1)
    a. Because the agent does not take any chances, it will not step onto the slippery spot. It finds the slippery spots in the first episode and does not go onto them again. Because there is a whole line of slippery spots it can only move up. This causes it to stay in the same spot the whole time. This means the reward will be zero.
    b. The agent gets a much better reward at the end of each episode. This is because it will take a chance of moving to a bad spot. Because all of the rewards require moving onto these spots it has a better chance of succeeding.
    c. Now the agent is receiving mainly negative rewards at the end of each episode. This is because it is too willing to take a chance on a slippery spot even when it has a package. This will cause it to drop a lot and receive negative reward.
2) A
    a. The results were mainly negative staying in the -10 to -20 range. This is a poor performance of the agent. The agent does not know where the thief is, and because the thief is moving, the agent cannot learn the location of the thief. This causes the agent to run into the thief a lot. Which causes negative reward.
    b. The results dramatically improve. Now the reward is staying above 250 for the most part. This is because the agent knows where the thief is and can avoid him. Early on the agent figures out that moving onto the location with the thief is undeniable and tries to avoid him in later simulations.
    c. .01 was the best rate. I started at .05 and increased it to .1. I found this was a worse result so I tested the middle of the two. This was also worse than .05 so I changed it to .025. This was better so I decided to keep going to .01. I discovered decreasing past .01 yielded worst results. So I tested the .011 and .009 and they were either equal or less than the result for .01 so .01 was the best.
3) The agent initially gets a negative reward. Then within the next 5 episodes it turns to a positive reward. The reward continues to grow until about the 250^{th} episode where if begins to fluctuate at a value around 320. The reward starts negative because the agent is learning will make a lot of bad moves to begin. The agent will then start to know what spots are good and will move to those locations more often. Eventually the agent gets a good estimate of the reward for each spot so it will not change too much.

Initial graph. Reward vs. Episode