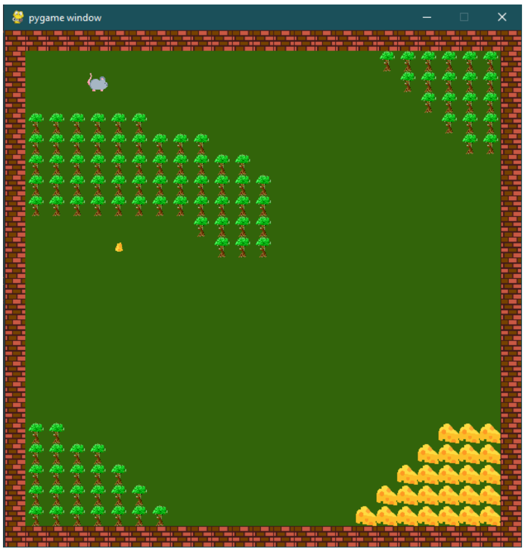
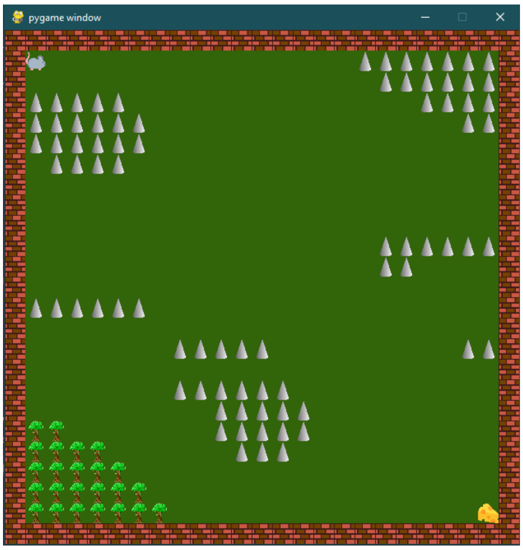
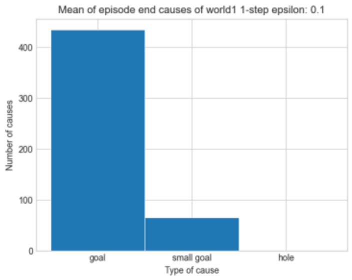
World1 is designed in a way that the agent is positioned close to a small reward state, and there is a high reward state further away.

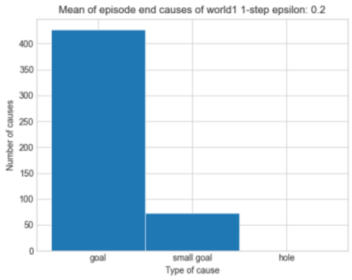
World2 is designed like a labyrinth, touching the spikes results in a penalty and ends the episode.

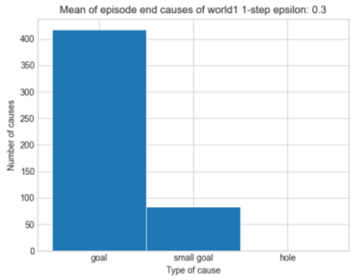


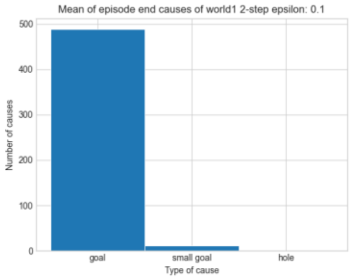


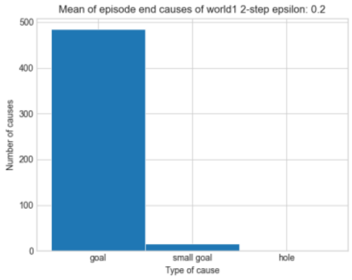
For world1 higher step count and lower epsilon results in better prioritization of the later higher reward.

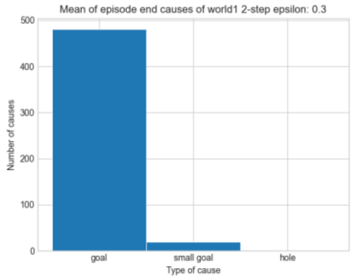


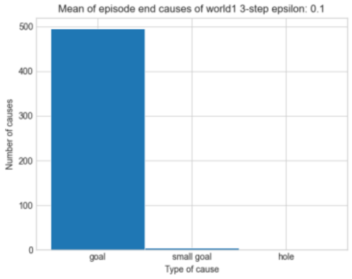


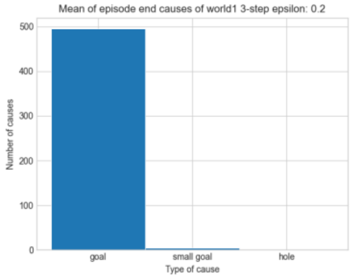


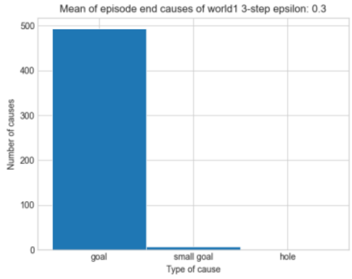




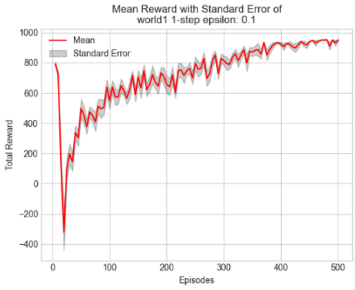


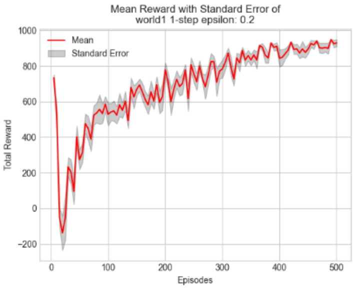


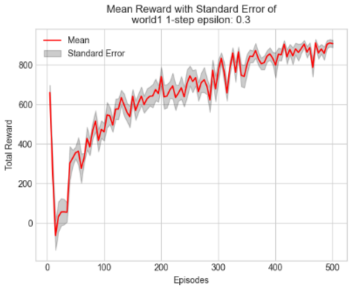


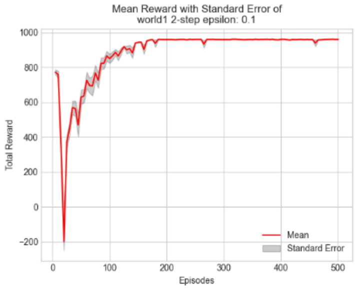


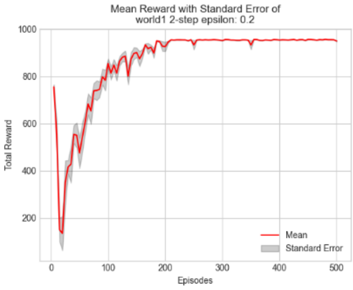
With more N-steps the agent learns much faster, and with higher epsilon causes higher standard error, and more spikes in the plot, due to random actions.

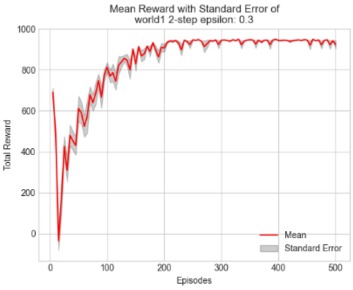


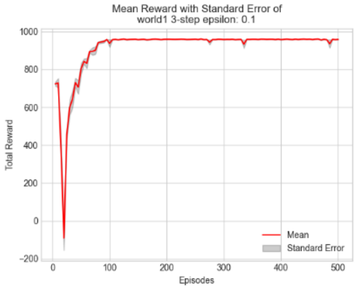


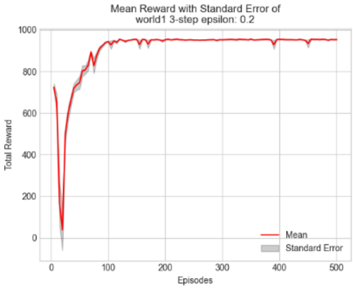


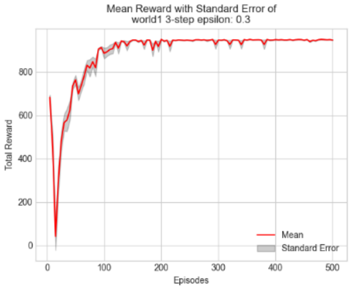




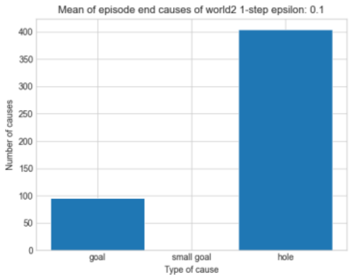


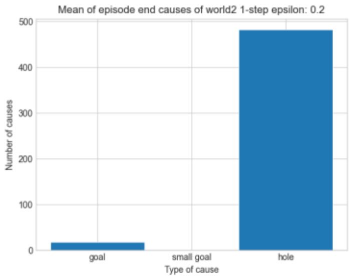


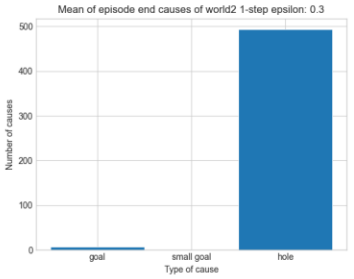


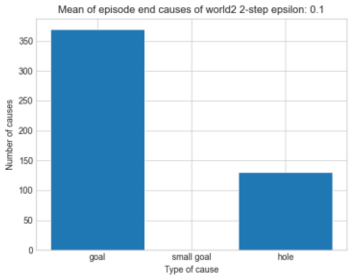


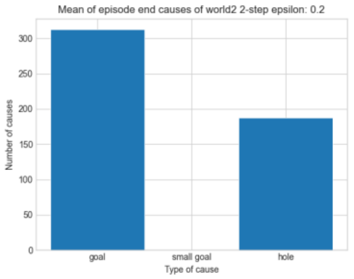
For world2, a higher step count and lower epsilon greatly increases the agents success of finding the goal.

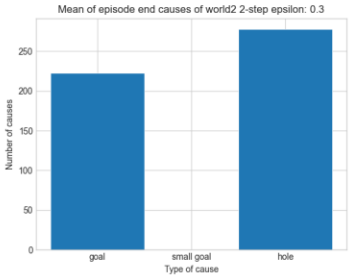


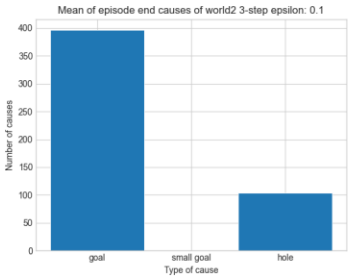


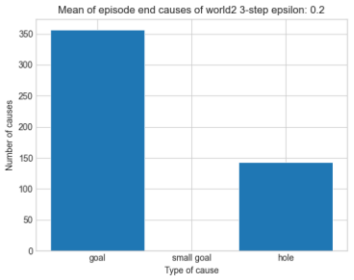


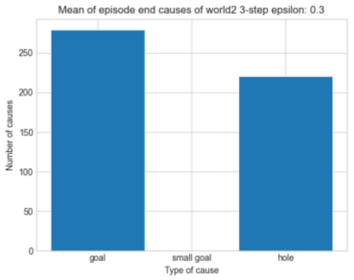






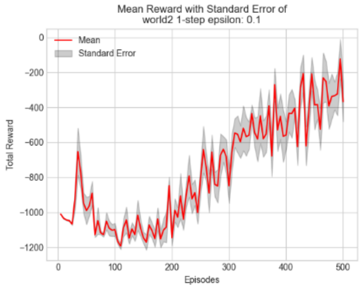


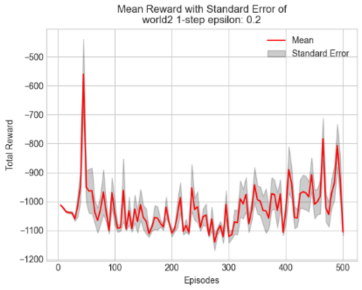


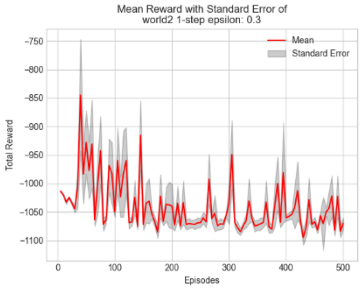


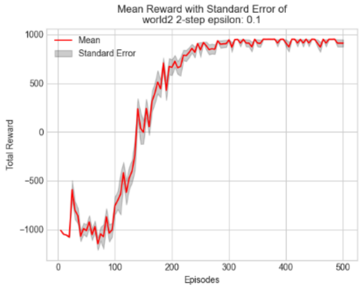
Similarly to world1, the agent learns much faster with higher step count, and lower epsilon.

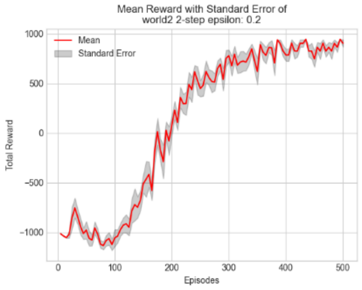
Higher epsilon makes the plot spikier due to random actions

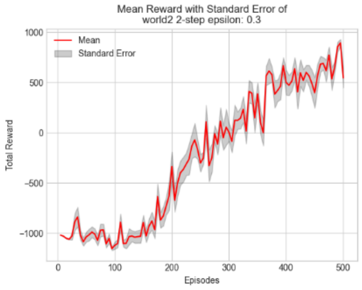

Mean Reward with Standard Error of
world2 1-step epsilon: 0.1
Mean
Standard Error
Total Reward
Episodes

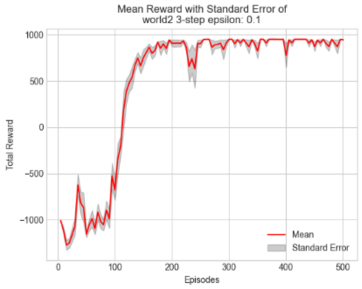

Mean Reward with Standard Error of
world2 1-step epsilon: 0.2
Mean
Standard Error
Total Reward
Episodes

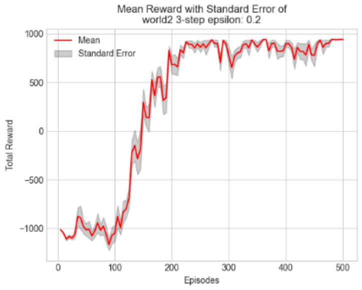

Mean Reward with Standard Error of
world2 1-step epsilon: 0.3
Mean
Standard Error
Total Reward
Episodes

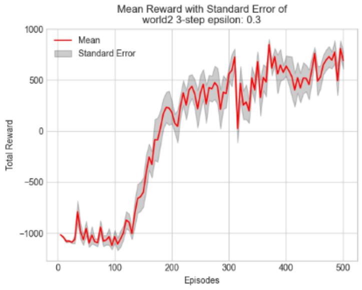

Mean Reward with Standard Error of
world2 2-step epsilon: 0.1
Mean
Standard Error
Total Reward
Episodes


Mean Reward with Standard Error of
world2 2-step epsilon: 0.2
Mean
Standard Error
Total Reward
Episodes


Mean Reward with Standard Error of
world2 2-step epsilon: 0.3
Mean
Standard Error
Total Reward
Episodes


Mean Reward with Standard Error of
world2 3-step epsilon: 0.1
Mean
Standard Error
Total Reward
Episodes


Mean Reward with Standard Error of
world2 3-step epsilon: 0.2
Mean
Standard Error
Total Reward
Episodes


Mean Reward with Standard Error of
world2 3-step epsilon: 0.3
Mean
Standard Error
Total Reward
Episodes