1.) With NDISKS=2, two goal states, and gamma=0.9, using passive learning (the user drives the agent via the "console"

(a) how many episodes are required, at a minimum, to achieve a policy which includes the silver path?

1 epsiode is required at minimum to achieve a policy which includes the silver path.

(b) how many additional episodes, at a minimum, are required to make the policy switch to following the golden path?

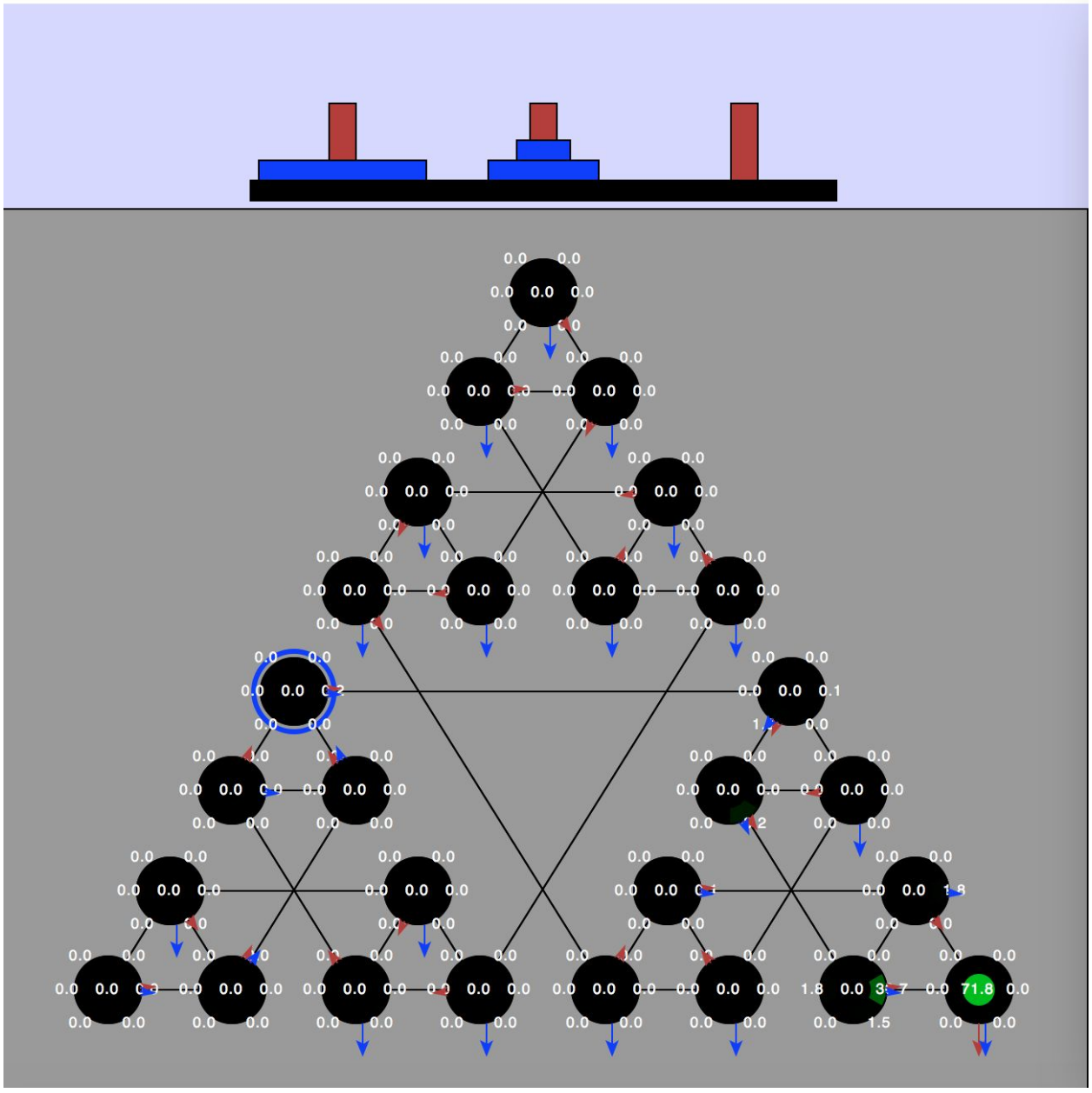
I need an addition 4 episodes at a minimum in order to make the policy switch to the golden path.

2.) With NDISKS=2, one goal state and gamma=0.9, If you use alpha=0.1 and epsilon=0.2 how many episodes of active Q-learning does it take to achieve a policy which matches the optimal policy along the golden path (optimal solution path)?

7 episodes so far. Transition count is now 108. In order to achieve a policy which matches the optimal policy and converges.

3.) First, using NDISKS=3, one goal state, and gamma=0.9, find a way to make your agent learn, starting with all Q values at 0, a policy that includes the golden path. This may require a large number of transitions and episodes, and custom schedules for controlling alpha and epsilon in the Q-Learning. Then describe the details of your strategy, beginning with how your agent adjusted the alpha value. Then describe how it adjusted epsilon. Report the number of episodes used and the number of transitions.

My agent adjusted 12 episodes so far. Transition count is now 1717. My agent adjusted the alpha state by multiplying by 0.98 each time. I do this in order to insure that the agent is learning from all the good Q value paths but does stop learning over time in order to ensure that it stops exploring. My agent does not adjust epsilon because I believe that 0.5 is a good epsilon rate, since it allows for mutation but also to follow the q_learning chosen action. The convergence criterion was satisfied too.


0.0
71.8