The MDP noise in each case is set to 0%.

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

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)?
Usually around 12 episodes. Sometimes more.

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. The learning must be all active, and not involve any user-driven transitions.
I set both Alpha and Epsilon to 0.5 initially. For Alpha value, I update its value by multiplying current alpha with 0.95 after each episode is complete (reaching the goal state). For Epsilon value, I update its value by multiplying current alpha with 0.99 after each episode is complete (reaching the goal state). Because as time goes on, we are more certain of the optimum degree of the policy and don't have to try random actions as at the beginning. However, I implement different decay rates for Alpha and Epsilon because it turns out very effective in finding the golden-path. I make the Epsilon decreases more slowly to avoid first few states, e.g. s0, from trapping in the local maxima. It still has a chance to go to random states after a few episodes have finished.

The result for finding the golden path is between 4 episodes and 31 episodes.