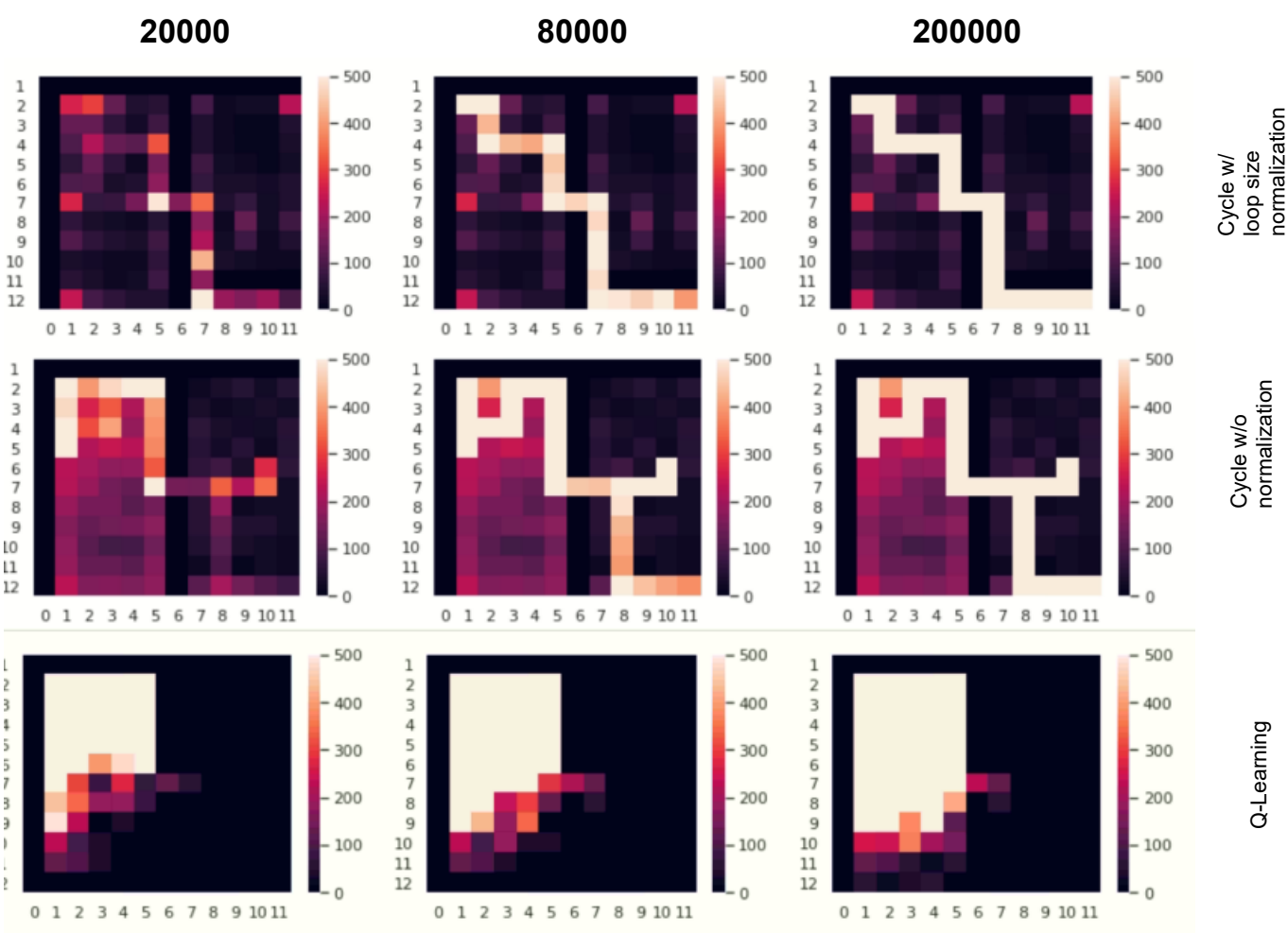
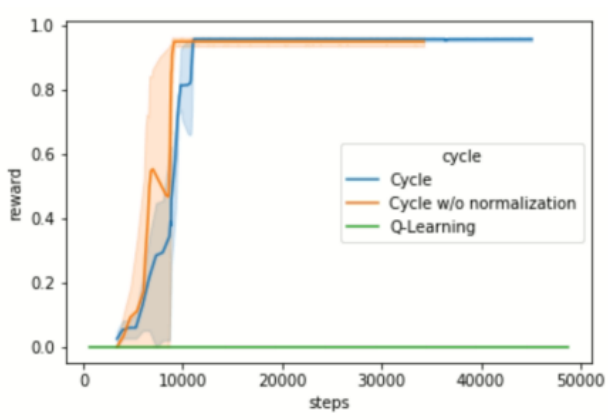
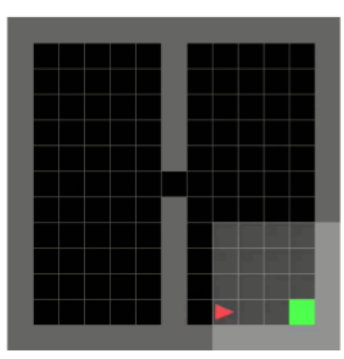
Results on SimpleCrossing 8x8 with one crossing

The cyclophobic agent is able to solve the environment while the Q-learning agent is unable to. Moreover, dividing the fixed cycle penalty -1, by the size of the loop greatly improves exploration.

Note that reward is only given at the green goal state!







Results on SimpleCrossing 24x24 with multiple crossings and rooms.

Again, the cyclophobic agent is able to solve the environment while the Q-learning agent barely able to get out of the second room. Therefore, cycles as inductive bias are extremely efficient for exploration.