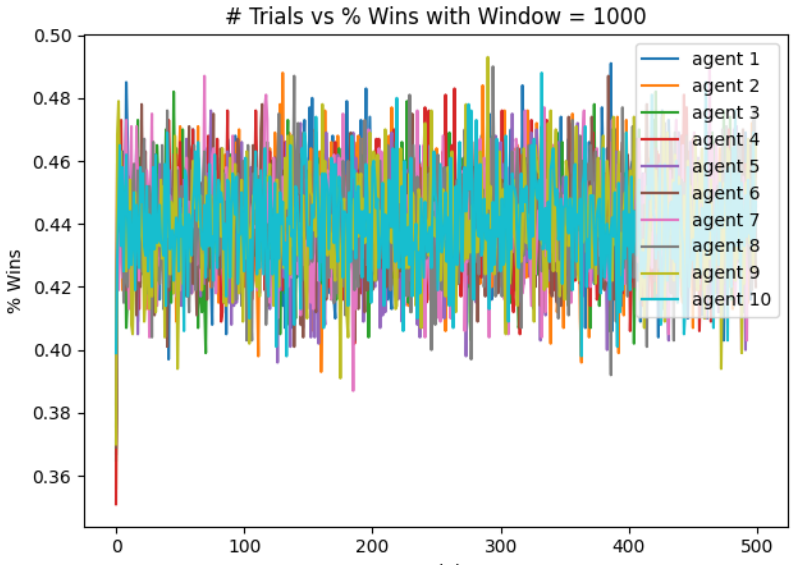
Kordel K. France
21 February 2023
EN 649.605.8VL

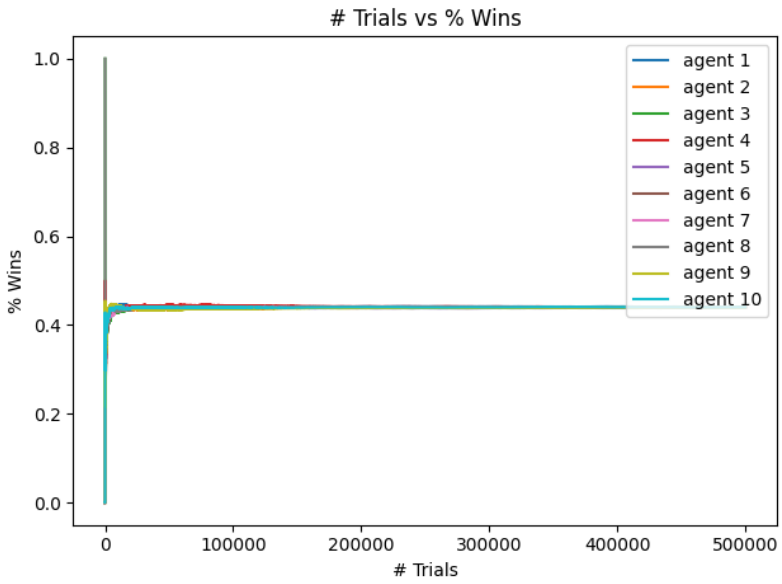
Project #2

I noticed results were relatively flat after the first 5000 iterations. This indicated that the agent had "learned all it could" early, which was good. However, my exploration rate ϵ remained constant throughout all 500,000 episodes of each agent. This is not desired because once the agent learns and its policy begins to converge to optimality, we have a lesser desire to explore as much. Due to this, I programmed the exploration rate to be a function of the number of episodes such that the agent would explore less as it learned more. So each agent starts off with an exploration rate of 1.0 and ends each trial with an exploration rate just slightly greater than zero.

The average win rate for each agent (not including draws) along with the summary of trials for over all 500,000 episodes can be seen below.

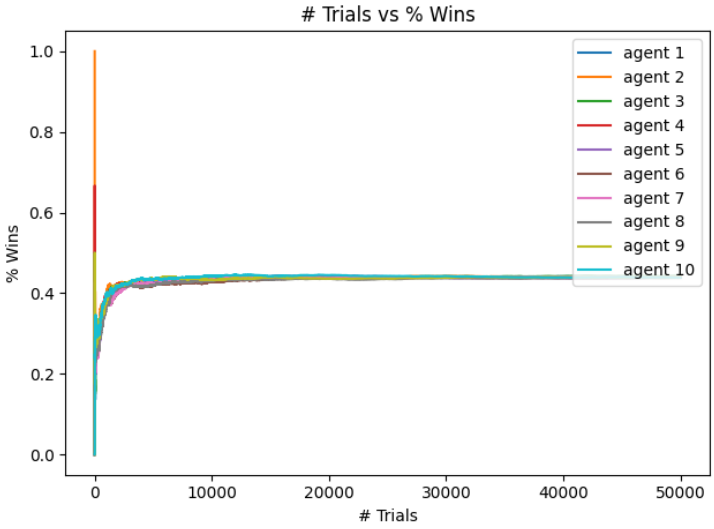
Agent #	Average Win Rate
1	44.08%
2	44.02%
3	43.97%
4	44.00%
5	43.93%
6	44.08%
7	44.05%
8	44.16%
9	44.01%
10	44.01%

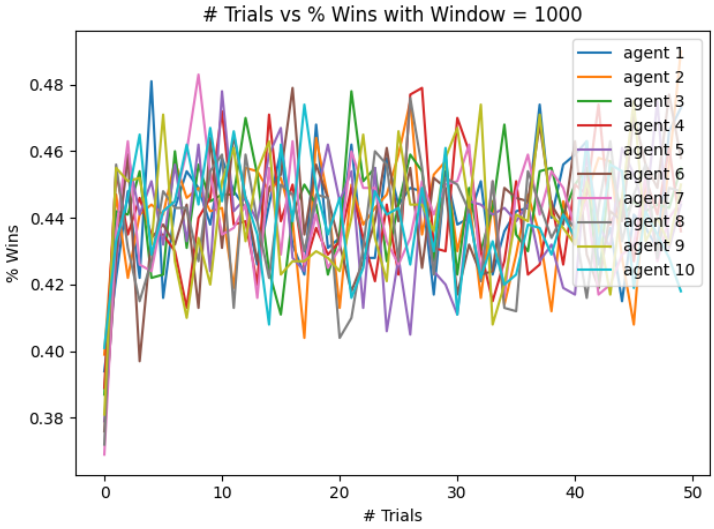




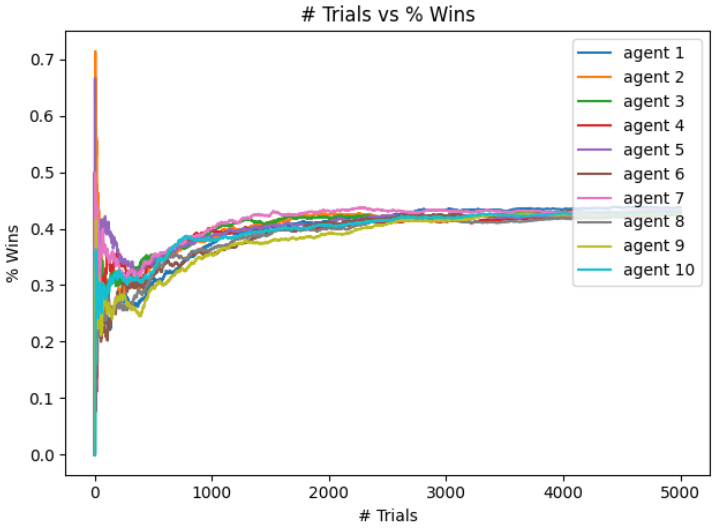
Kordel K. France
21 February 2023
EN 649.605.8VL

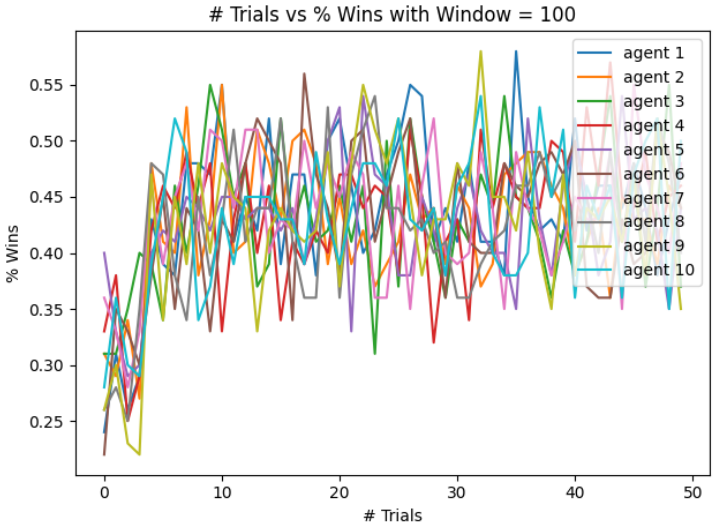
The summary of trials for each agent over 50,000 episodes can be seen below. Here, the impact of the throttled learning rate is more visible.





The summary of trials for each agent over 5000 episodes can be seen below, where the impact of the throttled learning rate is even more visible.





The agent seems to asymptotically converge to ~44% win rate after ~5000 episodes, regardless if 5000, 50,000, or 500,000 episodes occurred. Without the throttled learning rate, my win rate was on the order of ~31-34%.