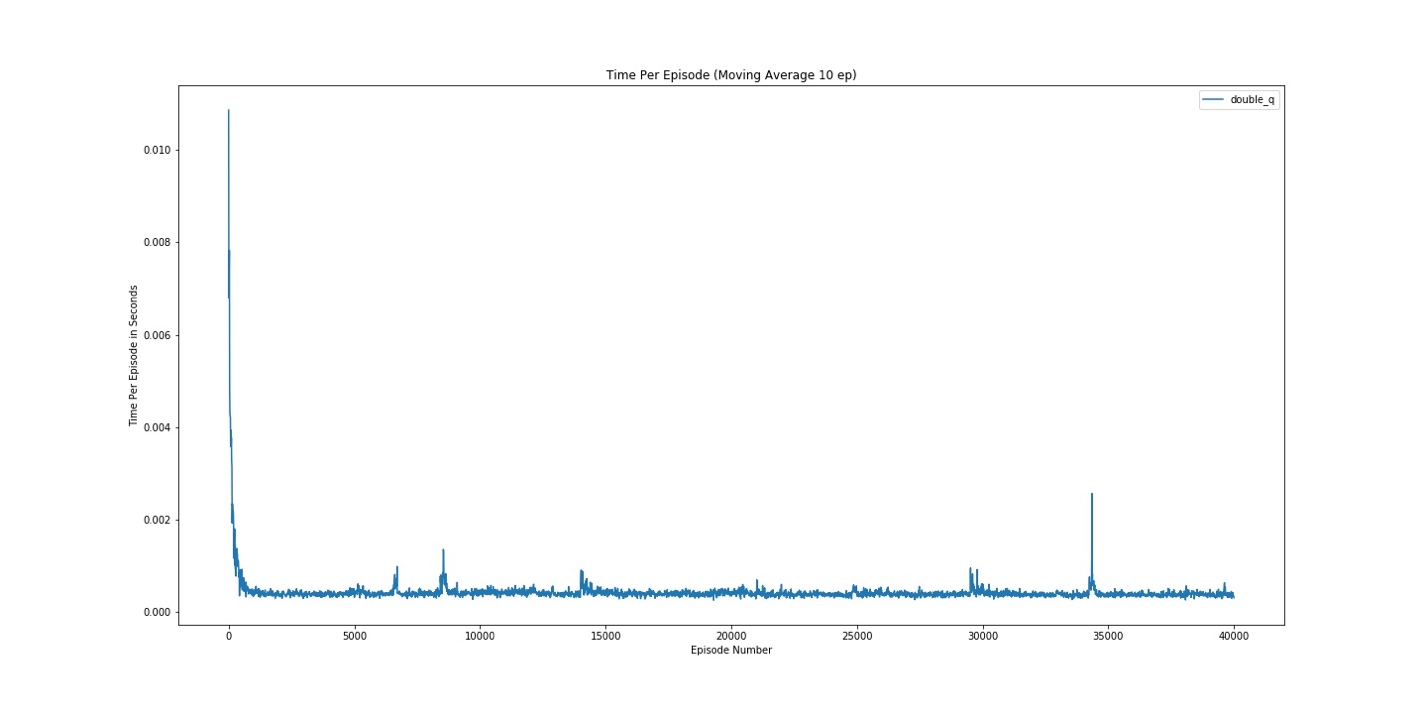


Figure 19: Double Q Time Per Episode (a = 0.05) Fixed Eps

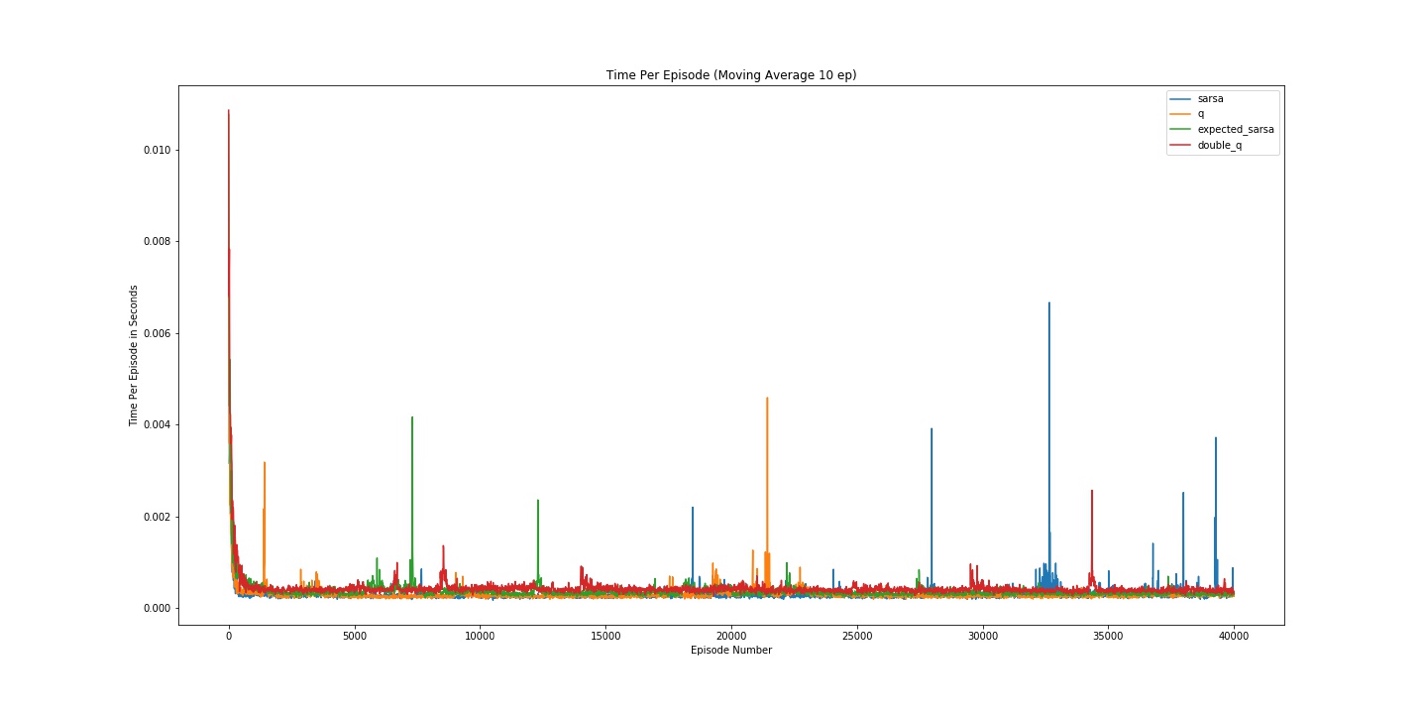


Figure 20: Combined Time Per Episode (a = 0.05) Fixed Eps

*Alpha 0.1*

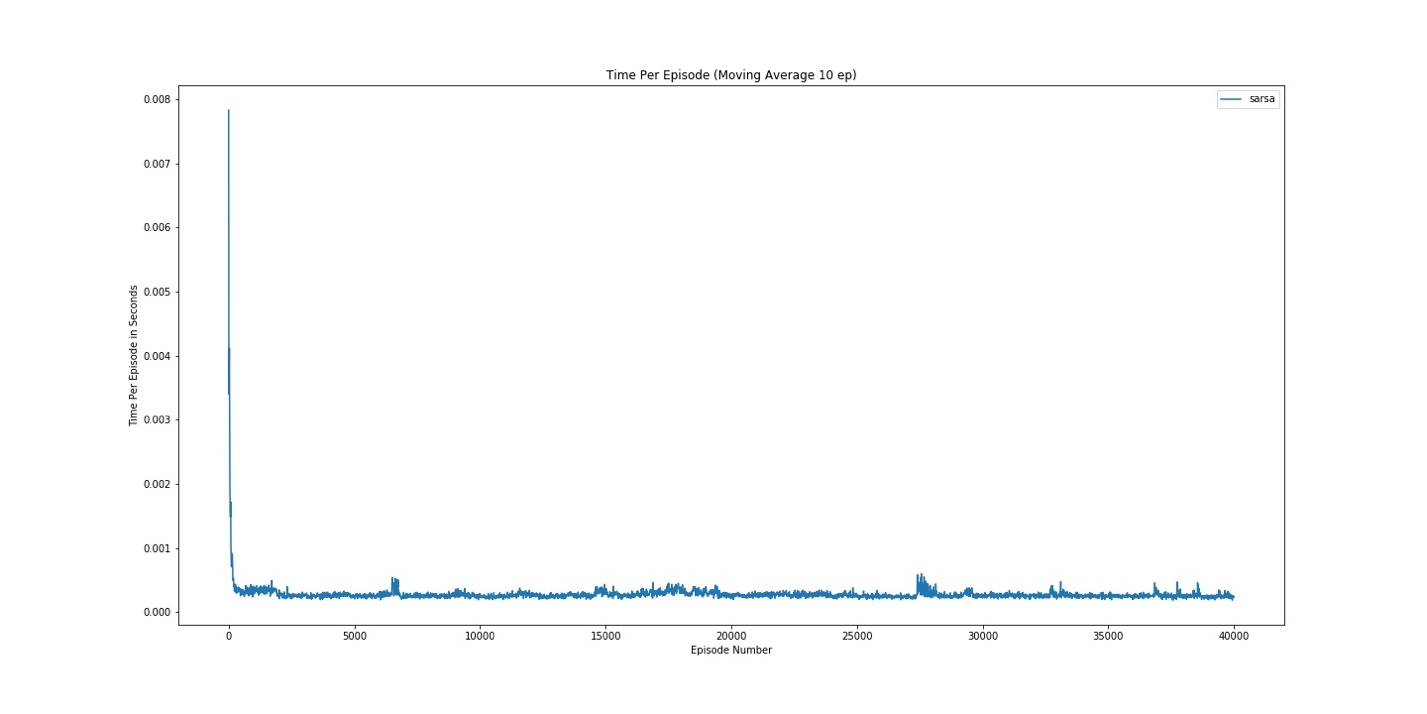


Figure 2: Sarsa Time Per Episode (a = 0.1) Fixed Eps

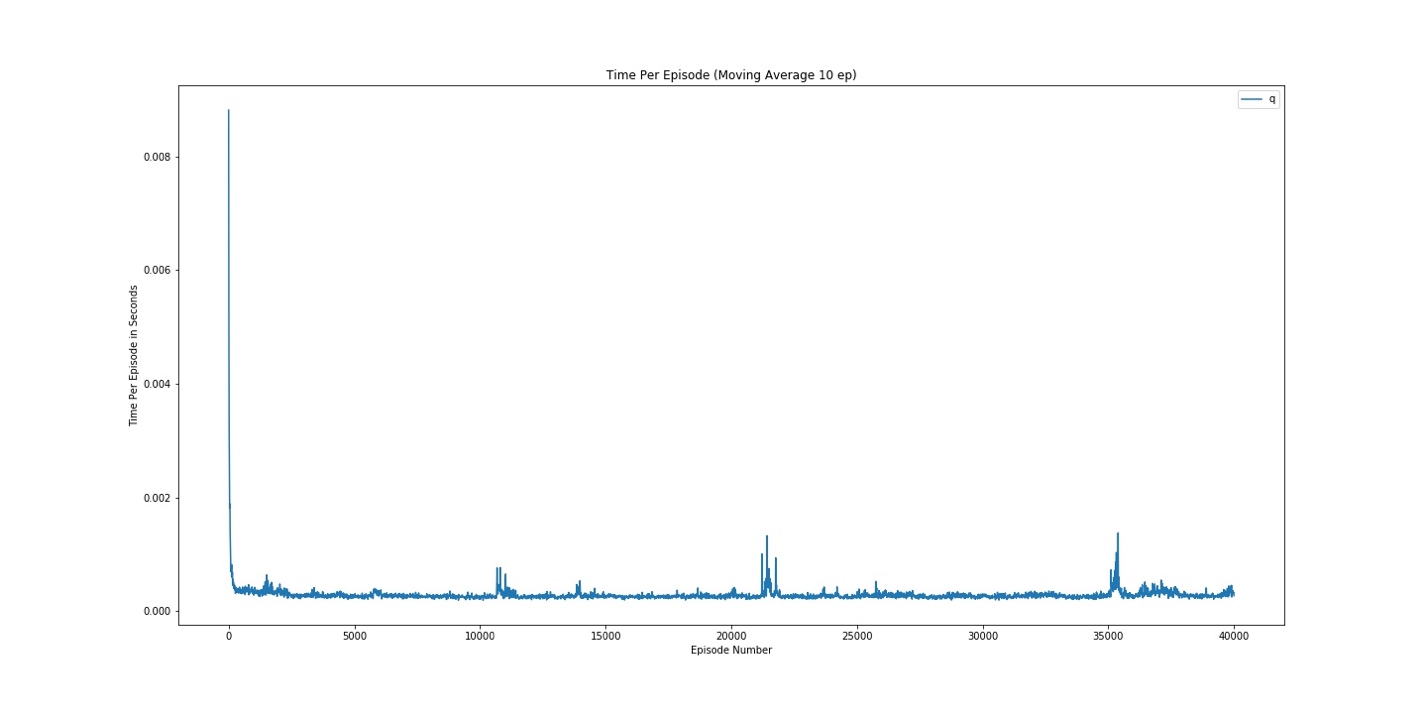


Figure 2: Q Time Per Episode (a = 0.1) Fixed Eps

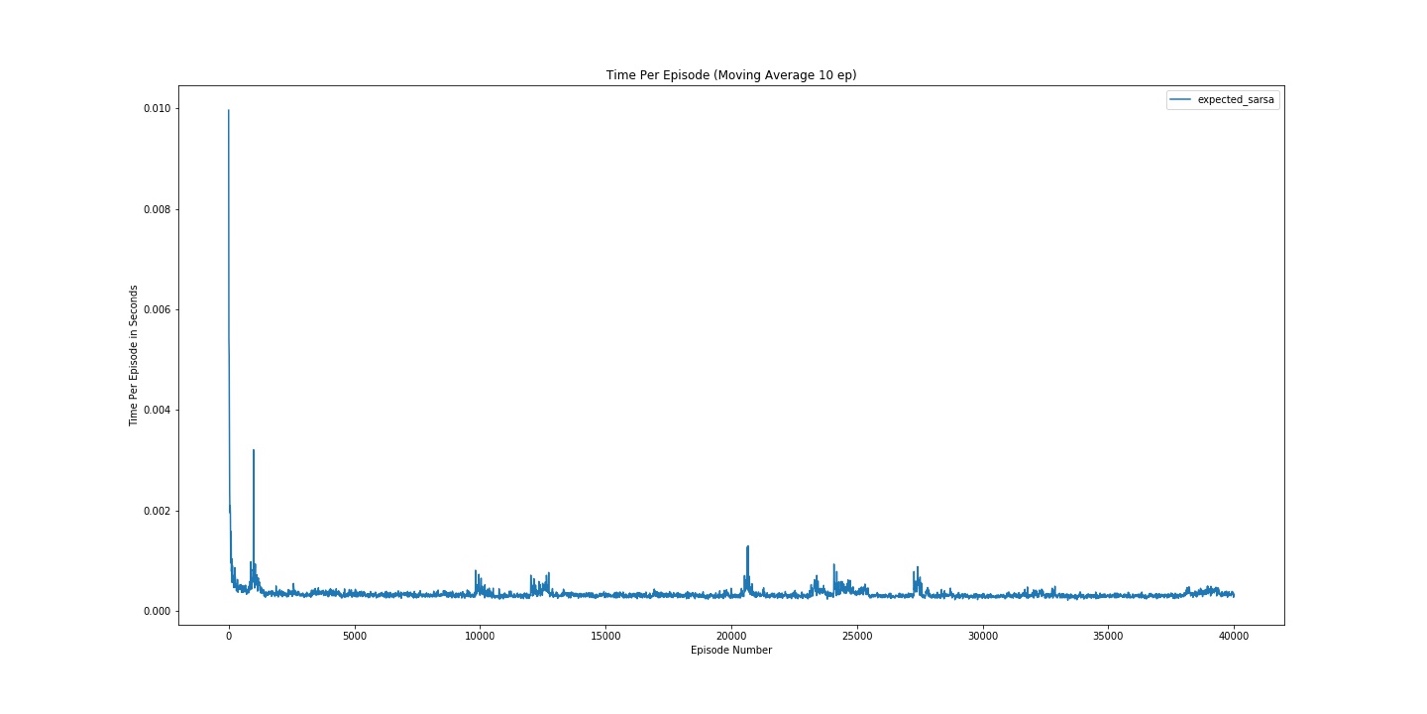


Figure 2: Expected Sarsa Time Per Episode (a = 0.1) Fixed Eps

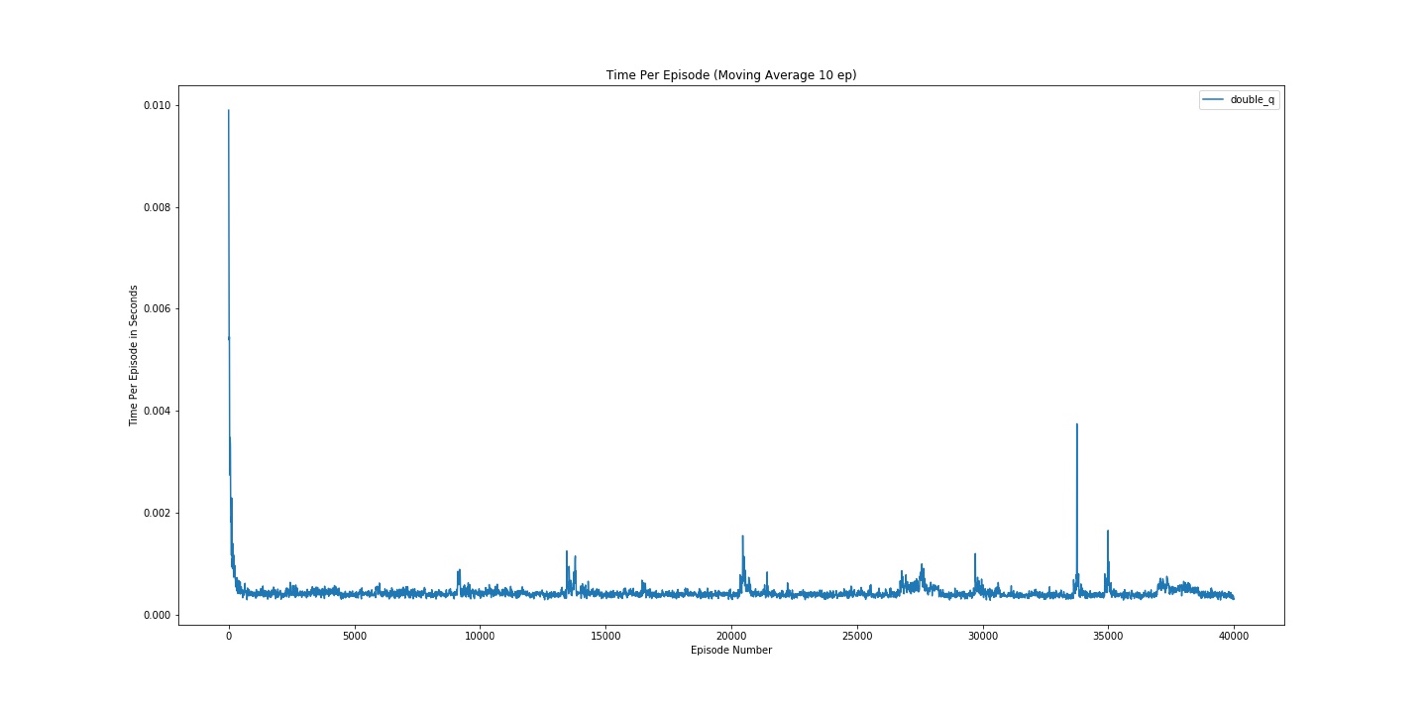


Figure 2: Double Q Time Per Episode (a = 0.1) Fixed Eps

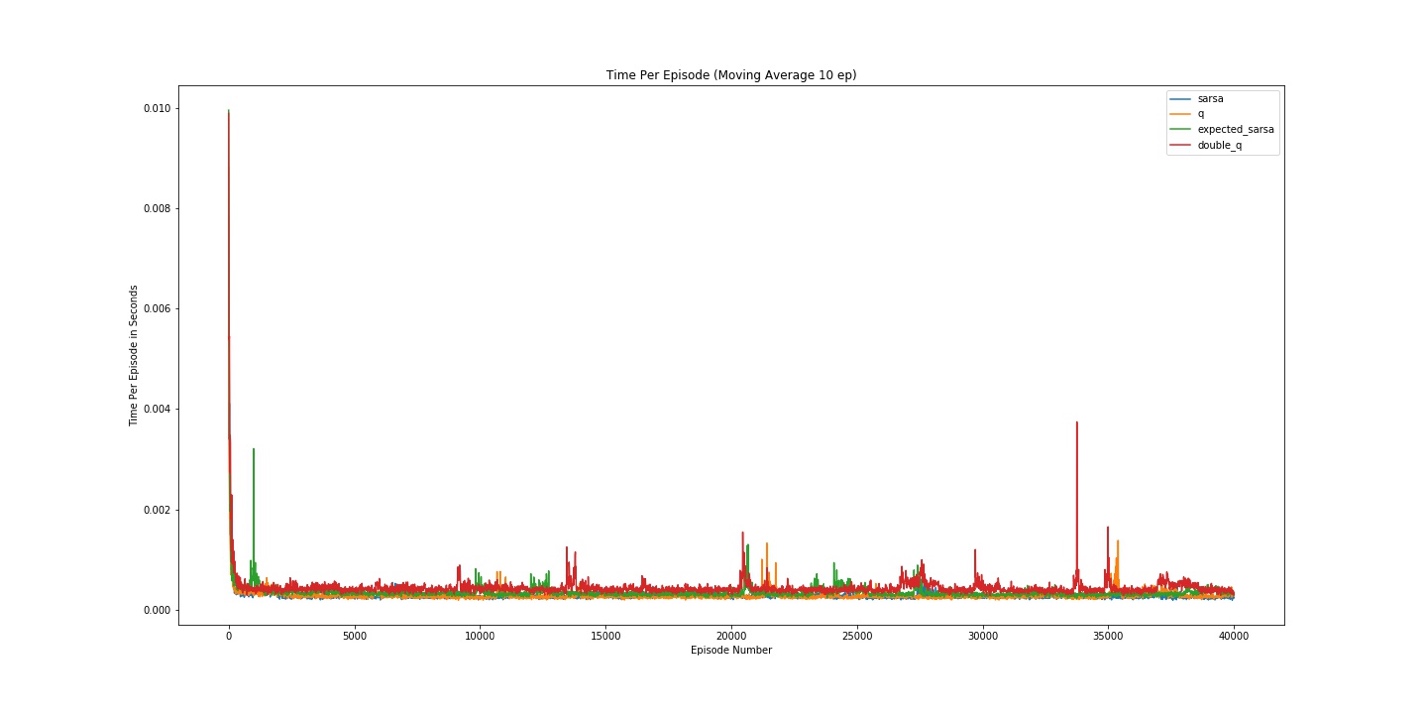


Figure 2: Combined Time Per Episode (a = 0.1) Fixed Eps

*Alpha 0.2*

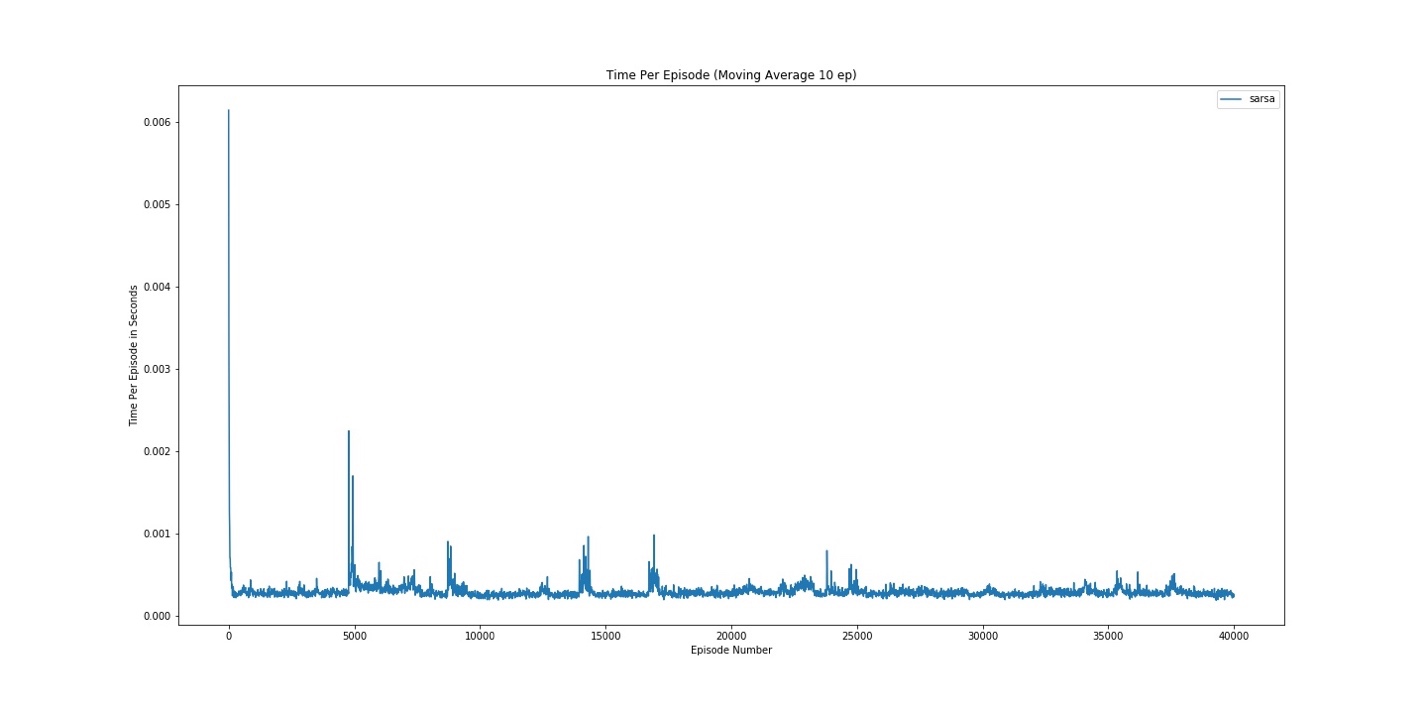


Figure 26: Sarsa Time Per Episode (a = 0.2) Fixed Eps

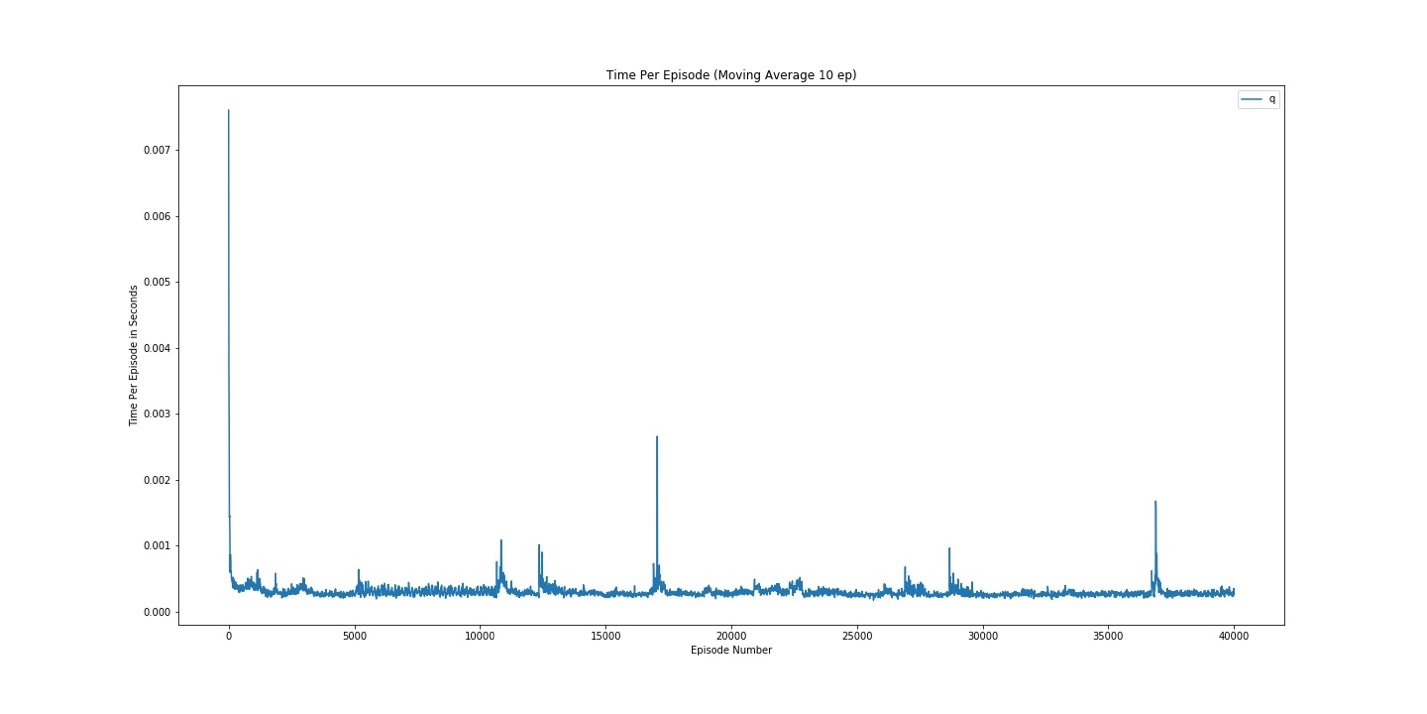


Figure 7: Q Time Per Episode (a = 0.2) Fixed Eps

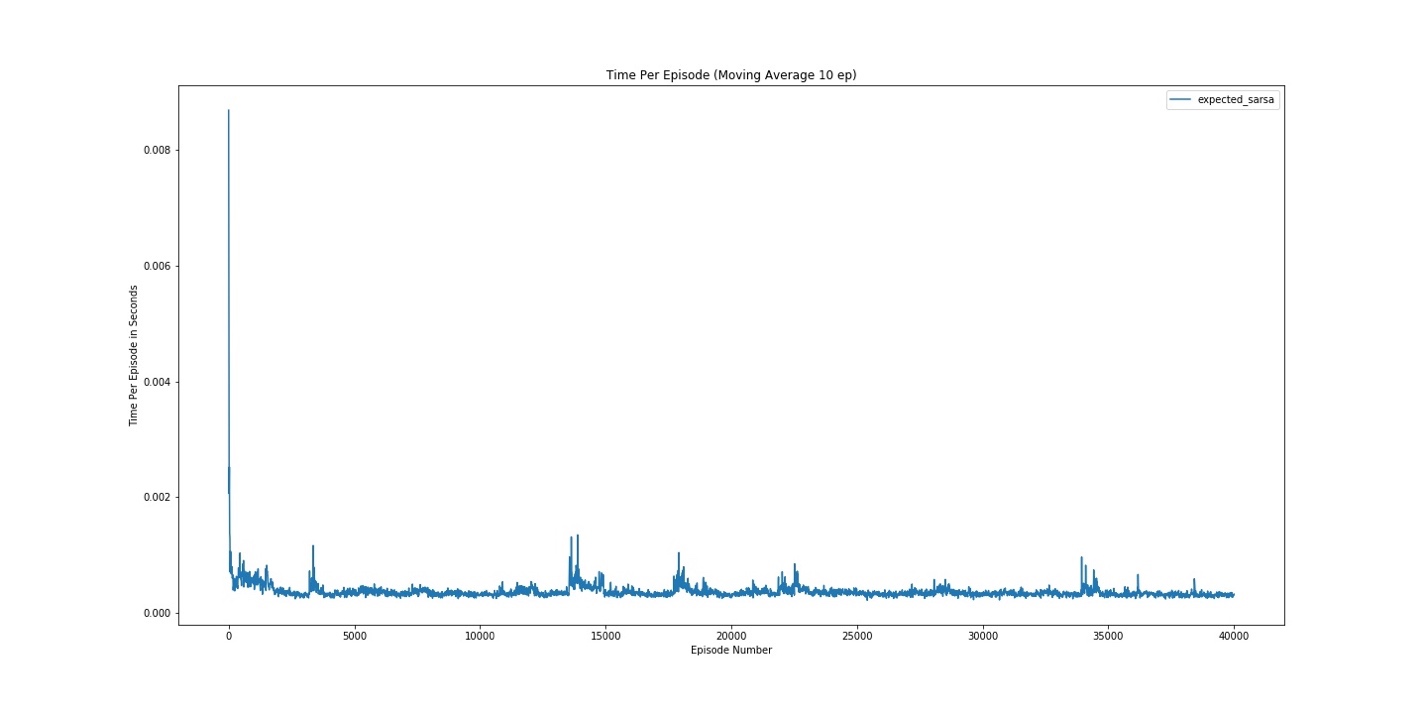


Figure 28: Expected Sarsa Time Per Episode (a = 0.2) Fixed Eps

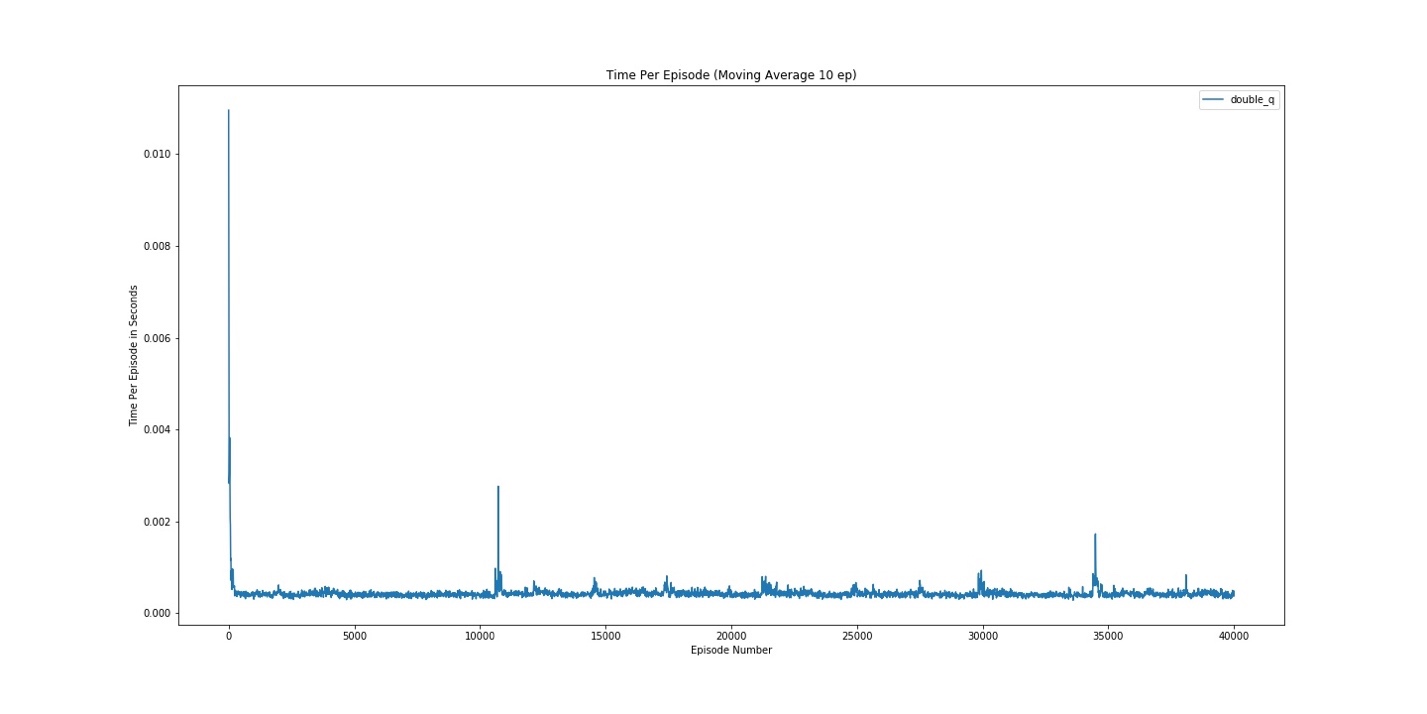


Figure 29: Double Q Time Per Episode (a = 0.2) Fixed Eps

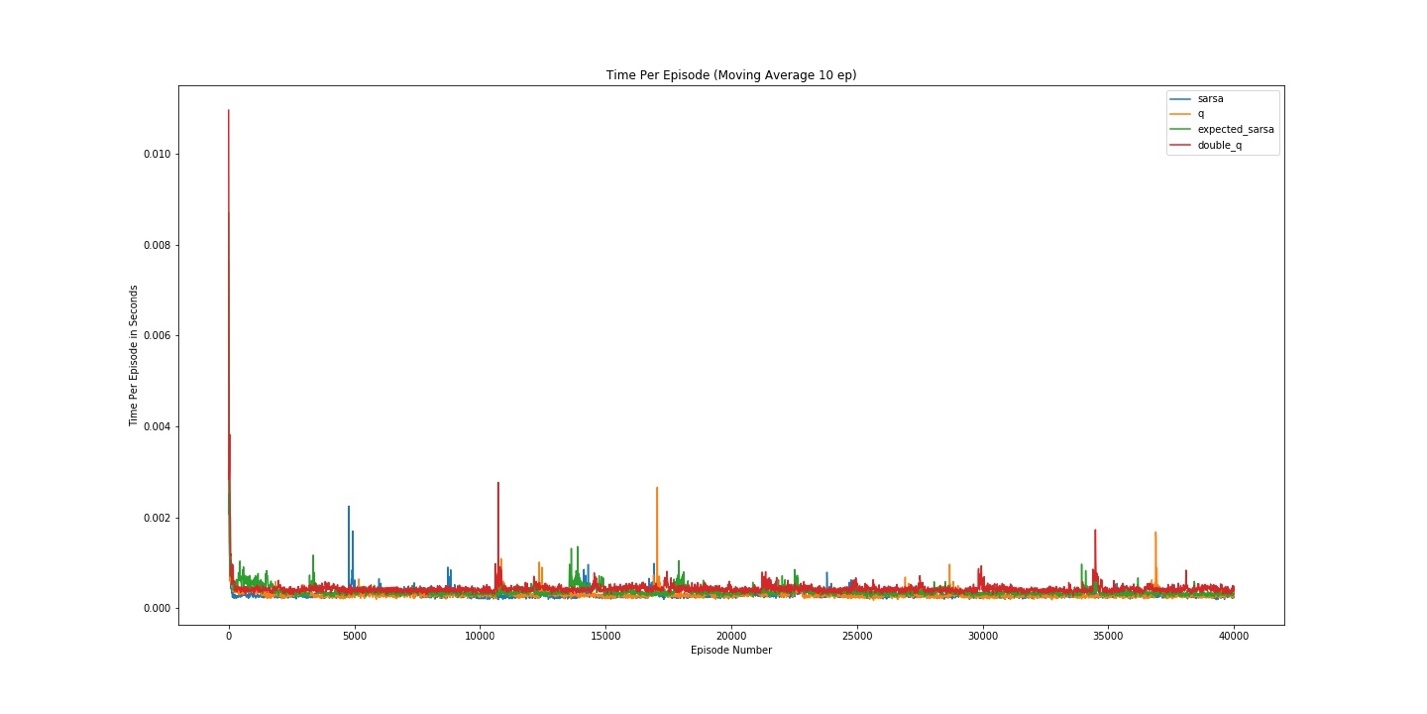


Figure 30: Combined Time Per Episode (a = 0.2) Fixed Eps

**Number of Steps per Episode Running Average**

*Alpha 0.05*

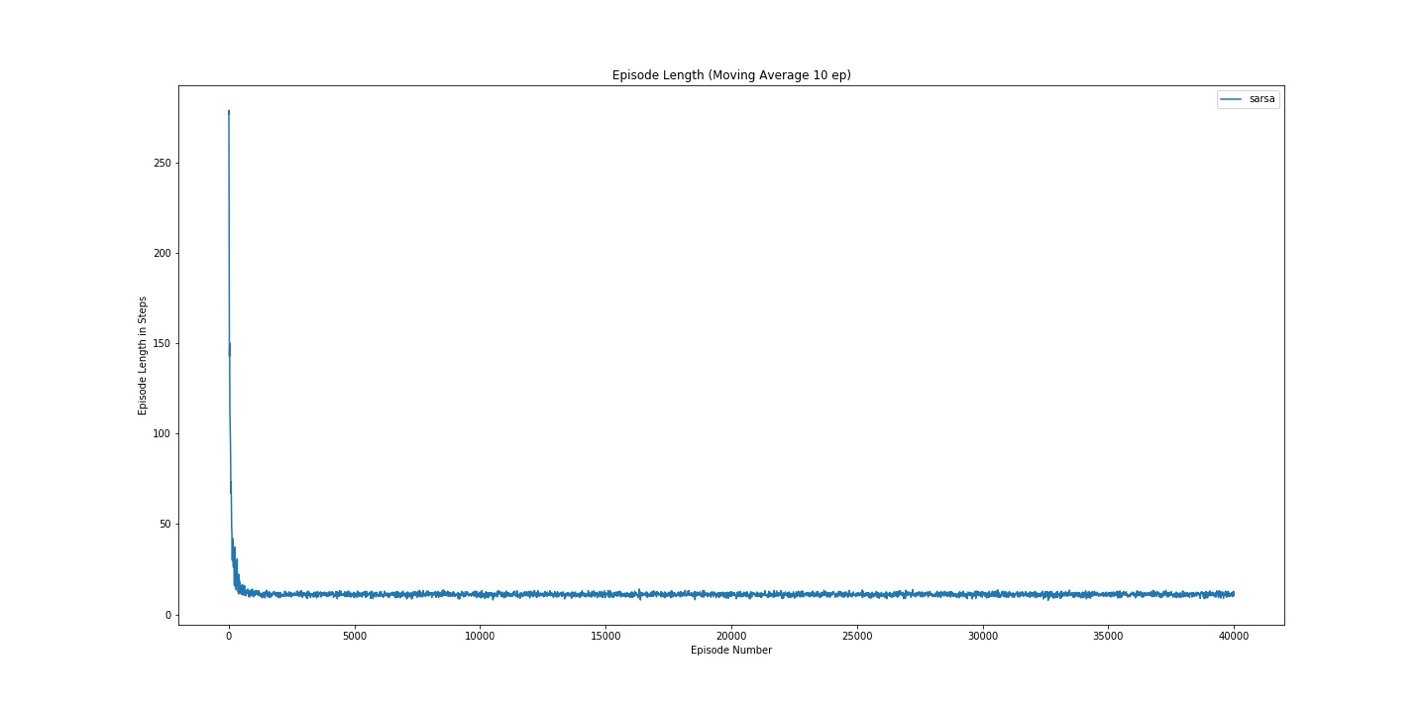


Figure 31: Sarsa Num Episodes (a = 0.05) Fixed Eps

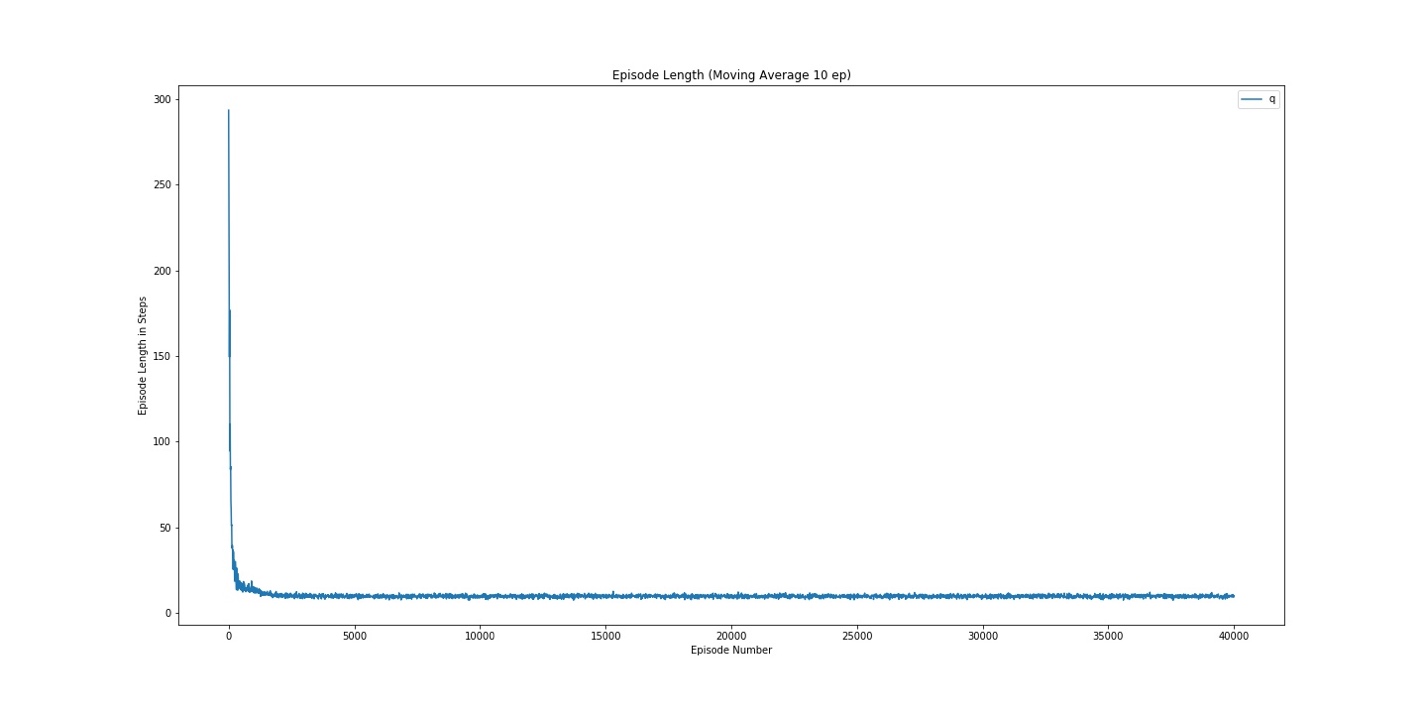


Figure 3: Q Num Episodes (a = 0.05) Fixed Eps

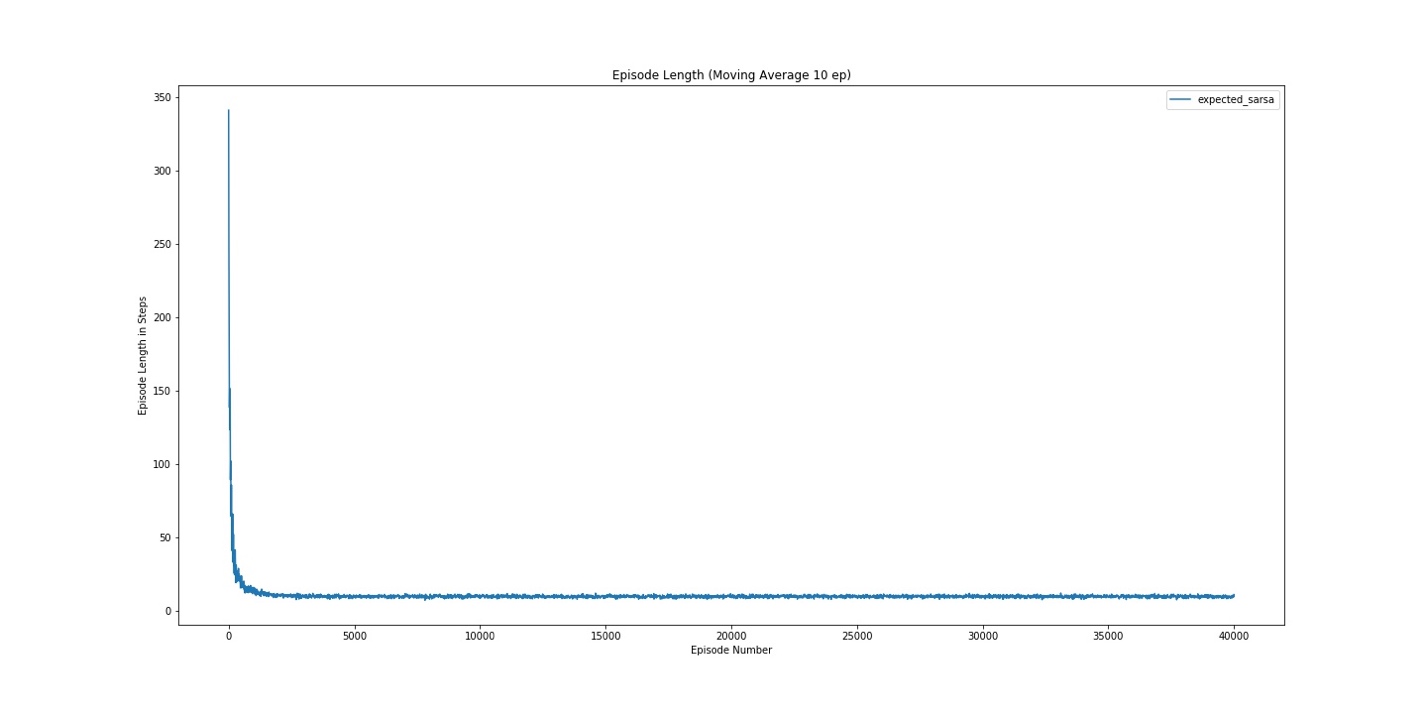


Figure 3: Expected Sarsa Num Episodes (a = 0.05) Fixed Eps

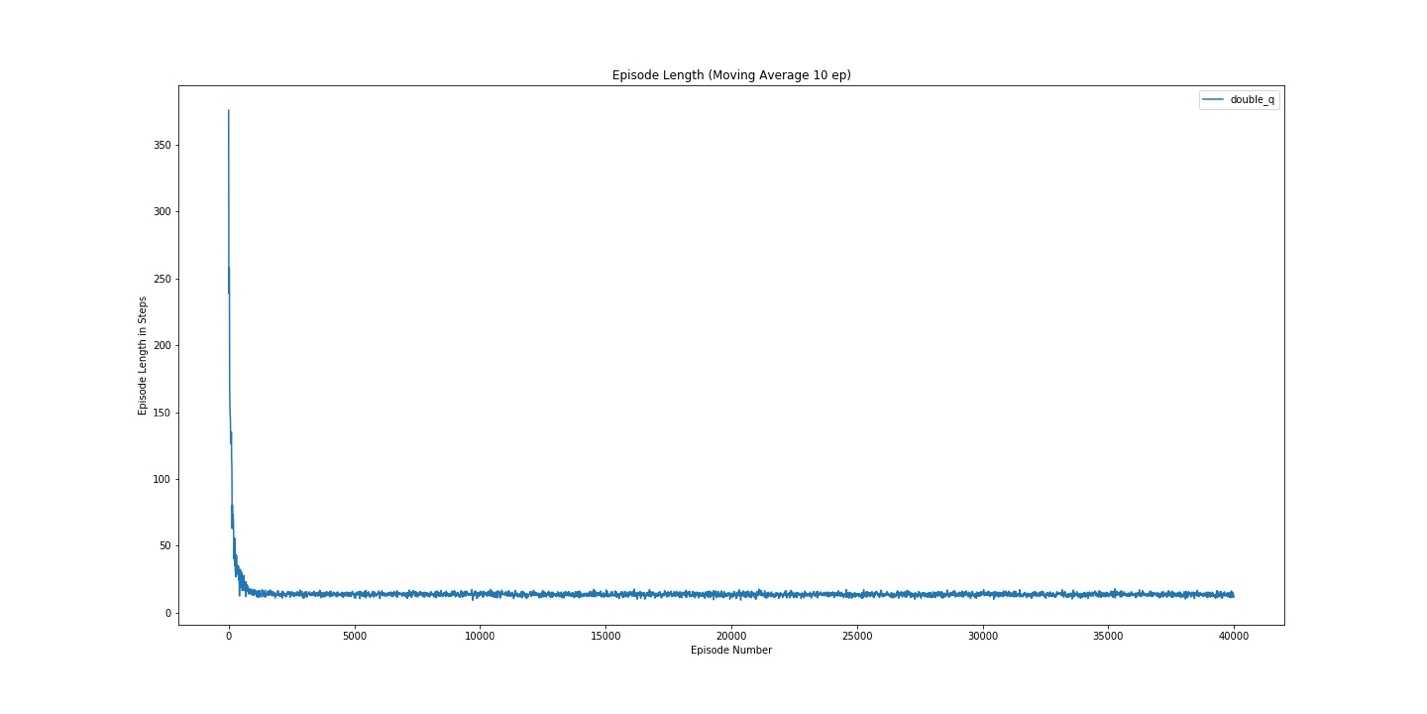


Figure 3: Double Q Num Episodes (a = 0.05) Fixed Eps

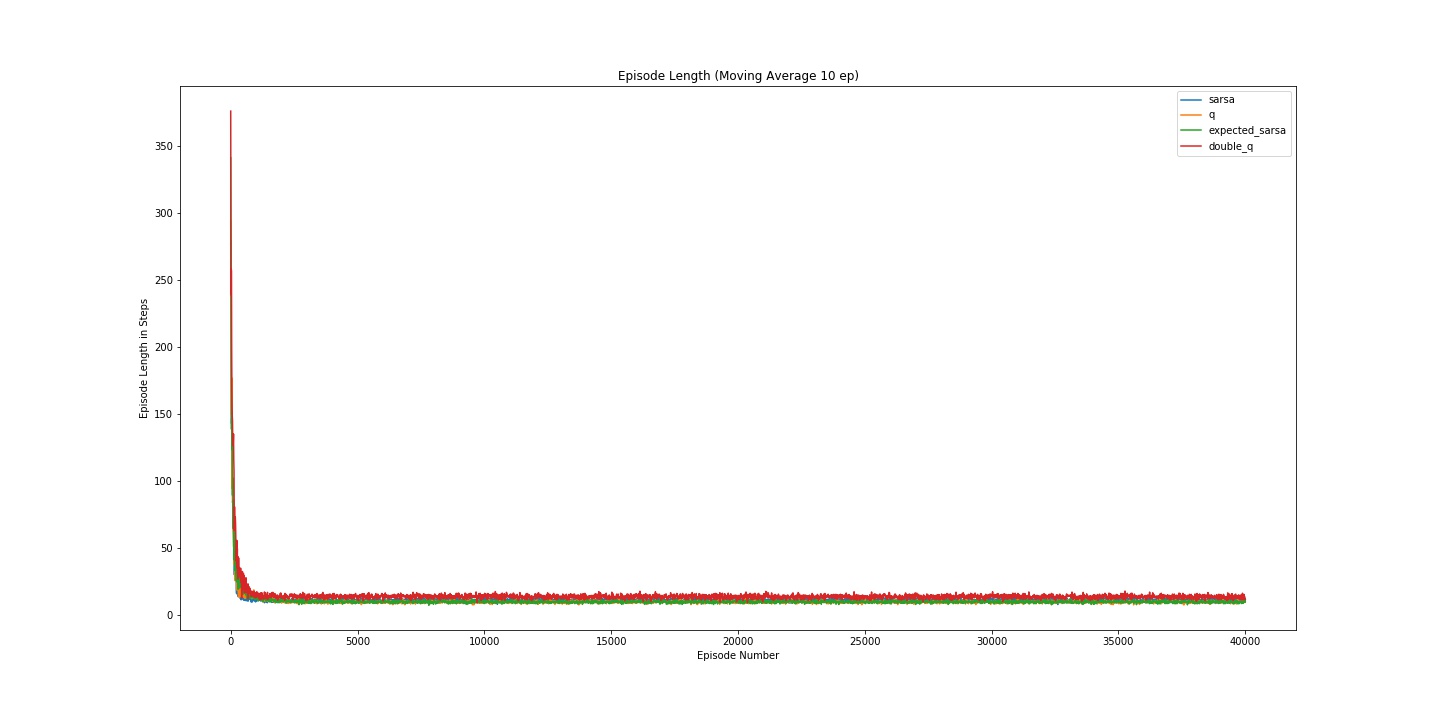


Figure 3: Combined Num Episodes (a = 0.05) Fixed Eps

*Alpha 0.1*

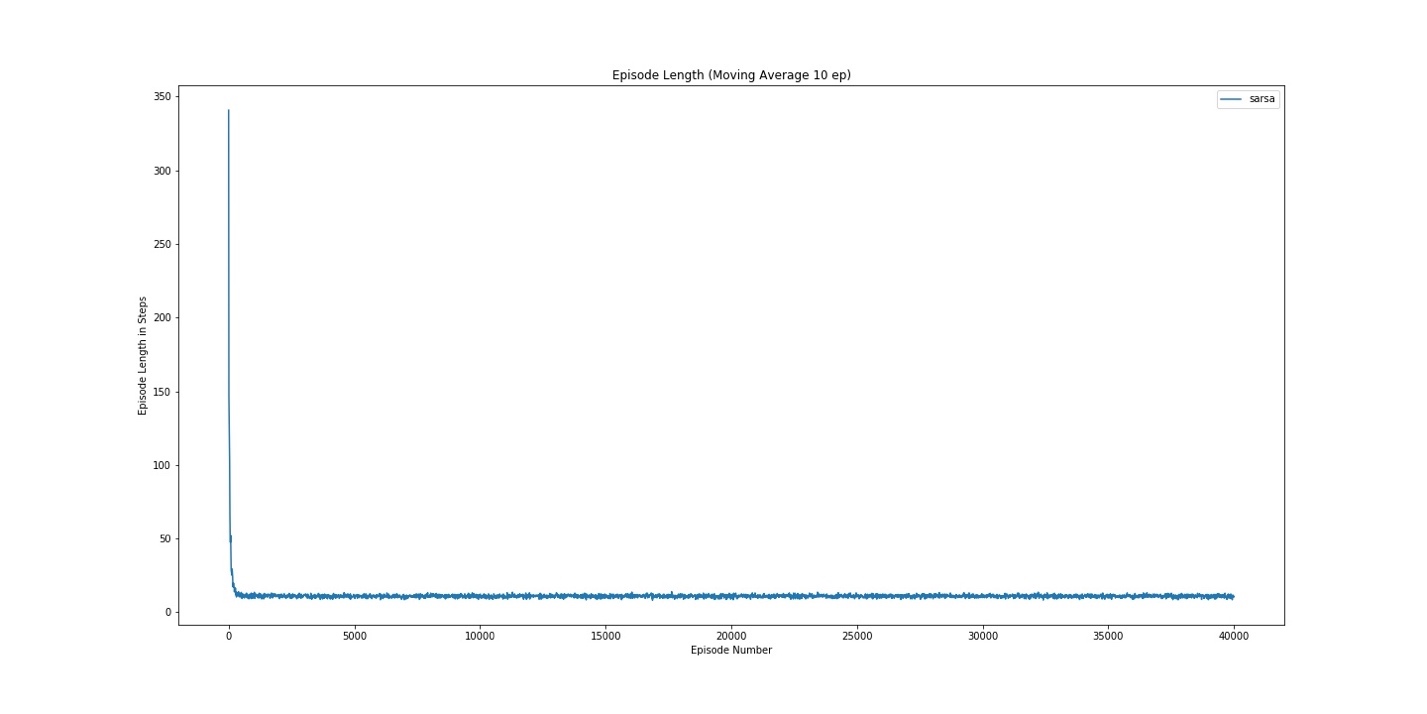


Figure 36: Sarsa Num Episodes (a = 0.1) Fixed Eps

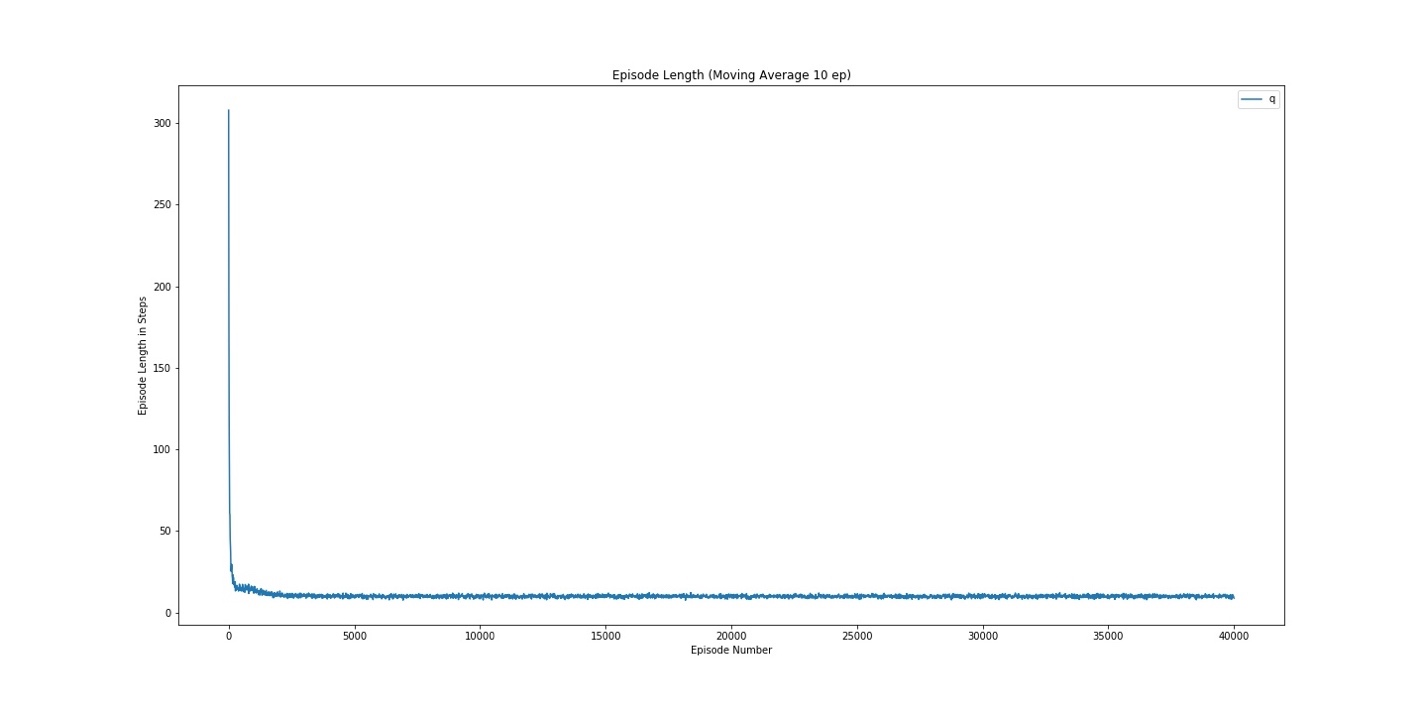


Figure 37: Q Num Episodes (a = 0.1) Fixed Eps

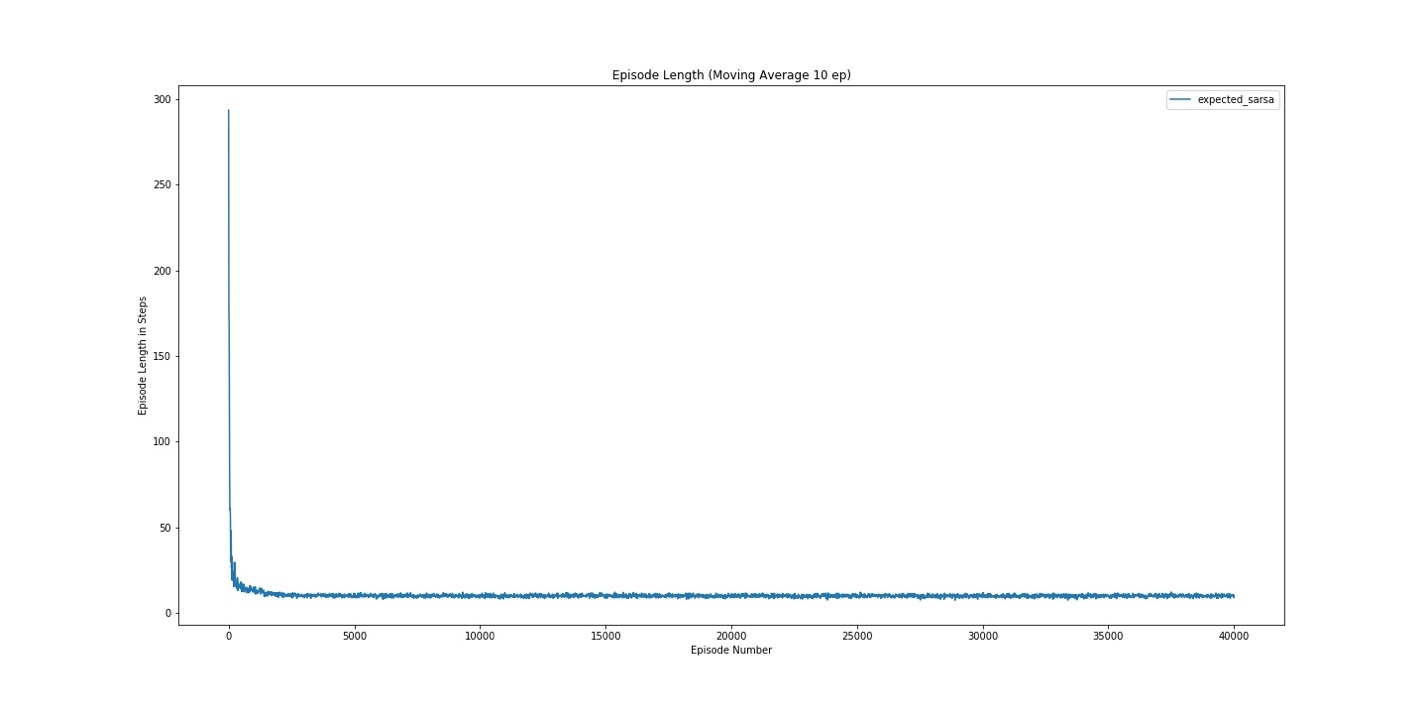


Figure 8: Expected Sarsa Num Episodes (a = 0.1) Fixed Eps

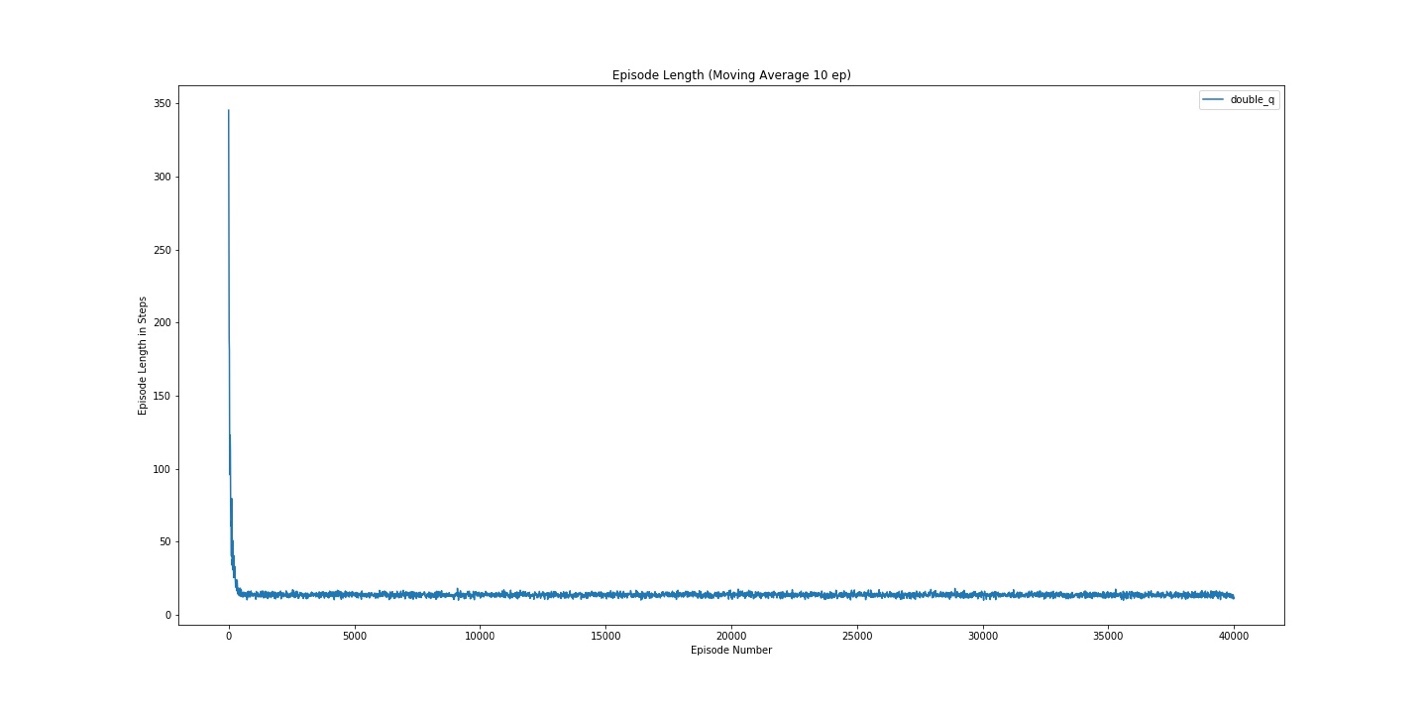


Figure 39: Double Q Num Episodes (a = 0.1) Fixed Eps

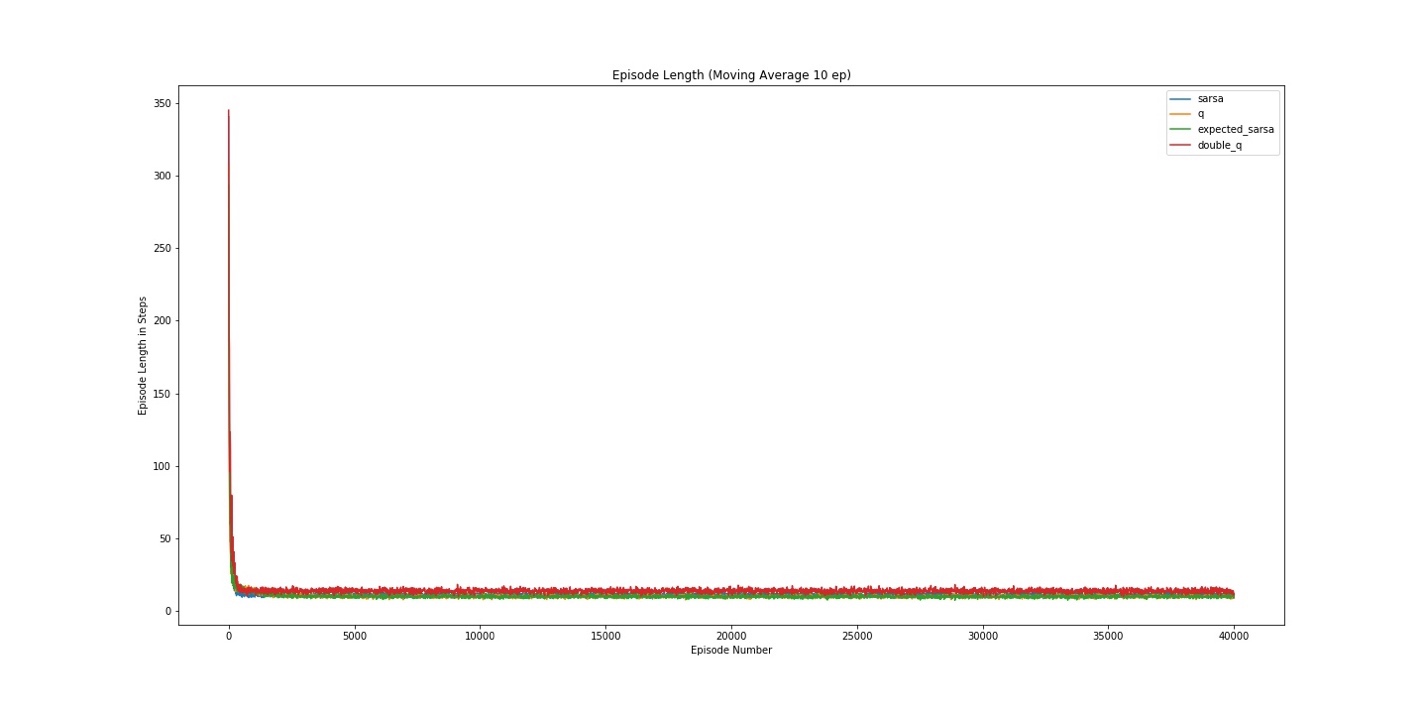


Figure 40: Combined Num Episodes(a = 0.1) Fixed Eps

*Alpha 0.2*

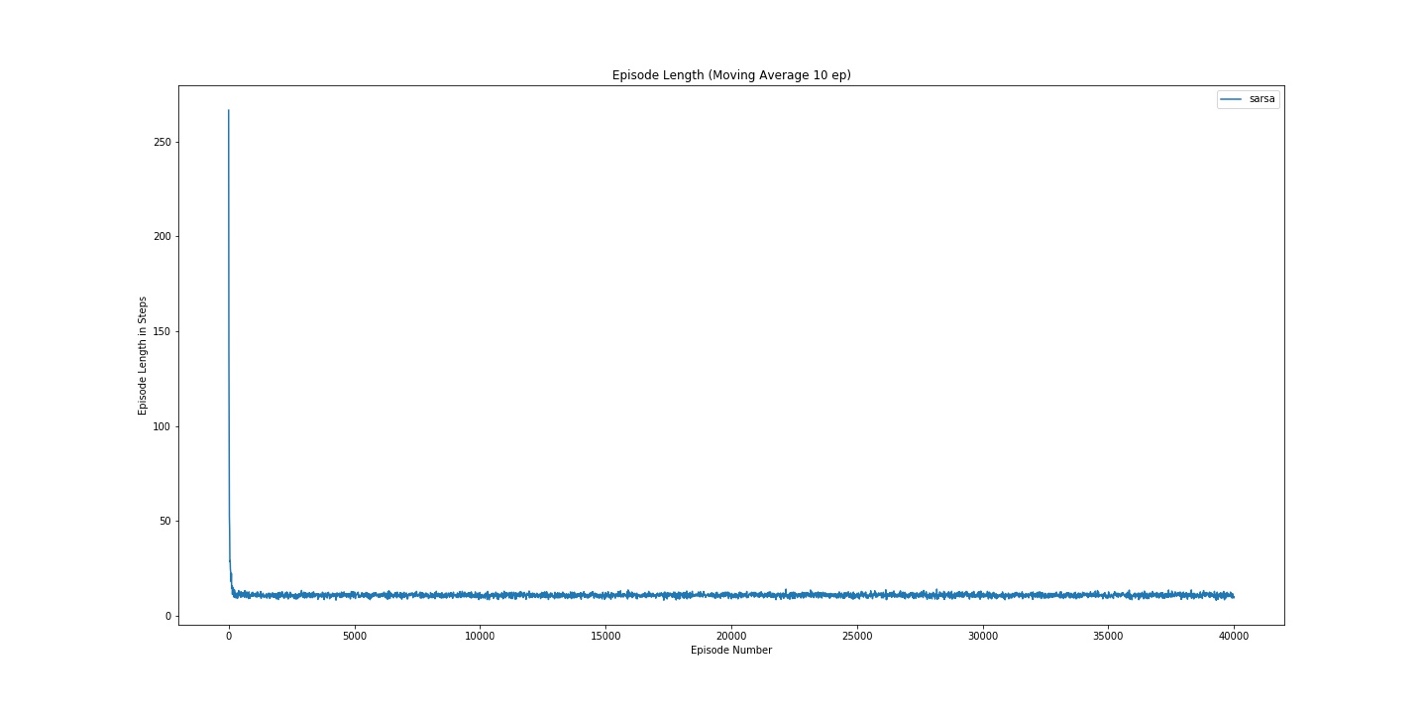


Figure 4: Sarsa Num Episodes (a = 0.2) Fixed Eps

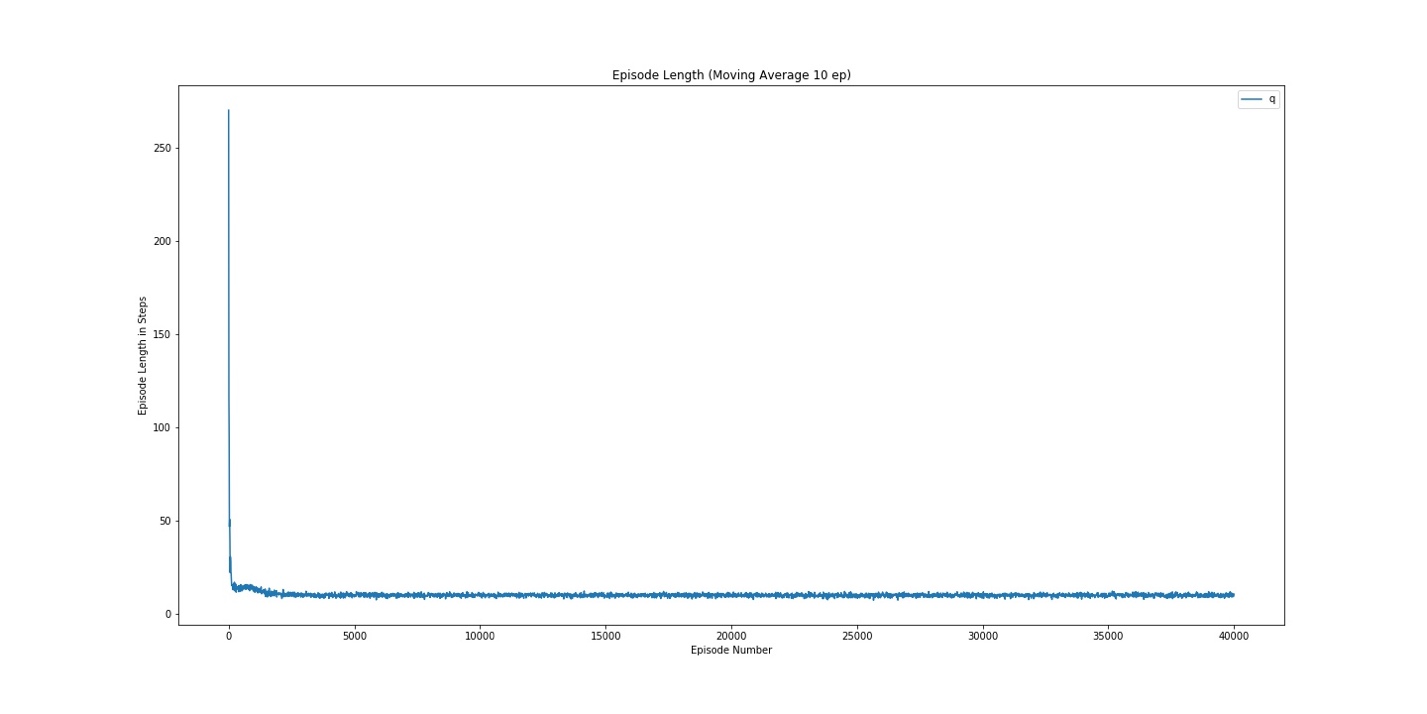


Figure 4: Q Num Episodes(a = 0.2) Fixed Eps

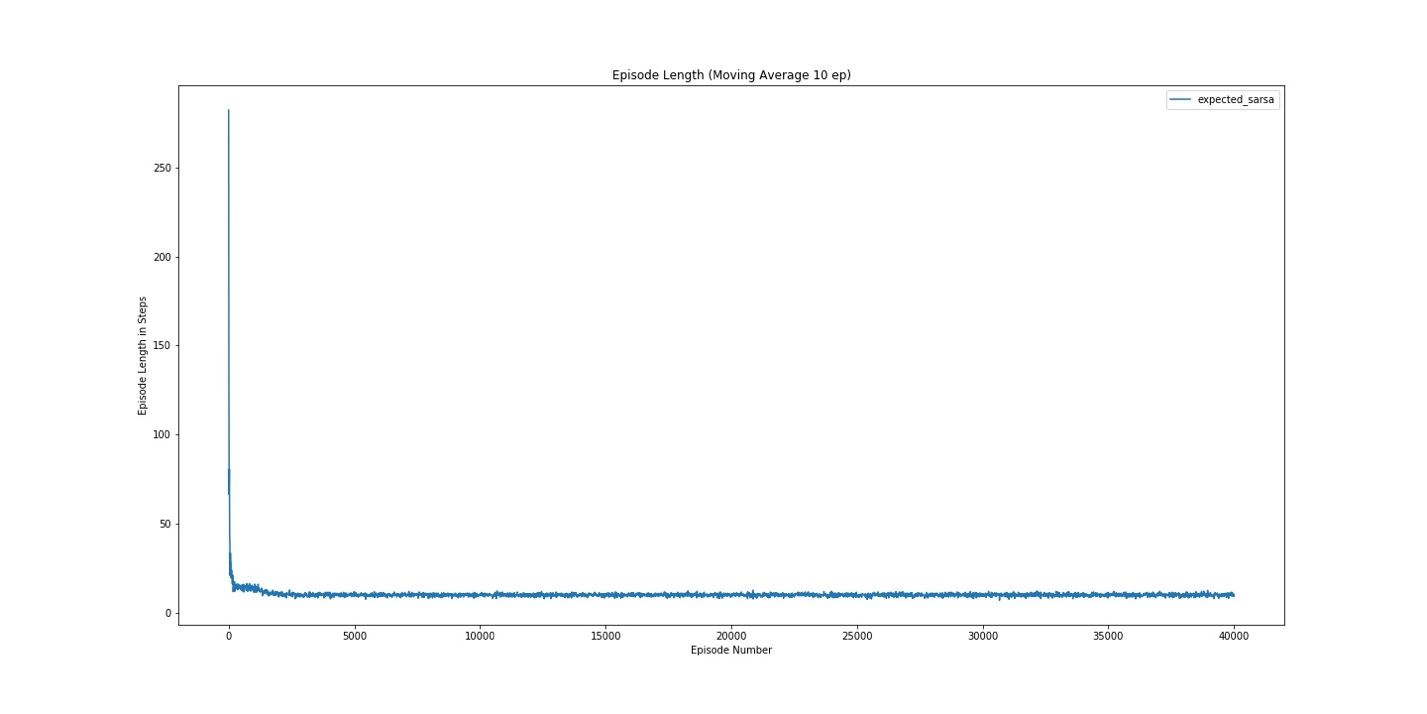


Figure 4: Expected Sarsa Num Episodes(a = 0.2) Fixed Eps

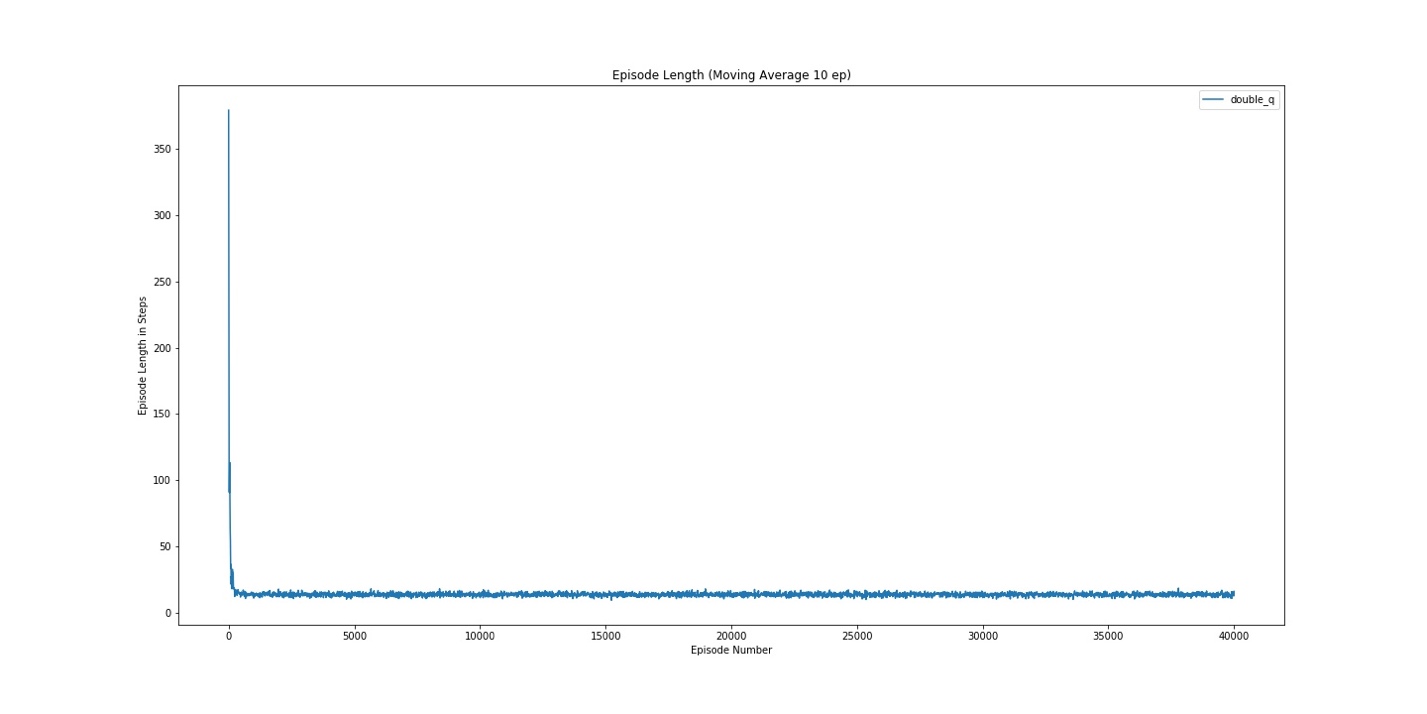


Figure 4: Double Q Num Episodes (a = 0.2) Fixed Eps

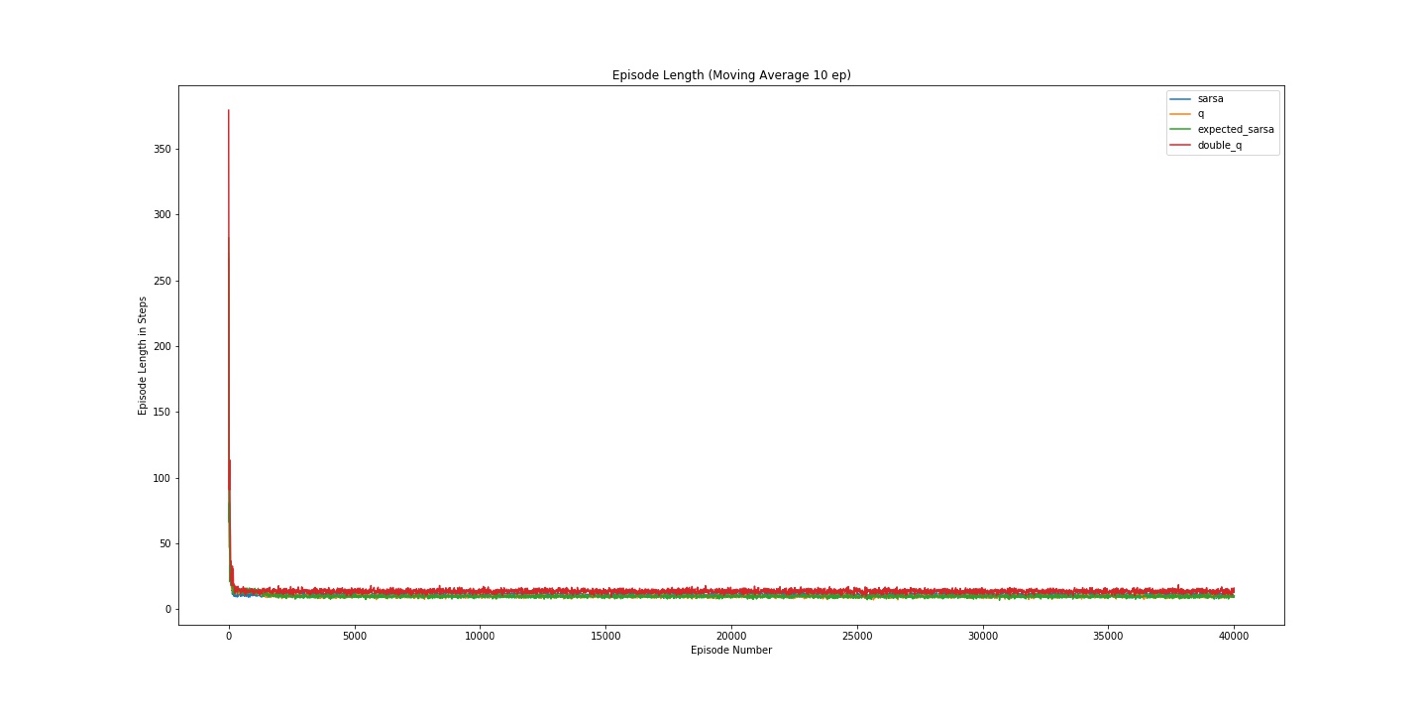


Figure 4: Combined Num Episodes (a = 0.2) Fixed Eps

**Variable Epsilon Results**

**Computation Time Per Episode Running Average**

*Alpha 0.05*

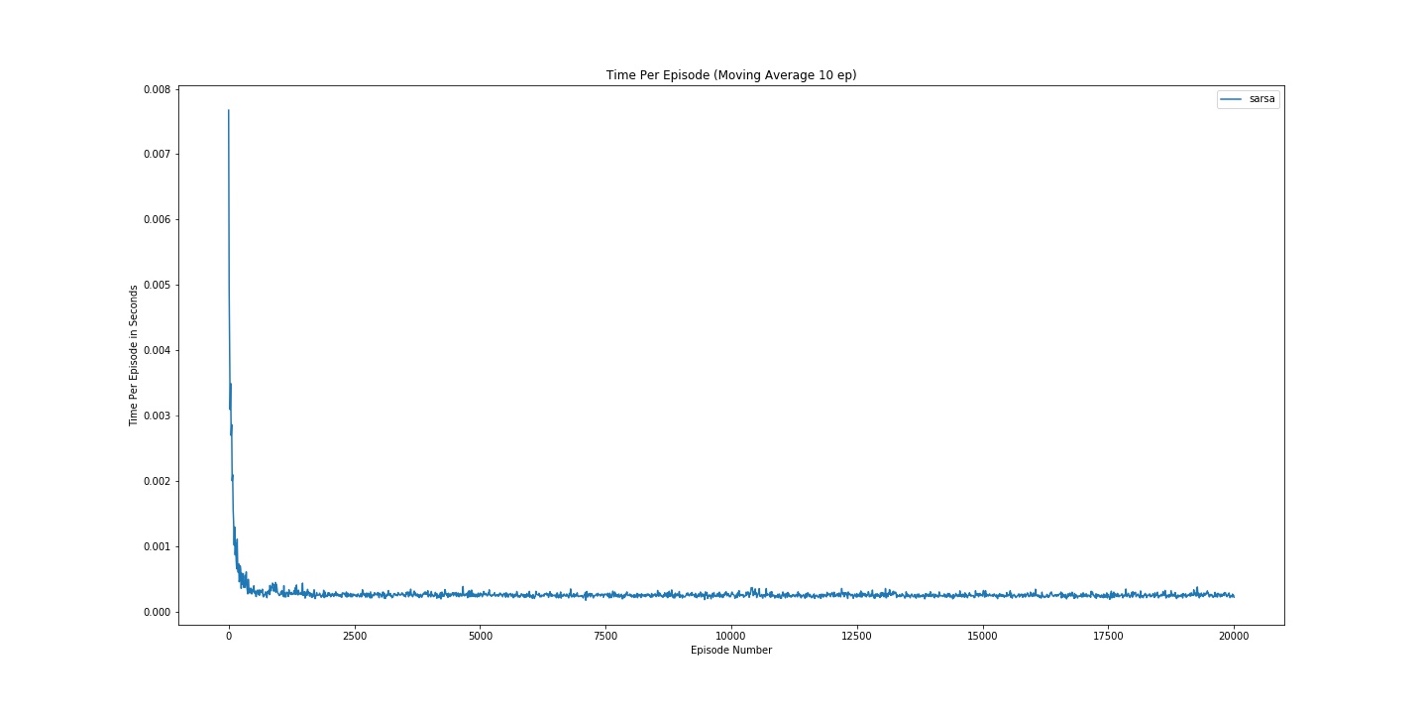


Figure 46: Sarsa Time Per Episode (a = 0.05) Variable Eps

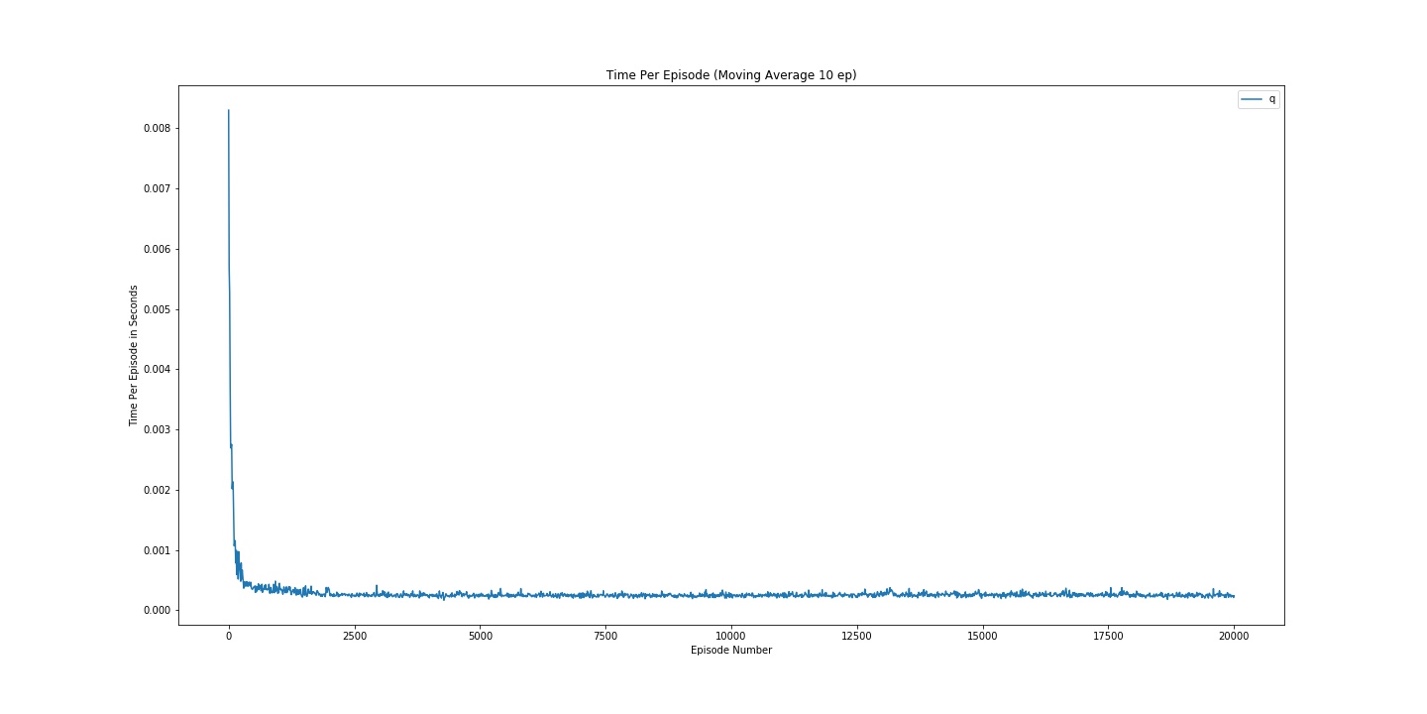


Figure 47: Q Time Per Episode (a = 0.05) Variable Eps

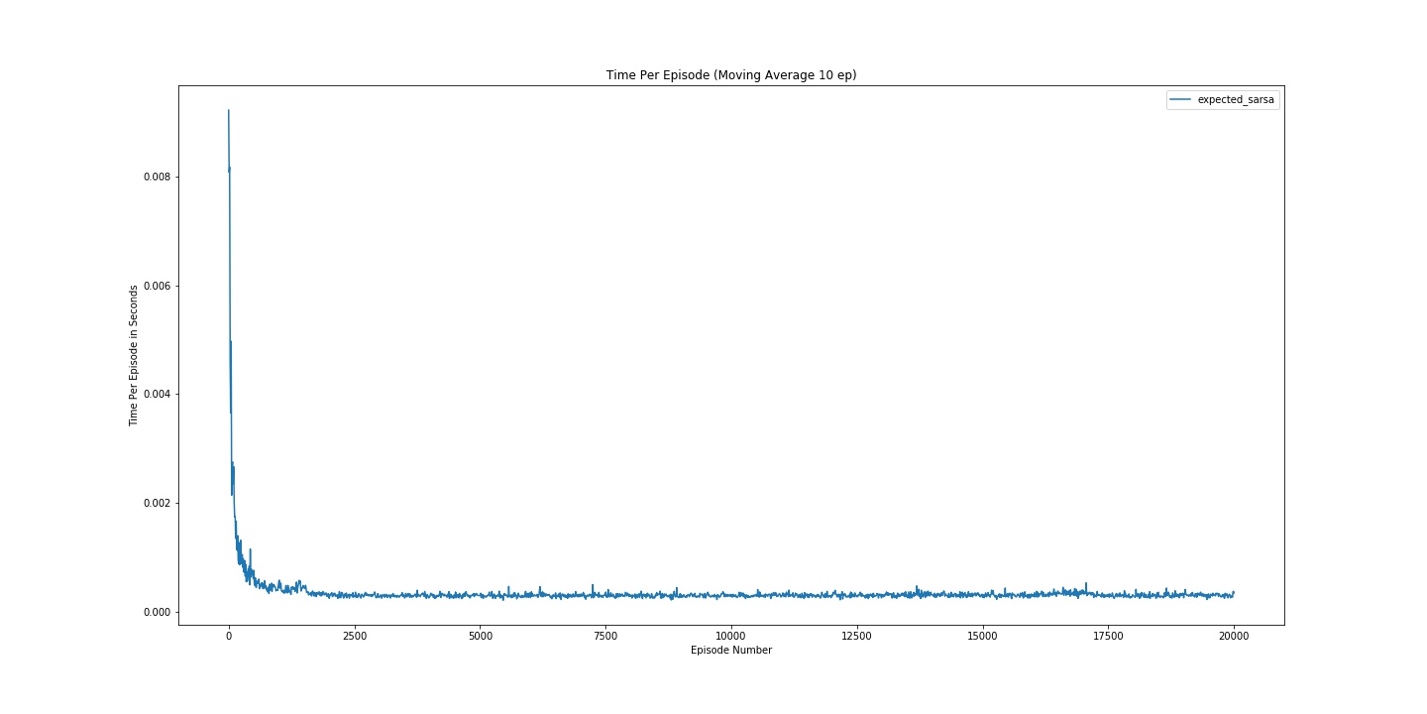


Figure 48: Expected Sarsa Time Per Episode (a = 0.05) Variable Eps

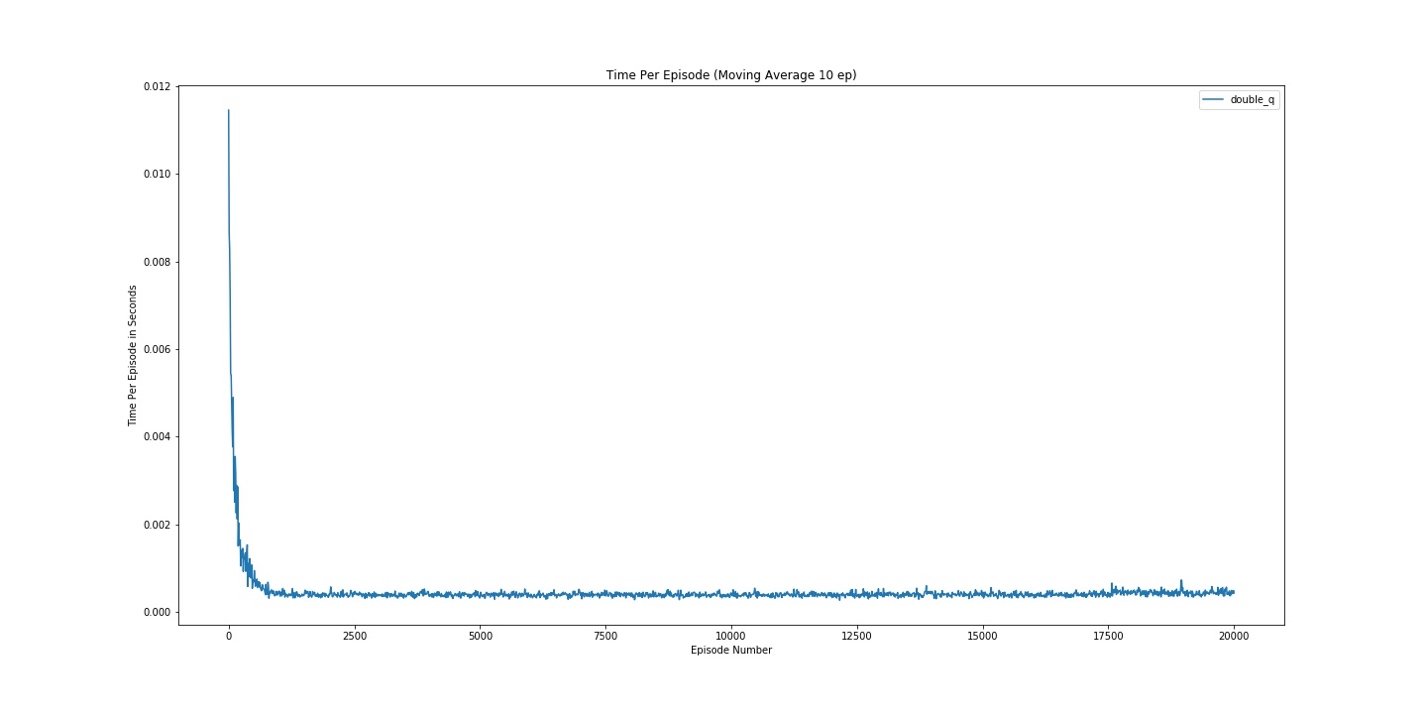


Figure 9: Double Q Time Per Episode (a = 0.05) Variable Eps

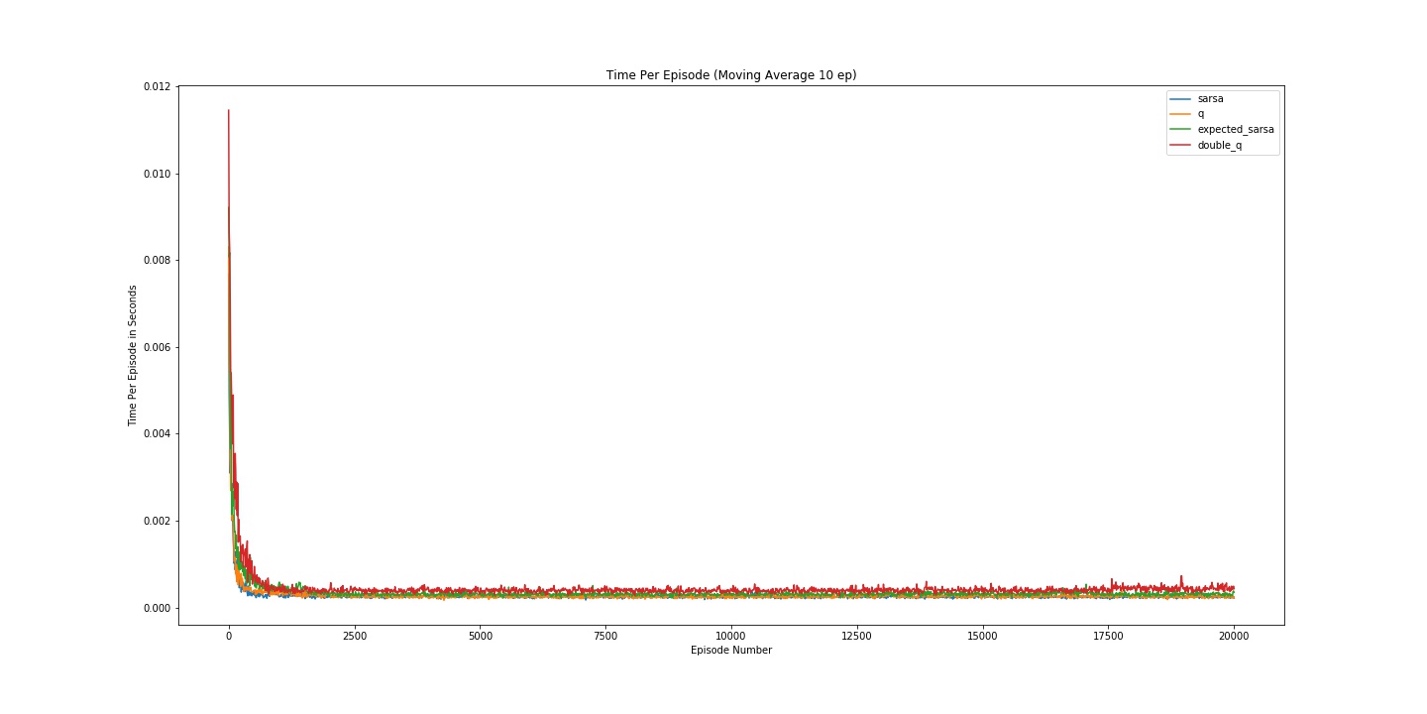


Figure 0: Combined Time Per Episode (a = 0.05) Variable Eps

*Alpha 0.1*

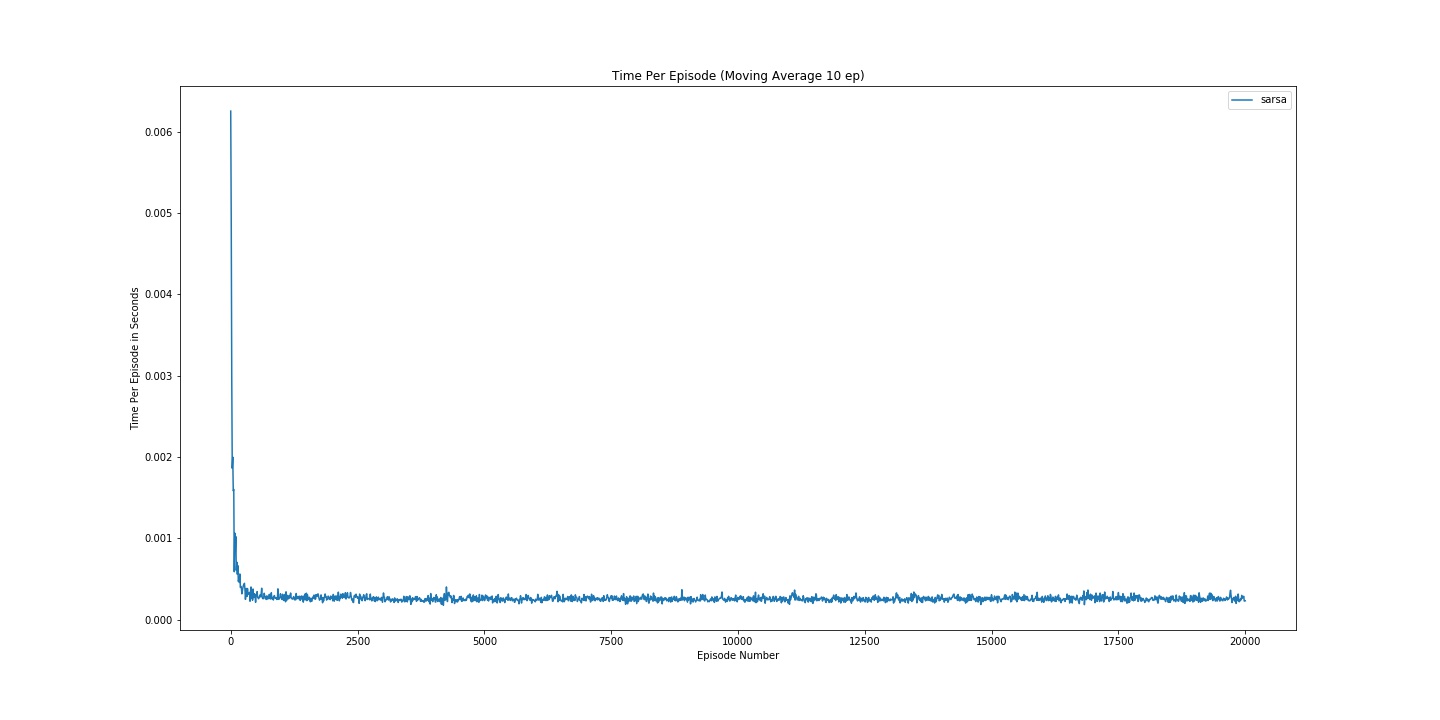


Figure 5: Sarsa Time Per Episode (a = 0.1) Variable Eps

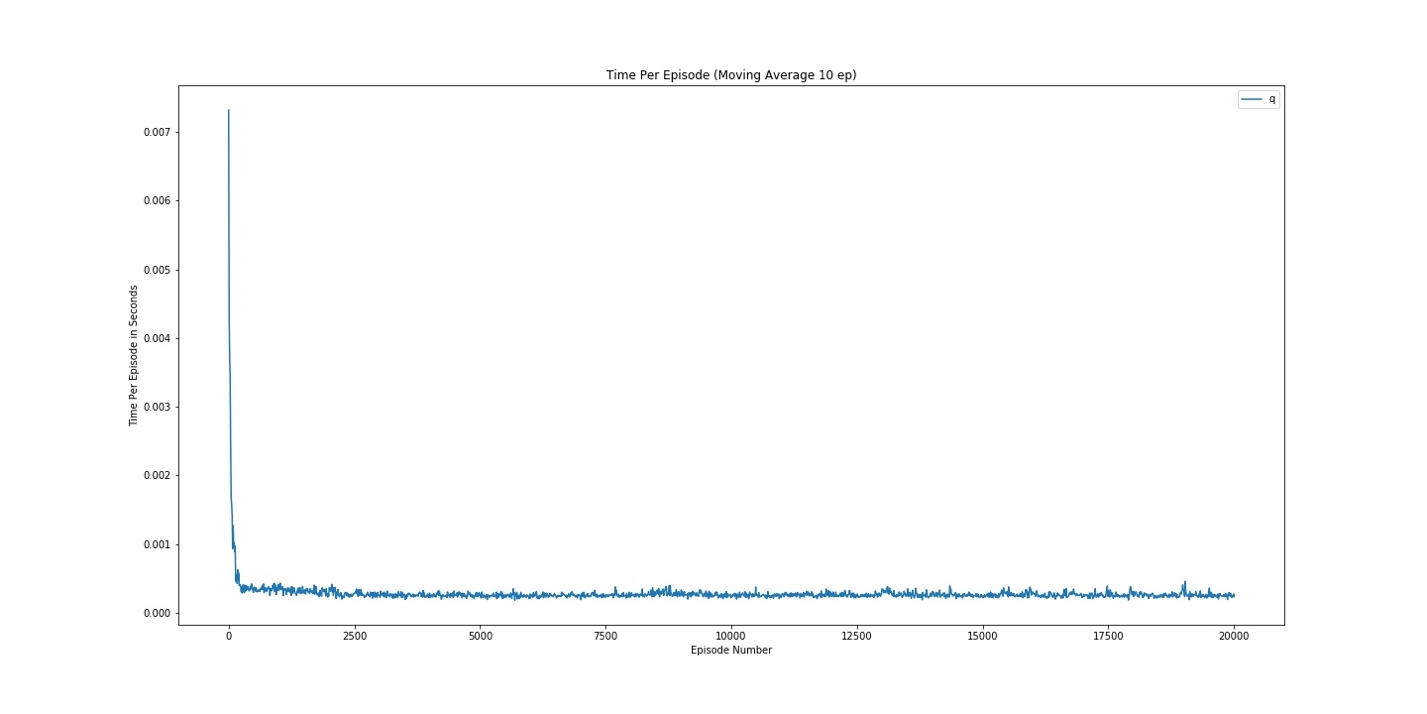


Figure 5: Q Time Per Episode (a = 0.1) Variable Eps

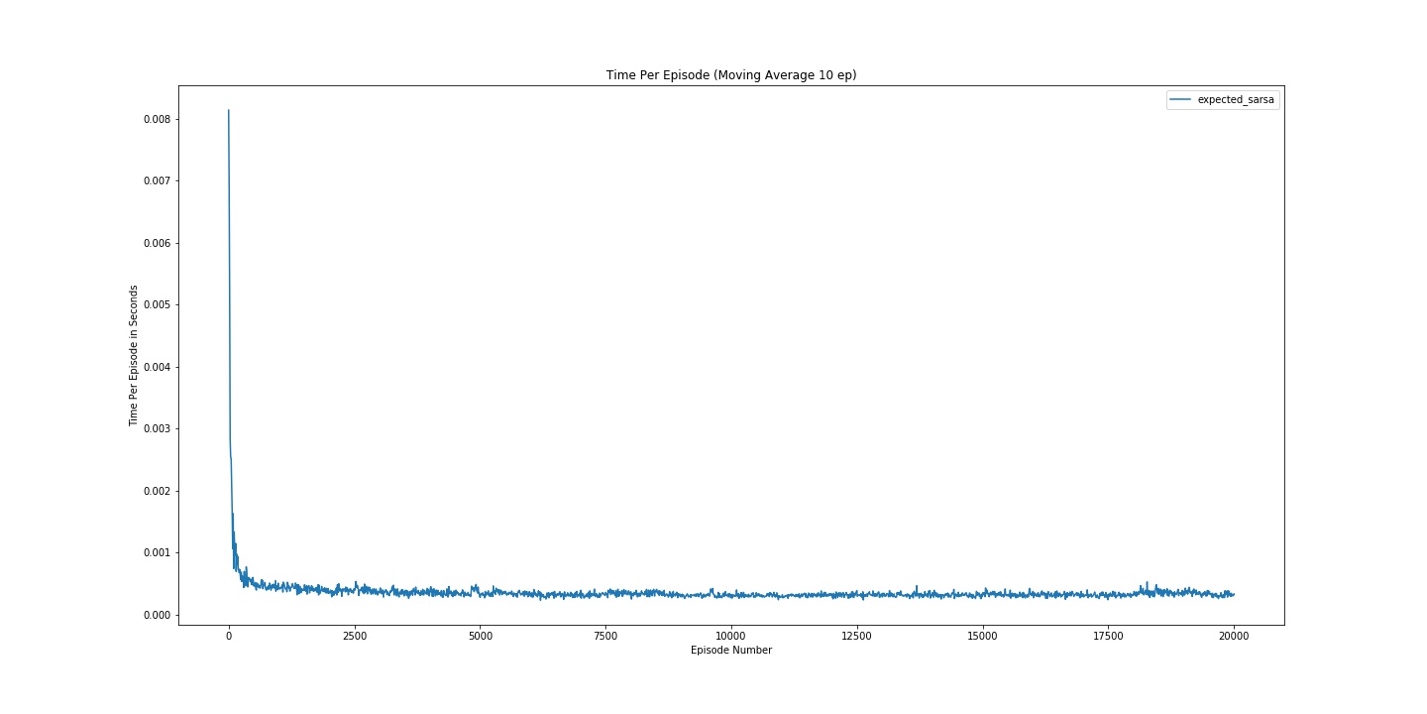


Figure 53: Expected Sarsa Time Per Episode (a = 0.1) Variable Eps

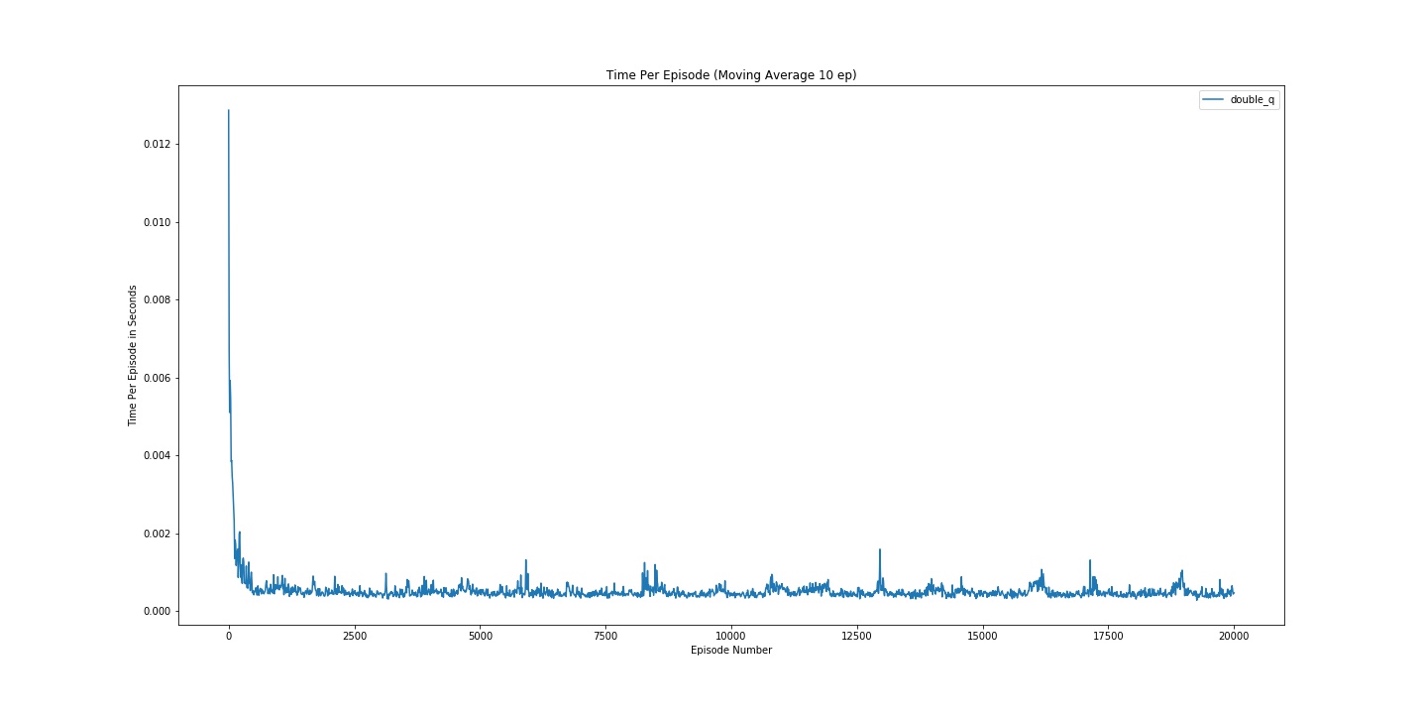


Figure 54: Double Q Time Per Episode (a = 0.1) Variable Eps

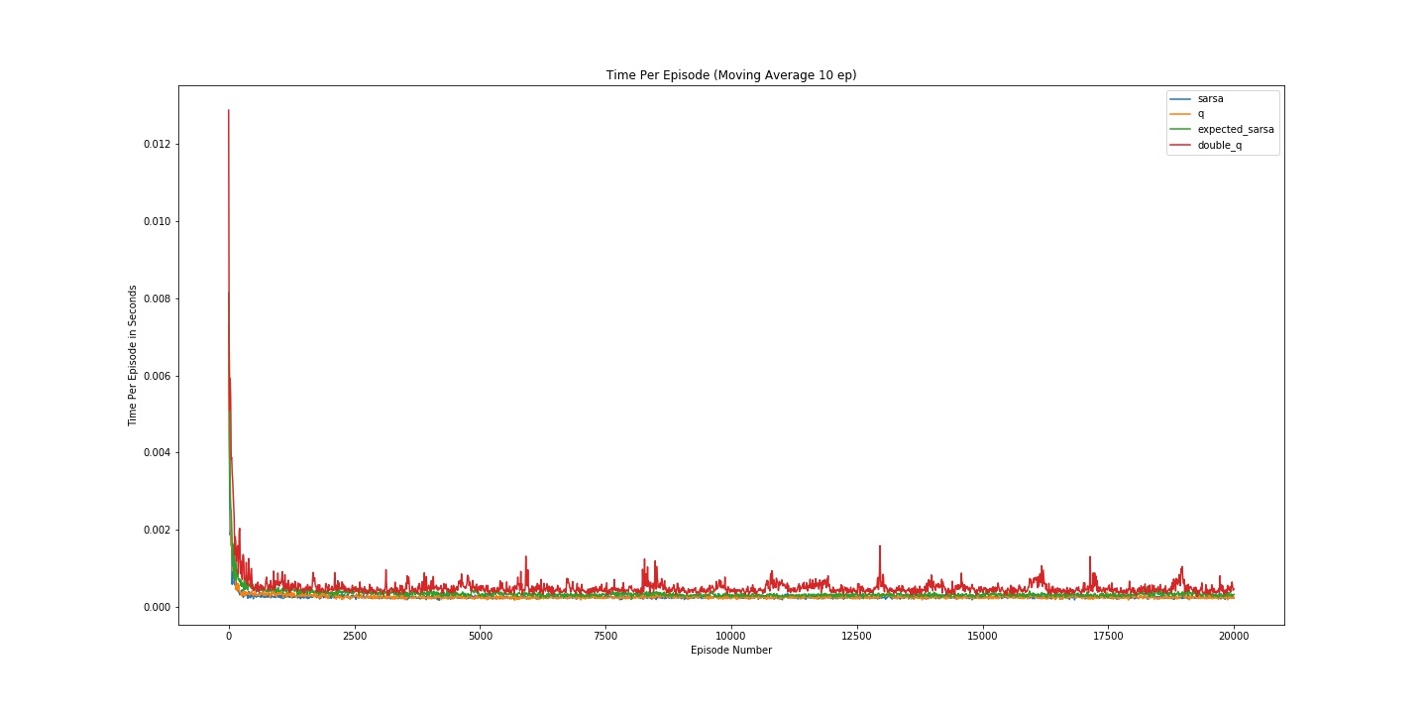


Figure 5: Combined Time Per Episode (a = 0.1) Variable Eps

*Alpha 0.2*

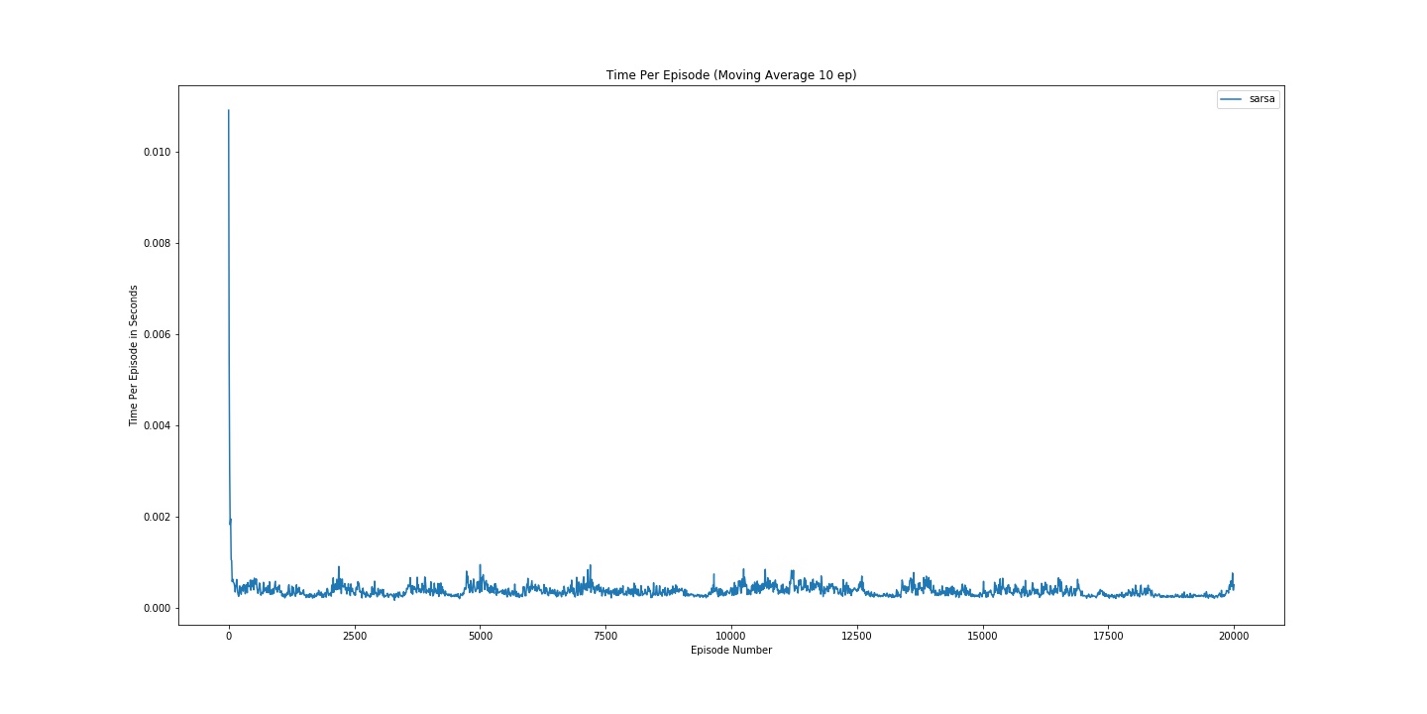


Figure 56: Sarsa Time Per Episode (a = 0.2) Variable Eps

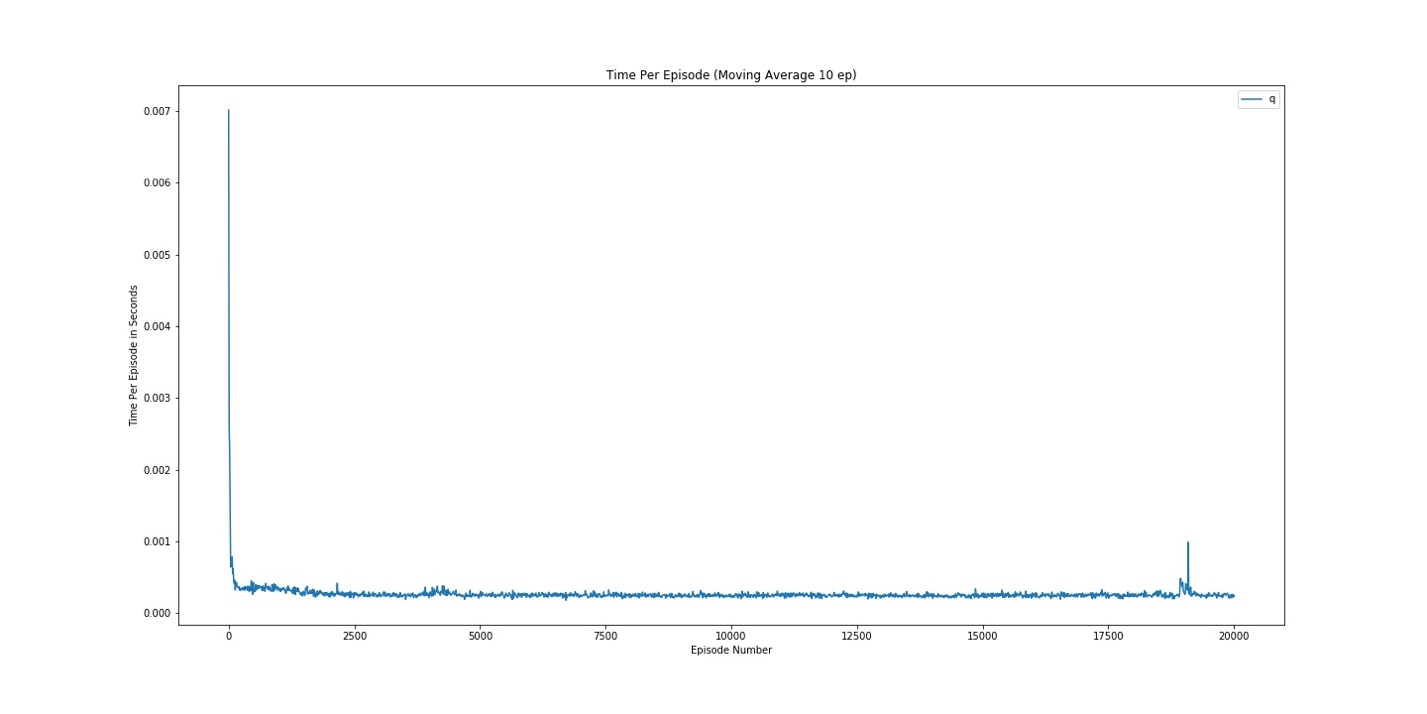


Figure 57: Q Time Per Episode (a = 0.2) Variable Eps

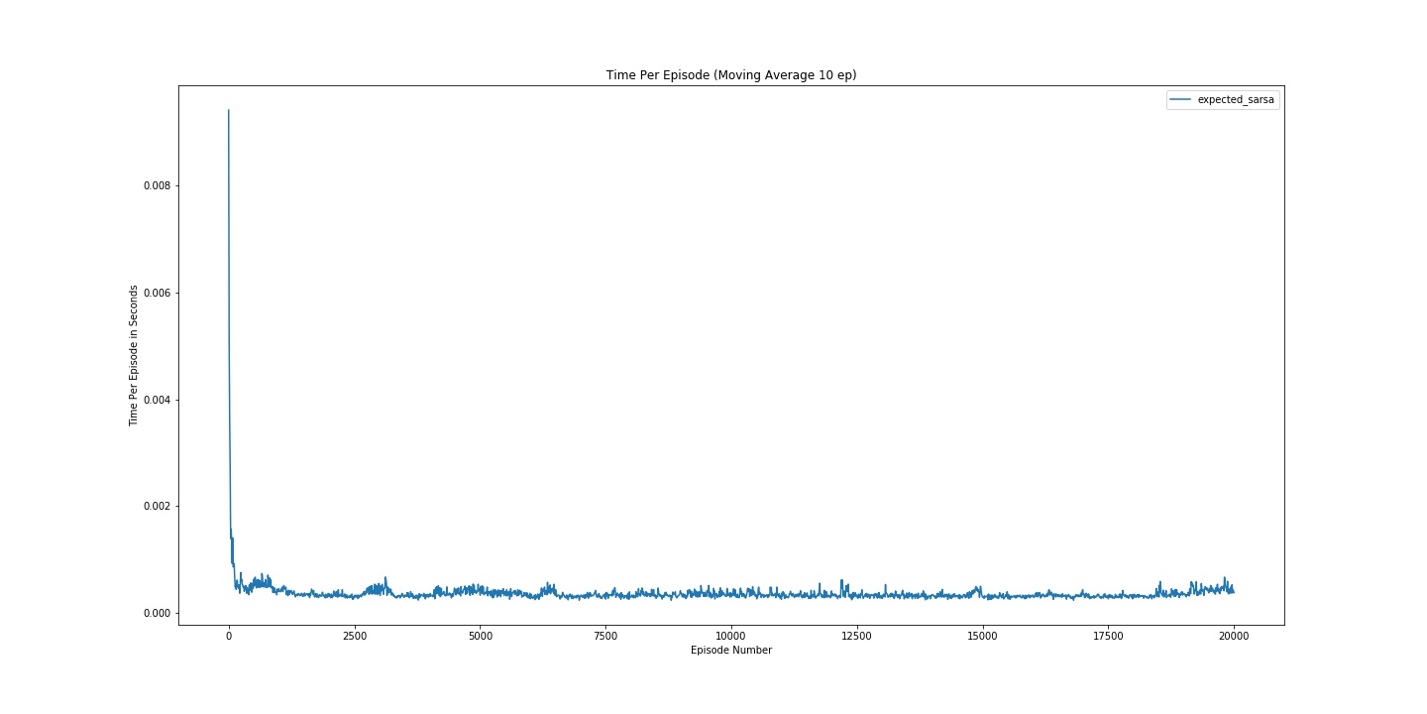


Figure 58: Expected Sarsa Time Per Episode (a = 0.2) Variable Eps

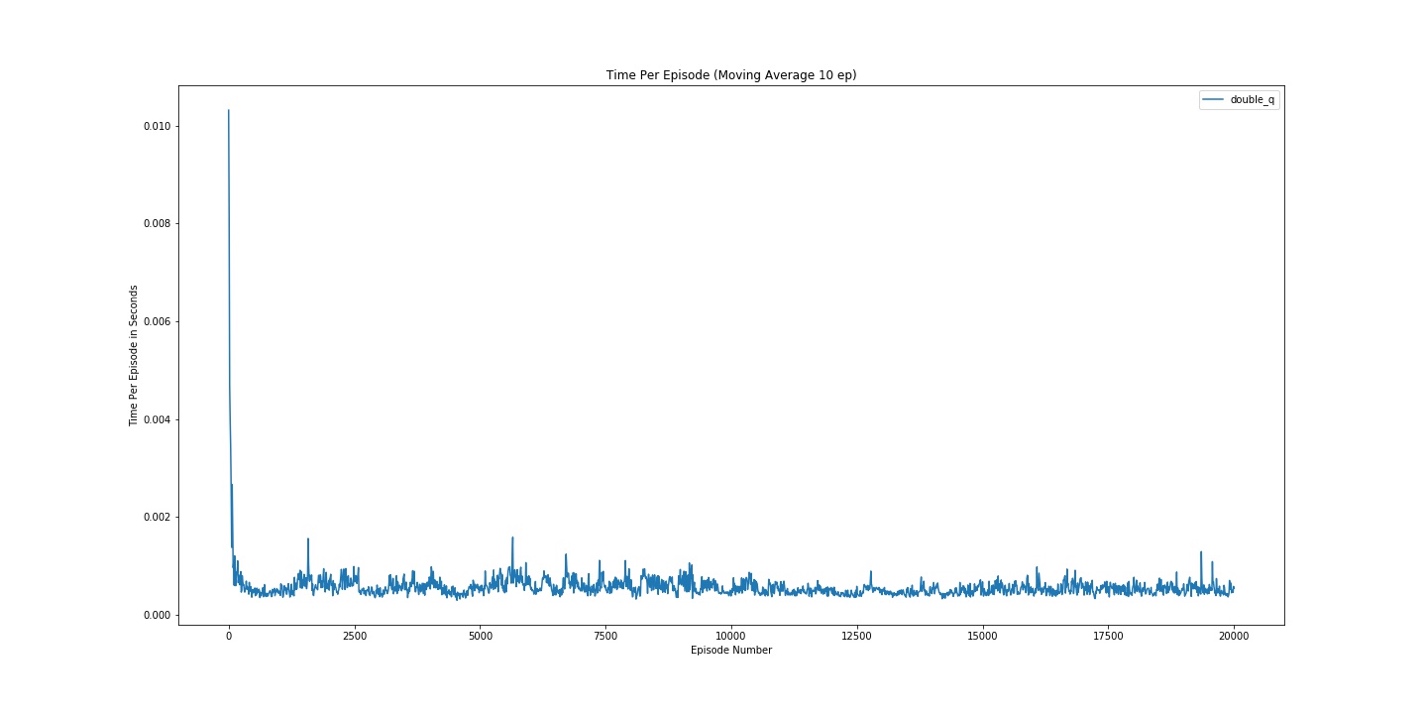


Figure 59: Double Q Time Per Episode (a = 0.2) Variable Eps

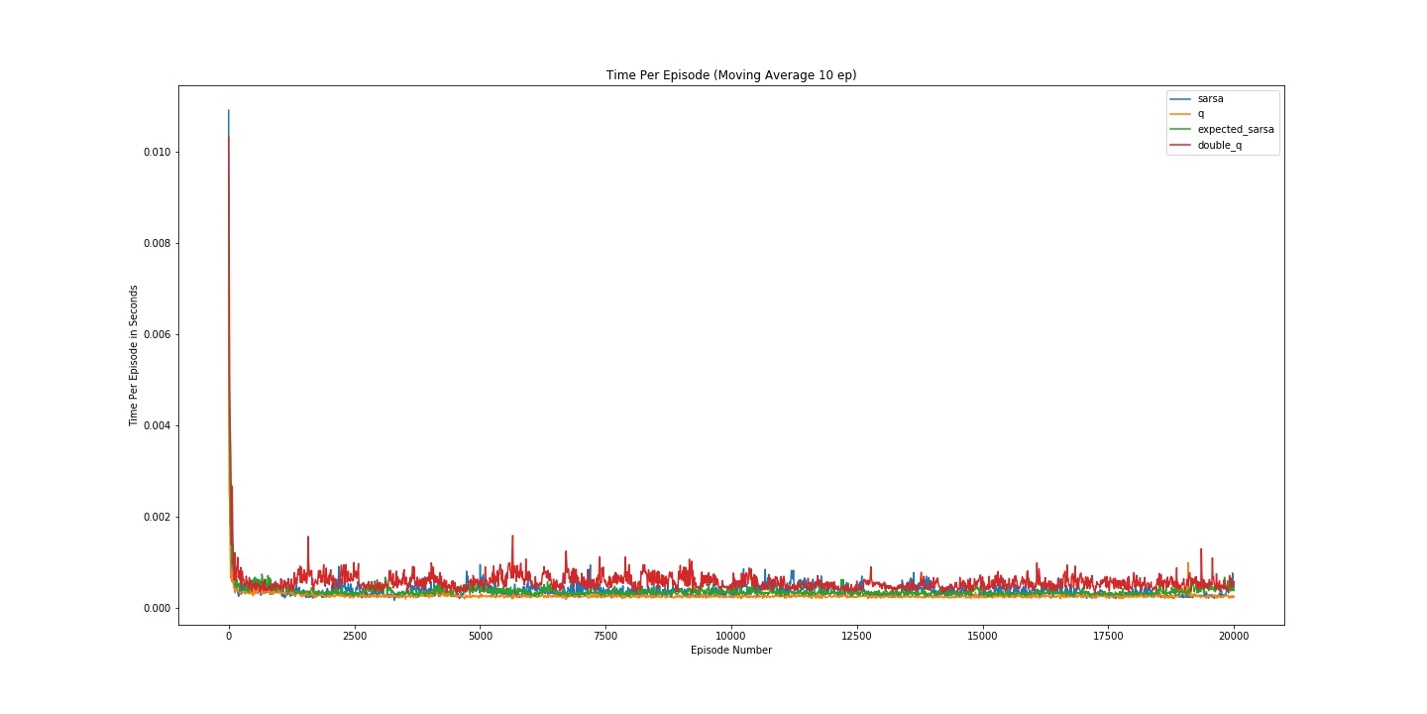


Figure 60: Combined Time Per Episode (a = 0.2) Variable Eps

**Number of Steps per Episode Running Average**

*Alpha 0.05*

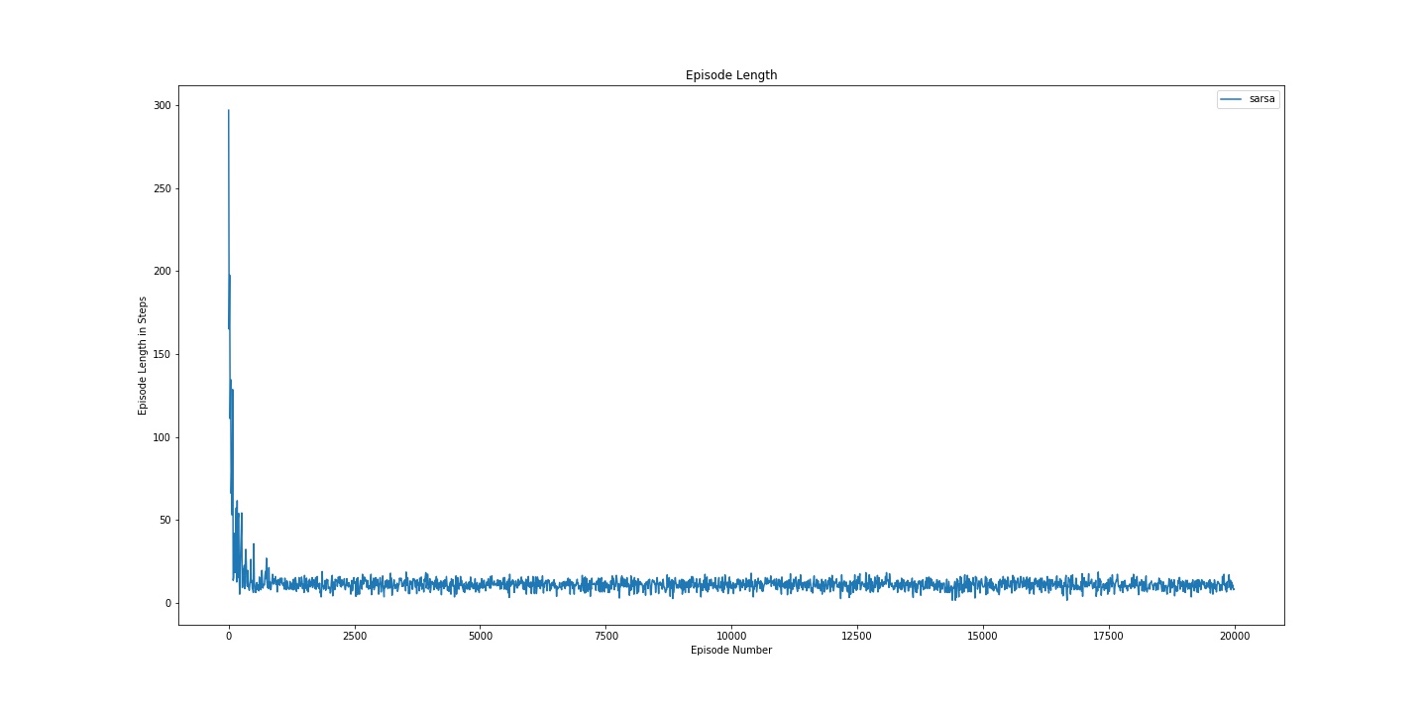


Figure 61: Sarsa Num Episodes (a = 0.05) Variable Eps

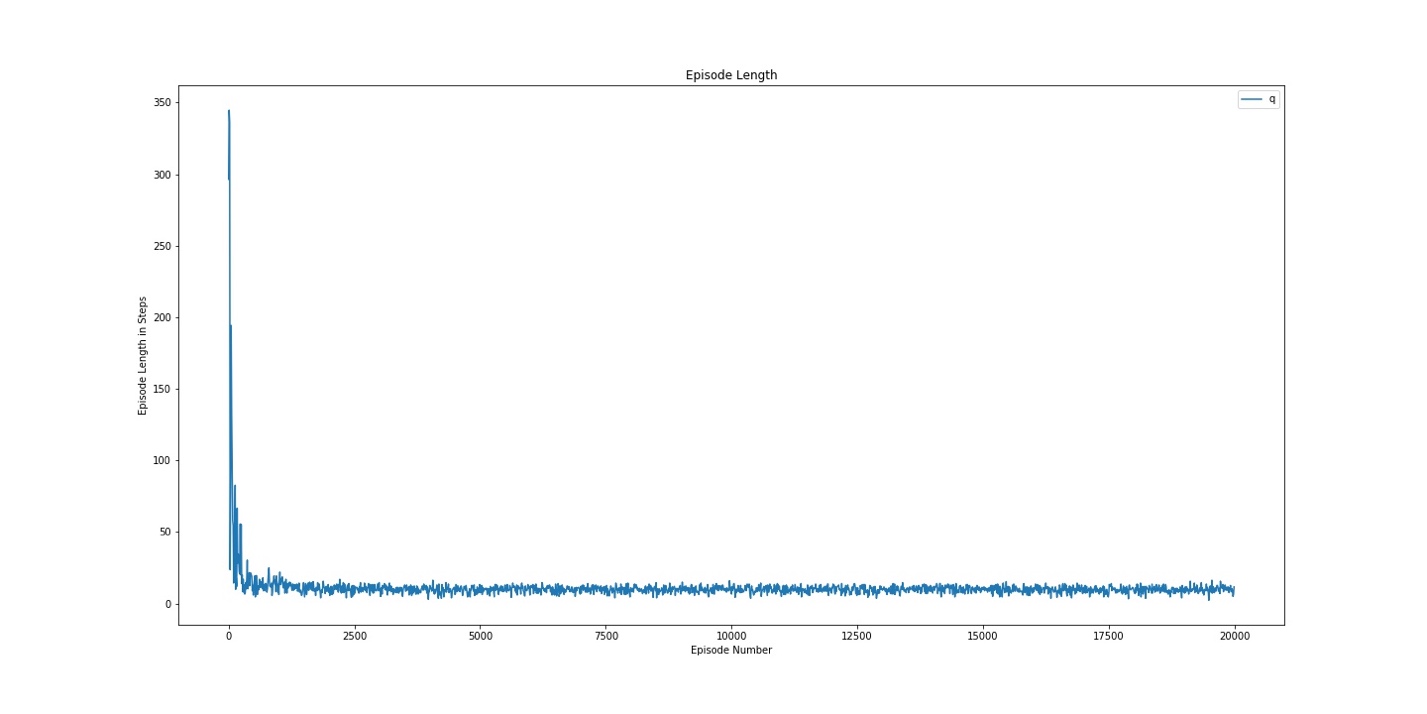


Figure 62: Q Num Episodes (a = 0.05) Variable Eps

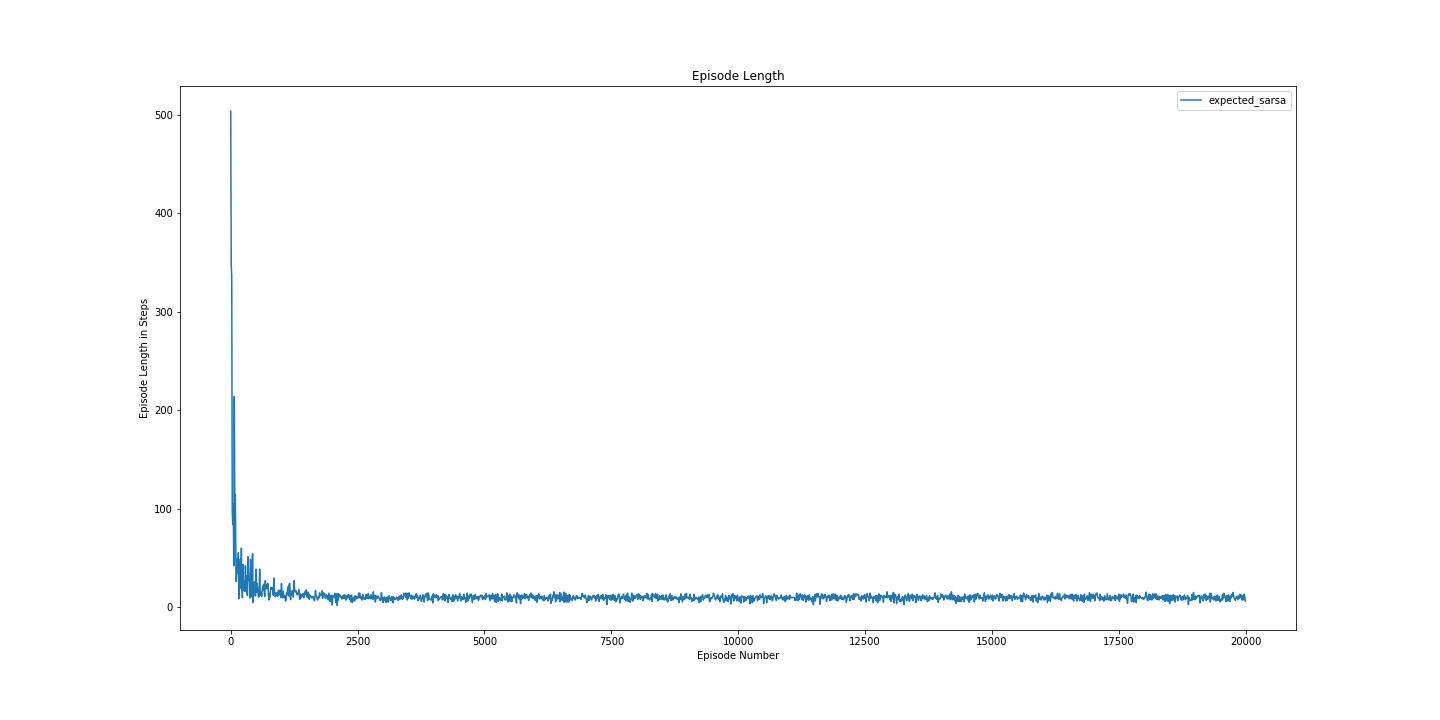


Figure 63: Expected Sarsa Num Episodes (a = 0.05) Variable Eps

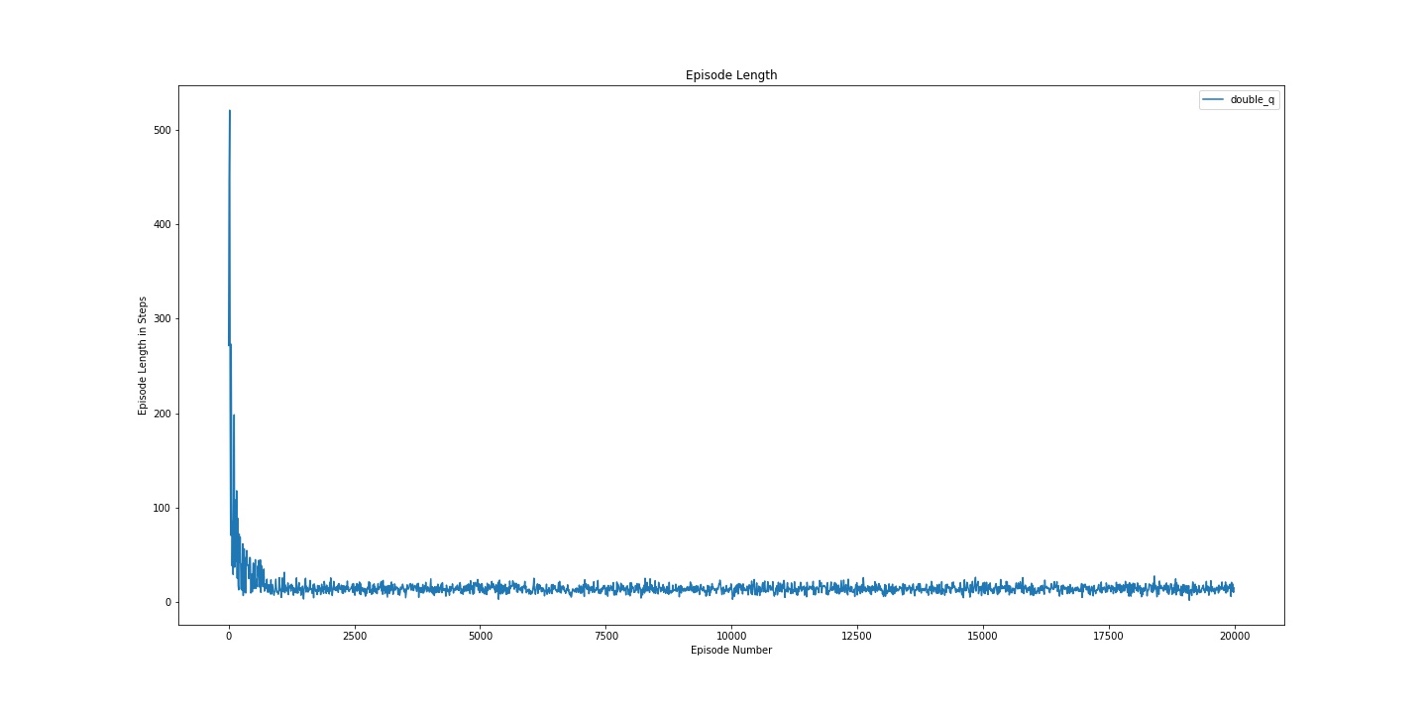


Figure 64: Double Q Num Episodes (a = 0.05) Variable Eps

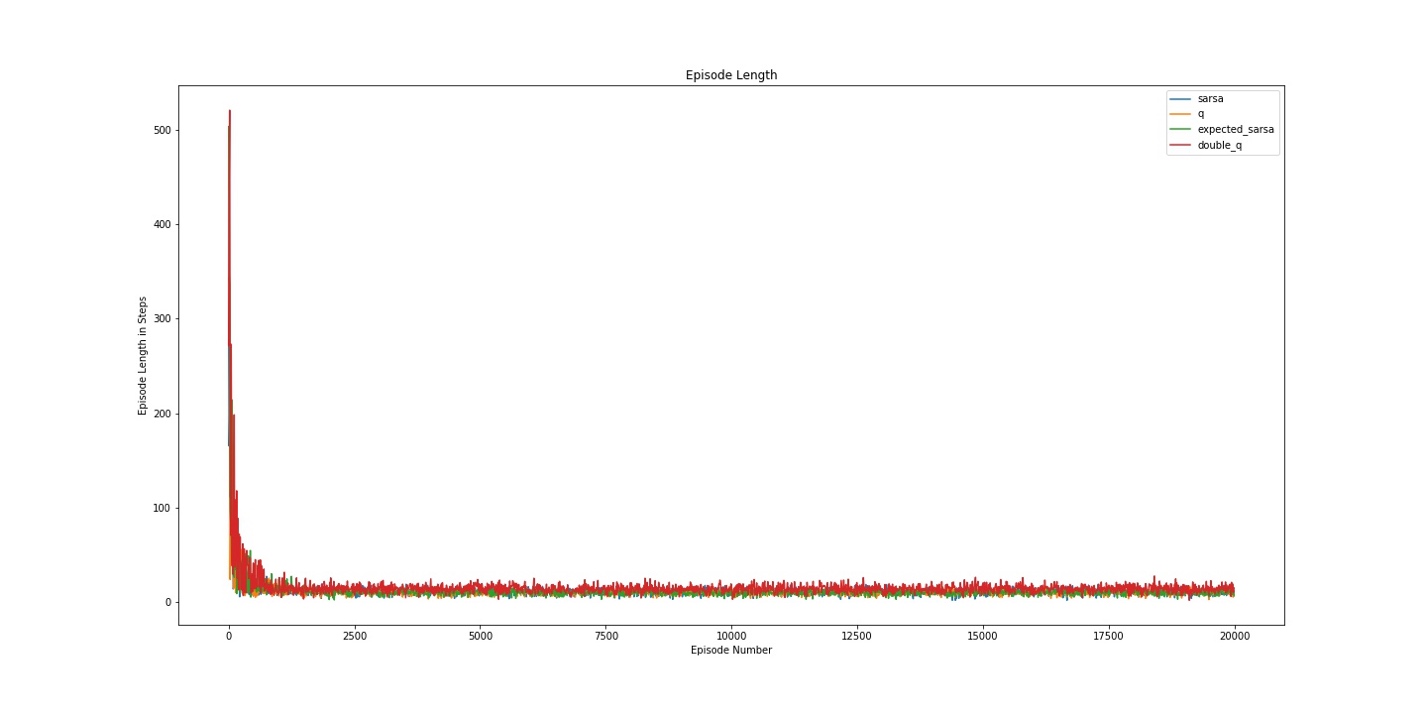


Figure 65: Combined Num Episodes (a = 0.05) Variable Eps

*Alpha 0.1*

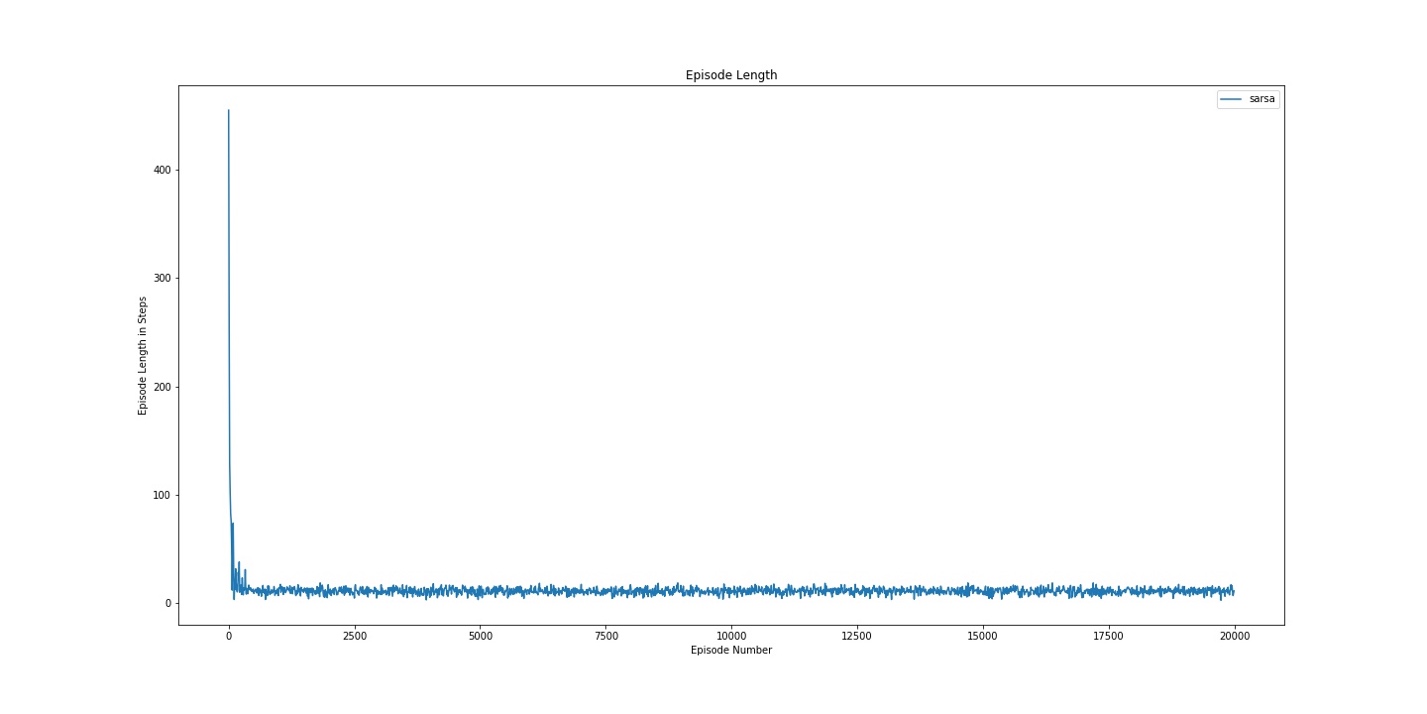


Figure 66: Sarsa Num Episodes (a = 0.1) Variable Eps

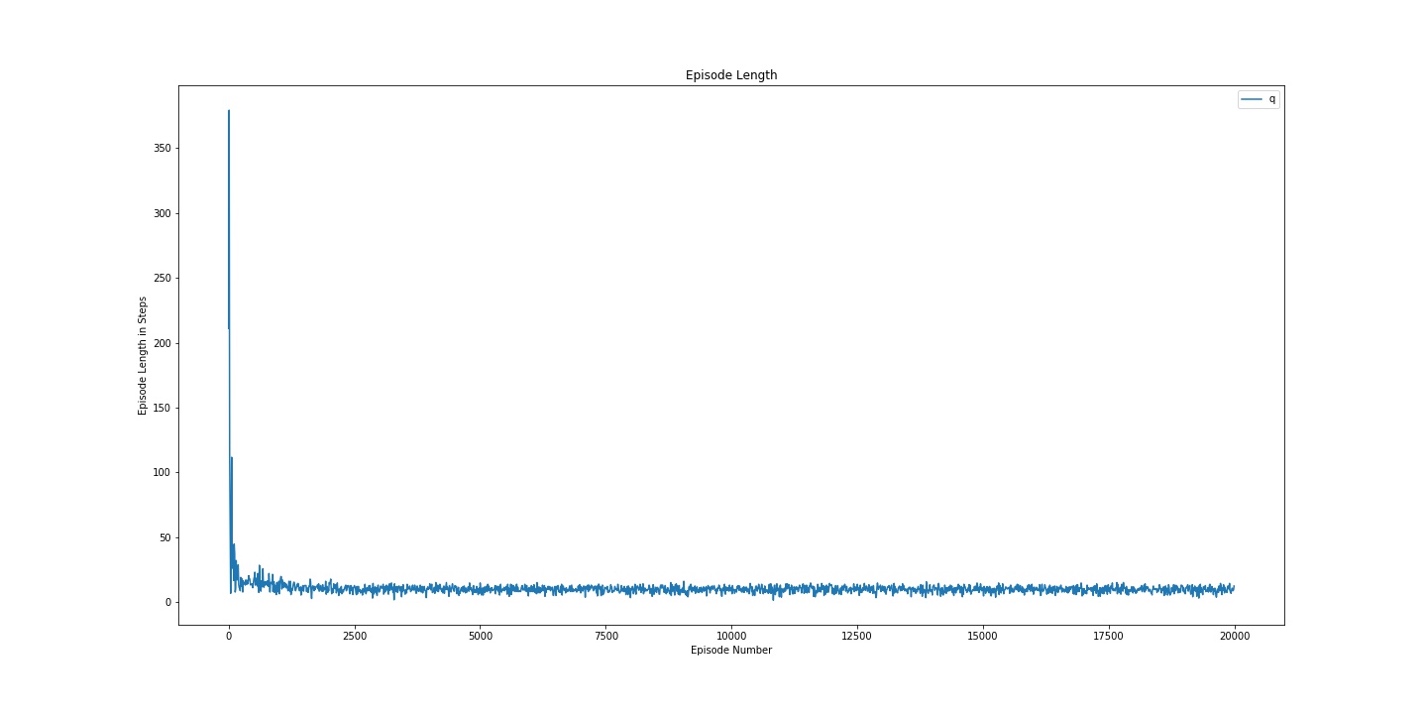


Figure 67: Q Num Episodes (a = 0.1) Variable Eps

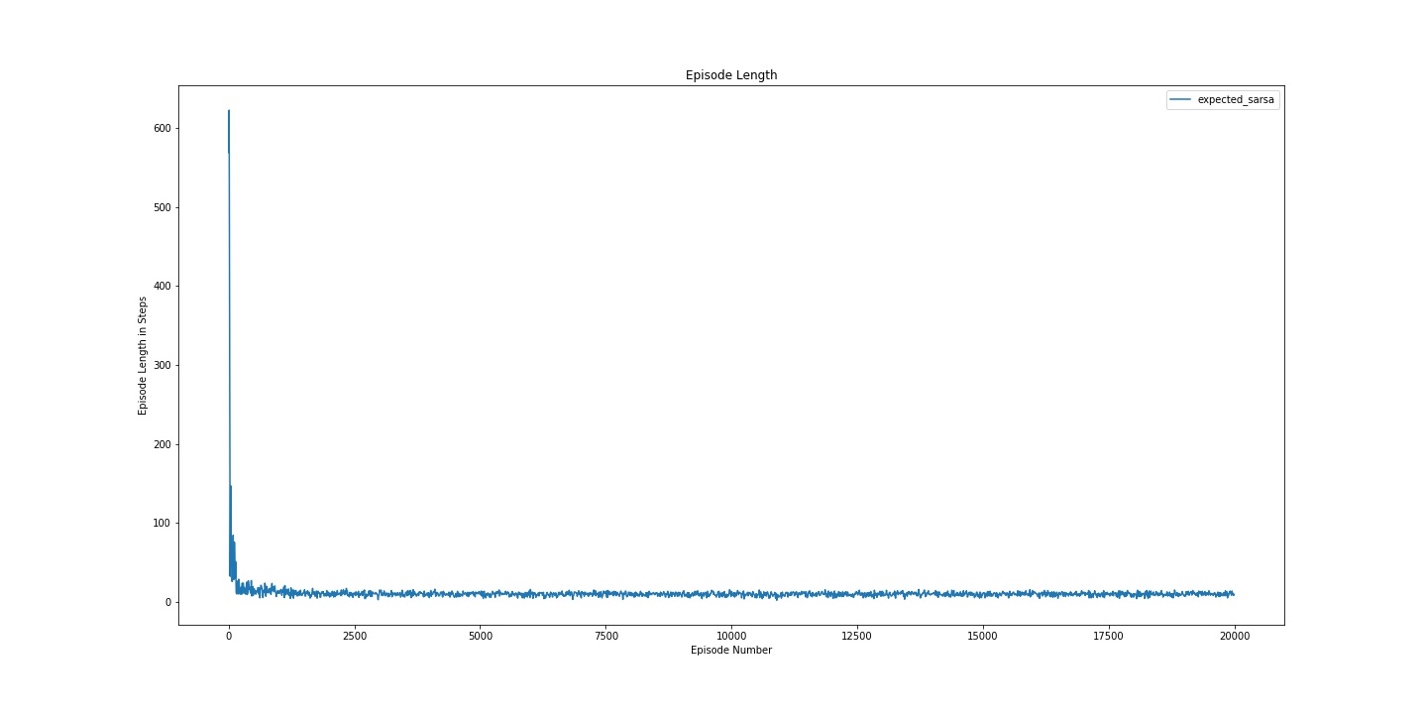


Figure 68: Expected Sarsa Num Episodes (a = 0.1) Variable Eps

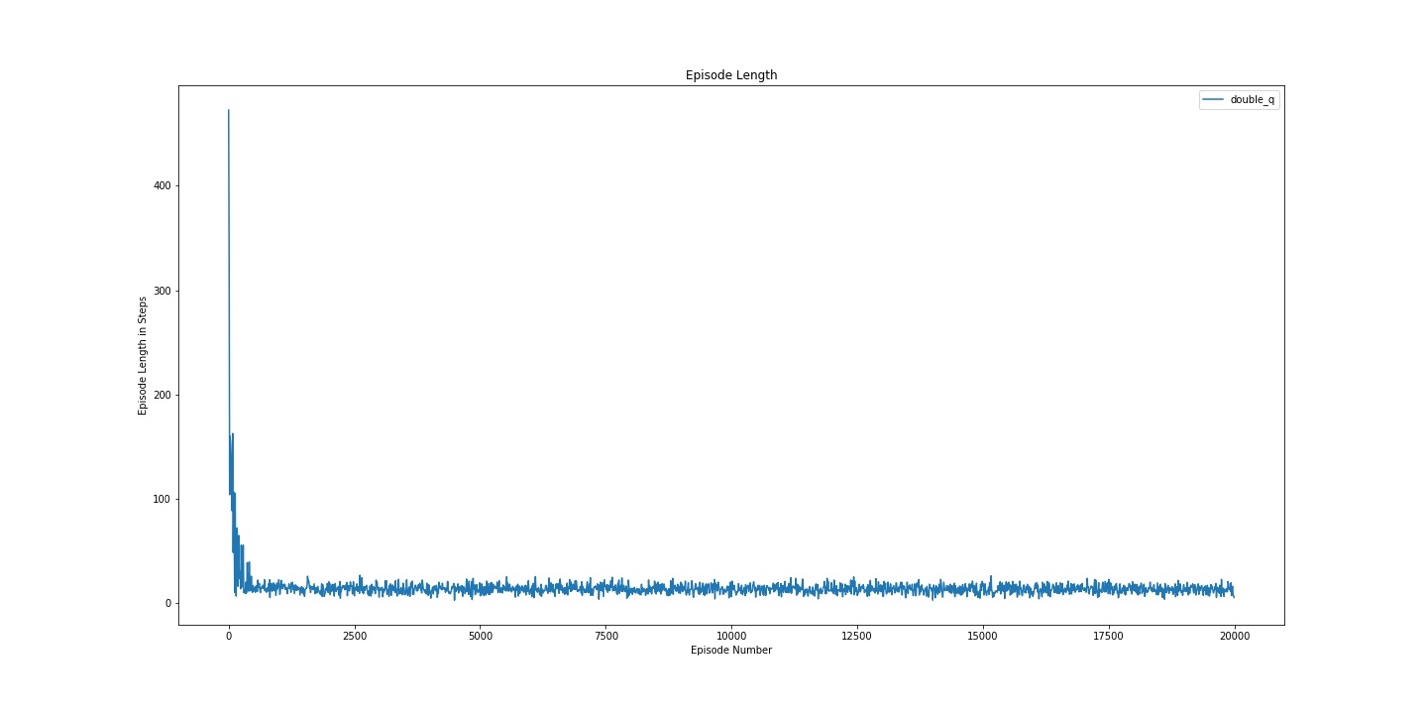


Figure 69: Double Q Num Episodes (a = 0.1) Variable Eps

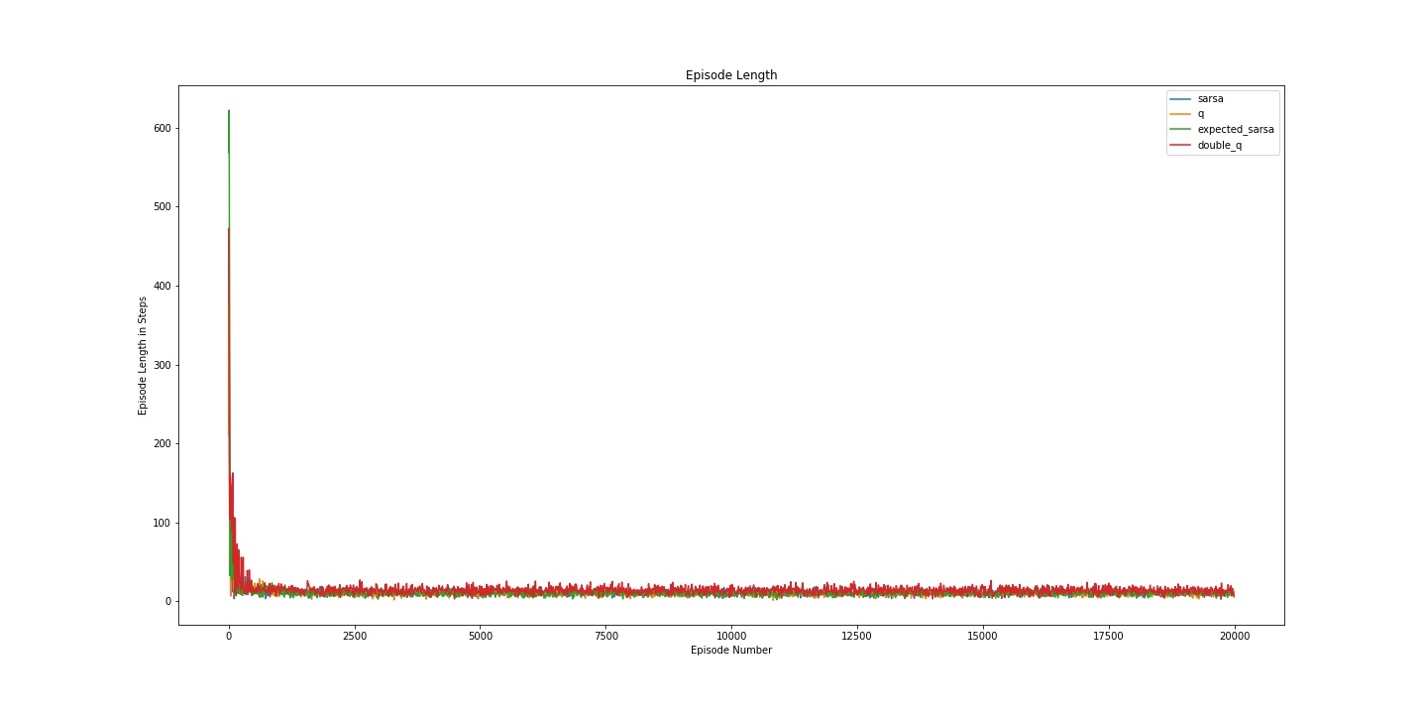


Figure 70: Combined Num Episodes(a = 0.1) Variable Eps

*Alpha 0.2*

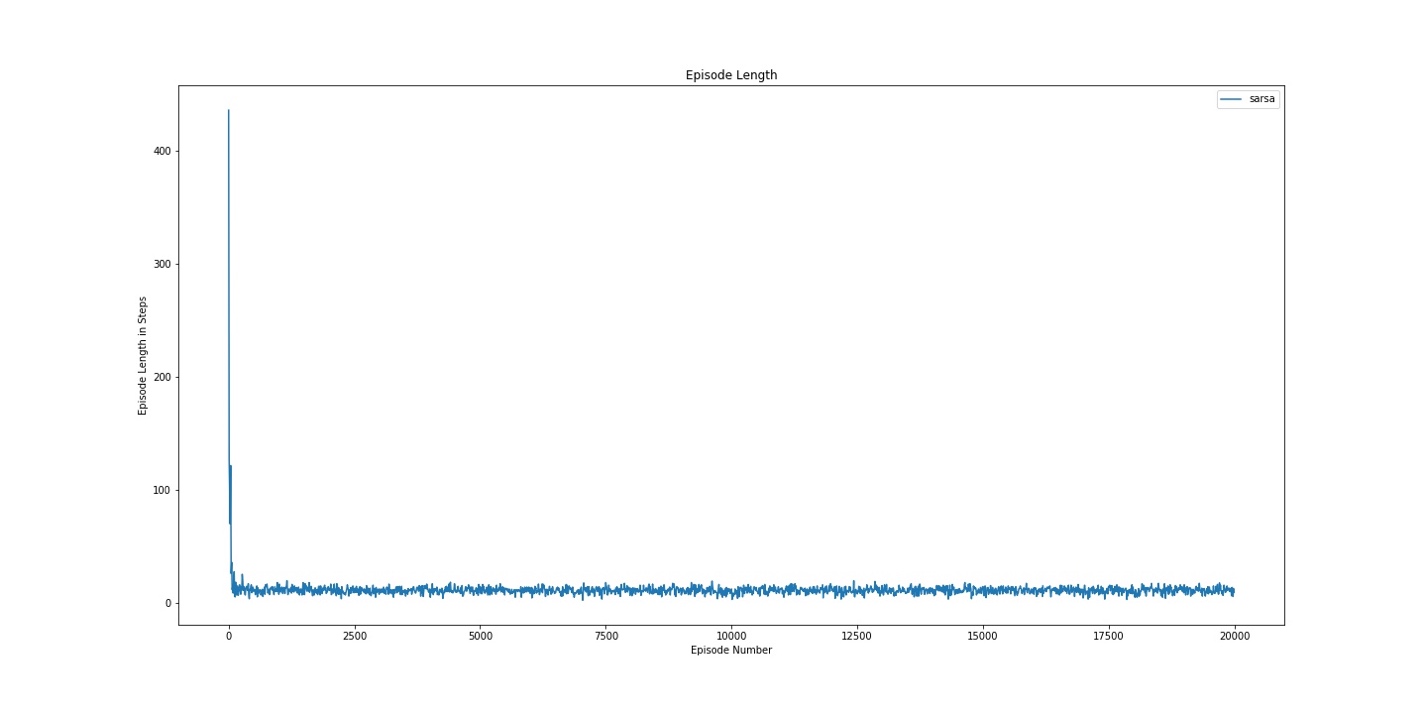


Figure 71: Sarsa Num Episodes (a = 0.2) Variable Eps

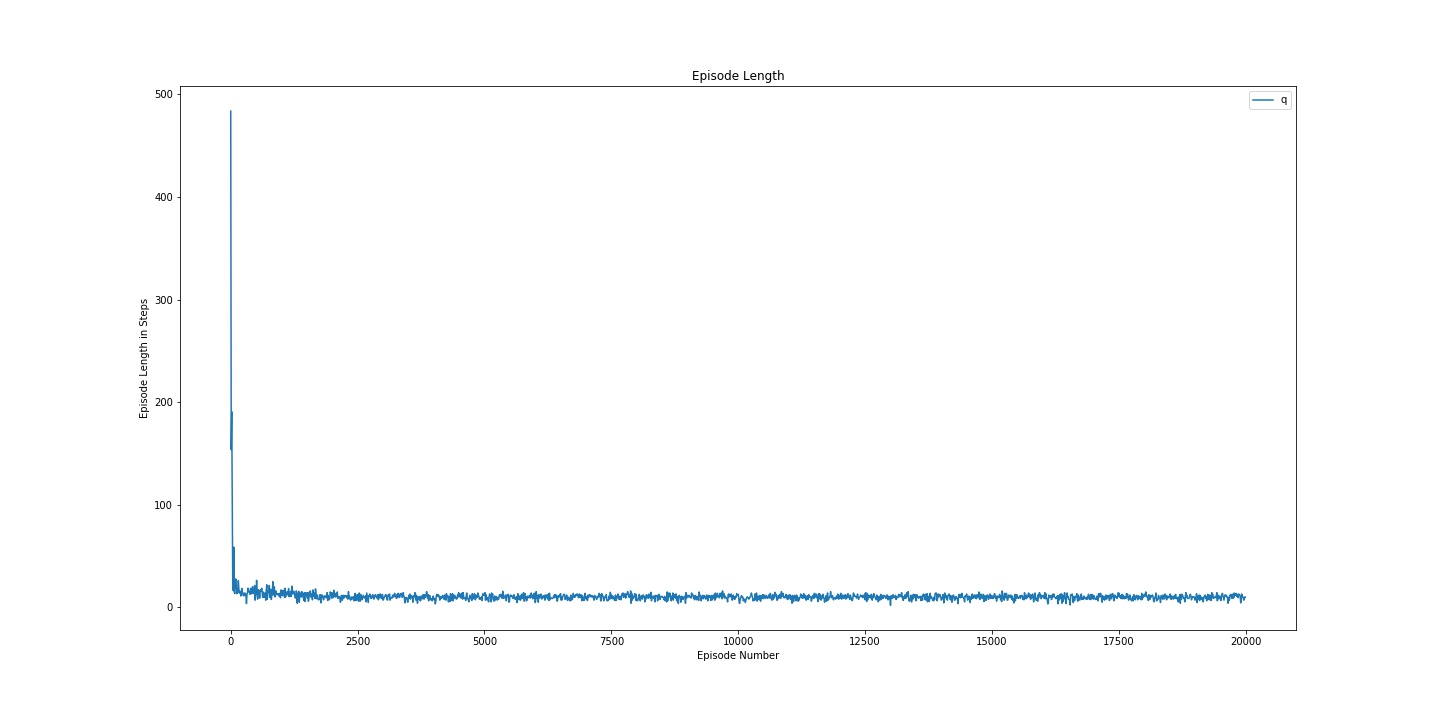


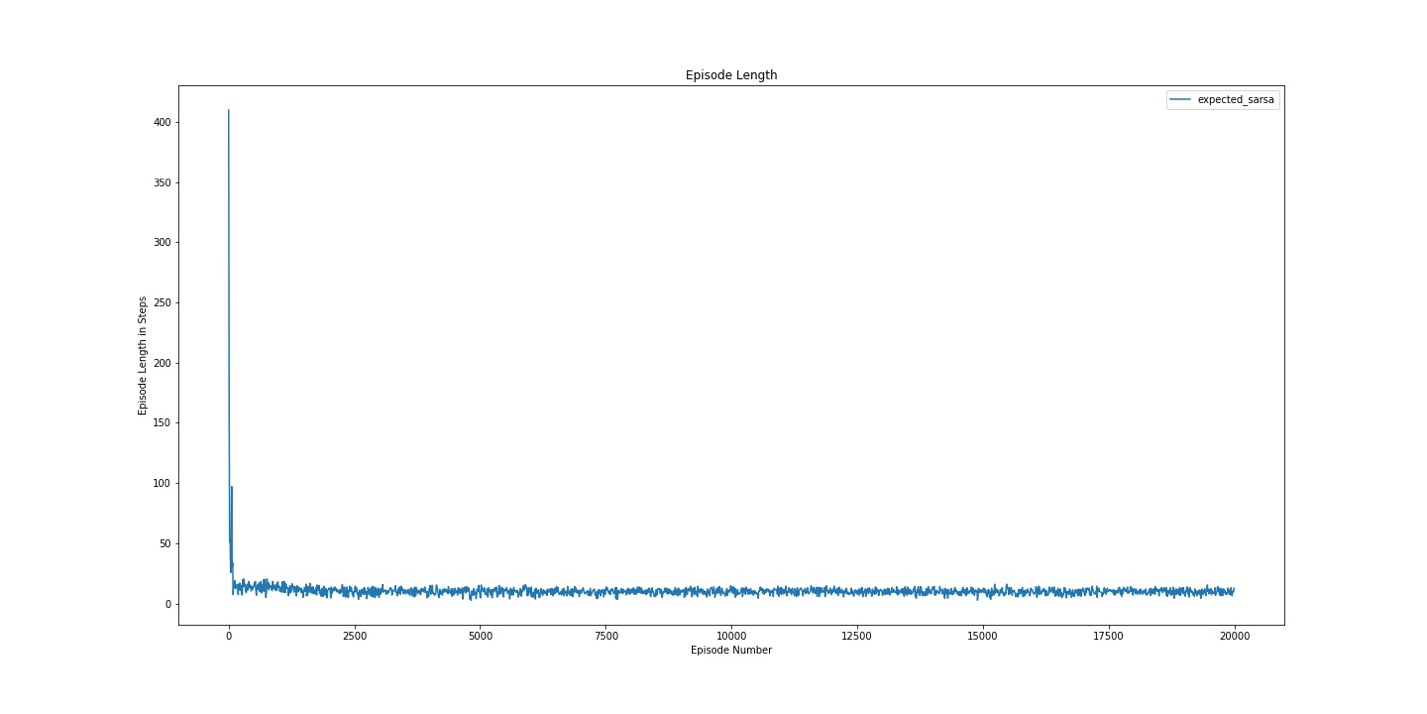
Figure 72: Q Num Episodes(a = 0.2) Variable Eps 

Figure 73: Expected Sarsa Num Episodes(a = 0.2) Variable Eps

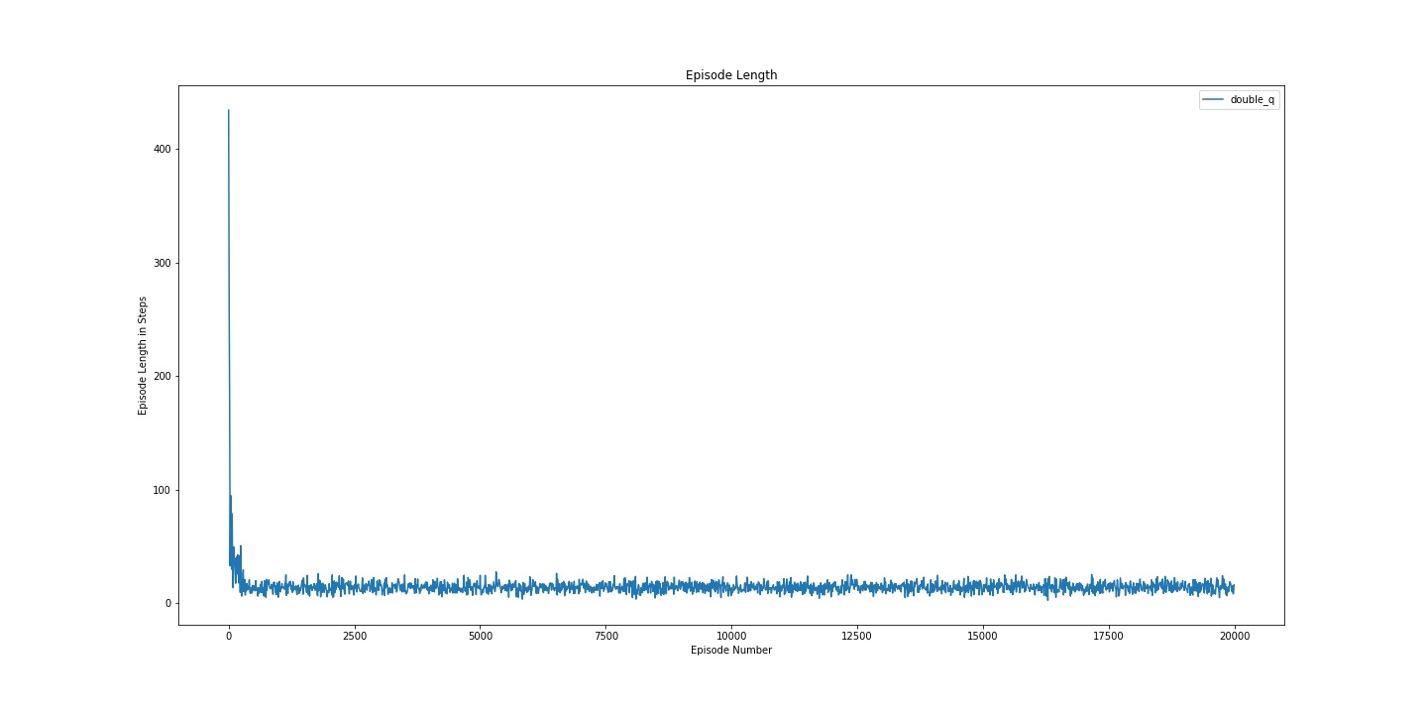


Figure 74: Double Q Num Episodes (a = 0.2) Variable Eps

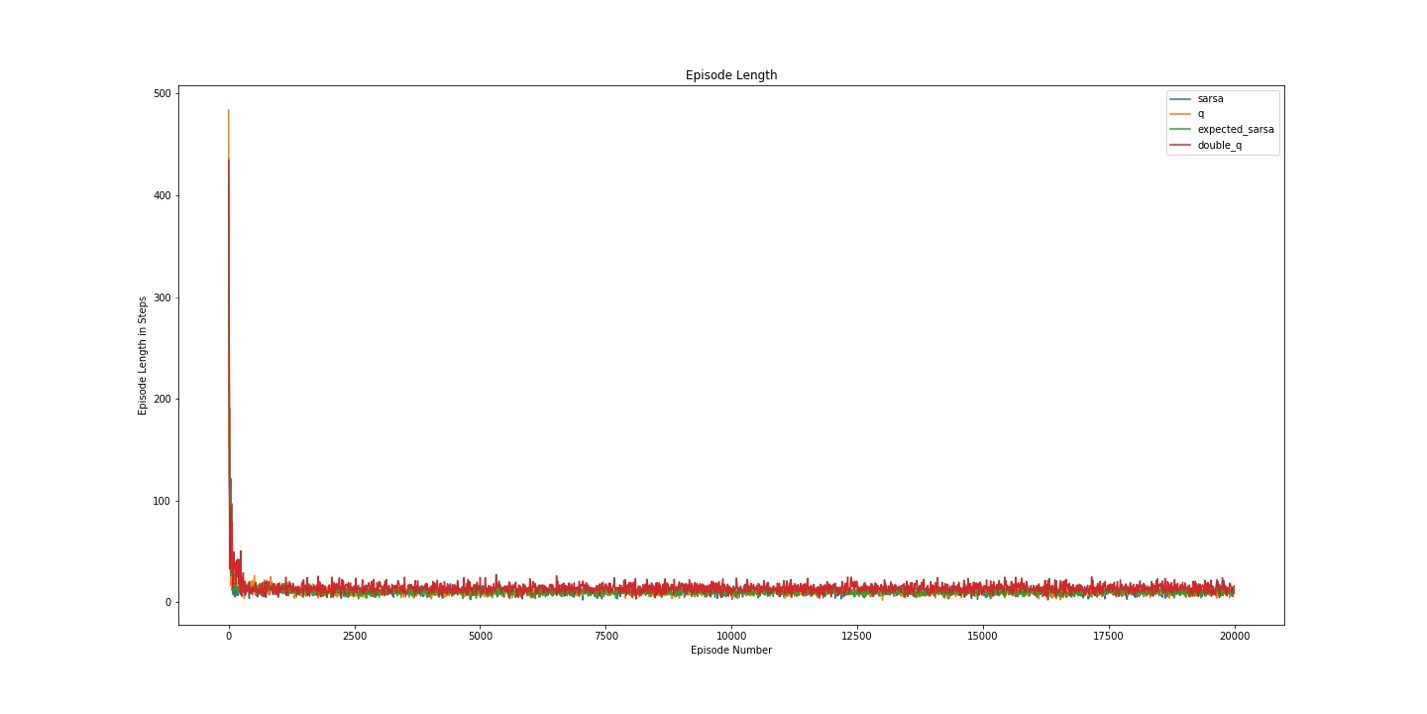


Figure 75: Combined Num Episodes (a = 0.2) Variable Eps