For world1, where the agent should ignore smaller reward to prioritize higher later reward.

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.

For world2, where the agent should solve a labyrinth, but going into the walls of the labyrinth ('o' states) will result in a high penalty and end the episode.

Higher step count and lower epsilon will greatly increases the agents success of finding the goal.