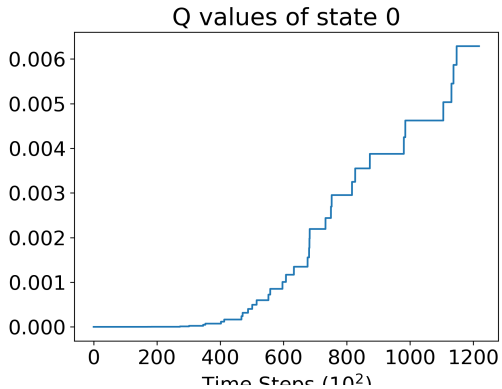
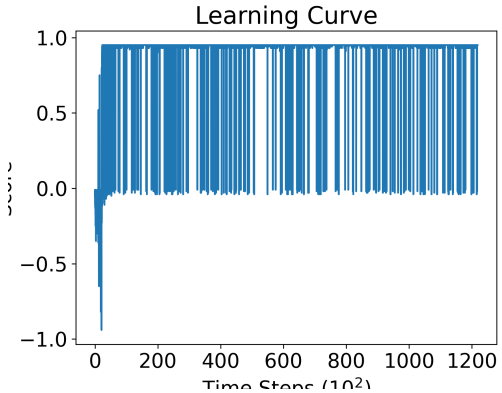
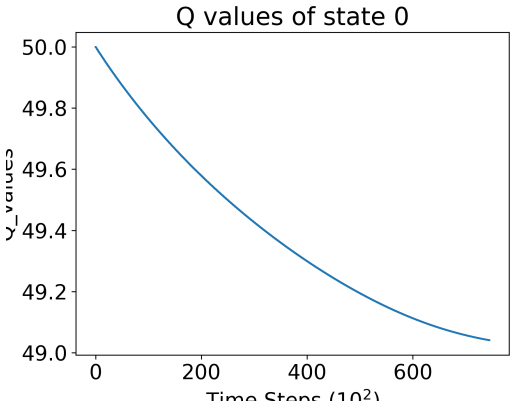
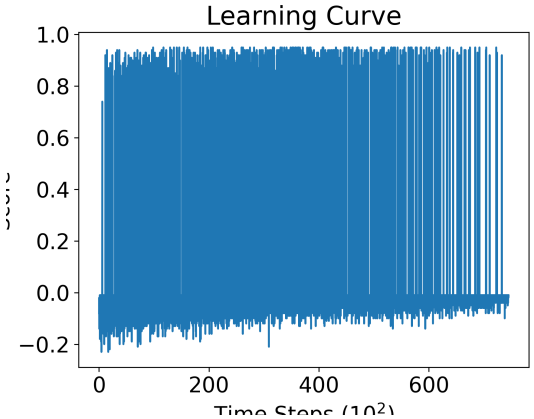
Qtable Initialization 0



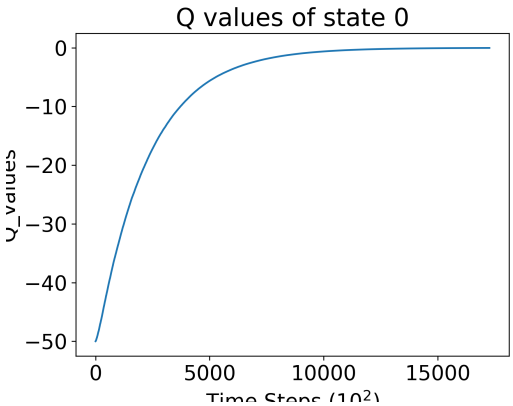


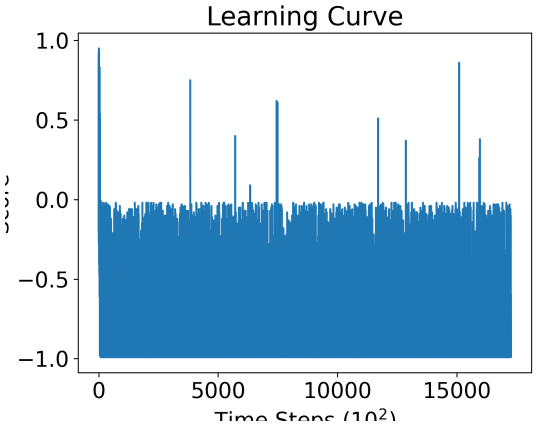
Qtable Initialization 50



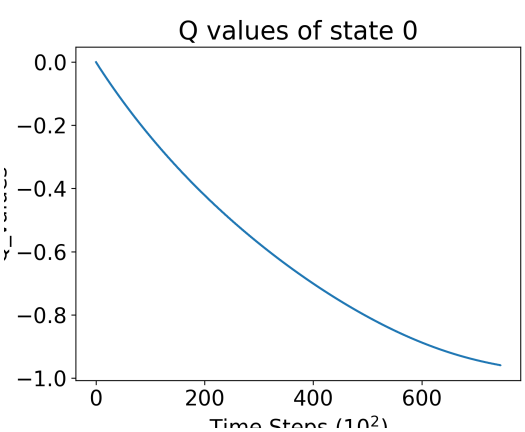


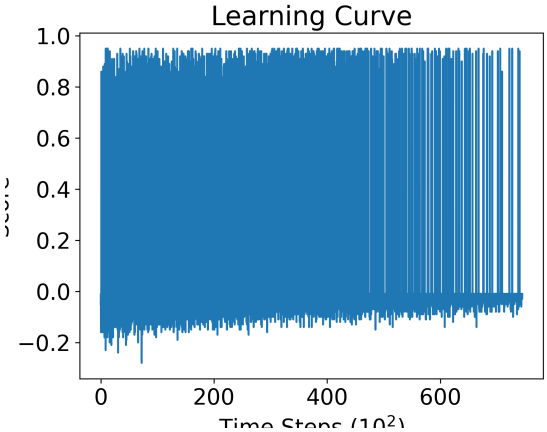
Qtable Initialization -50





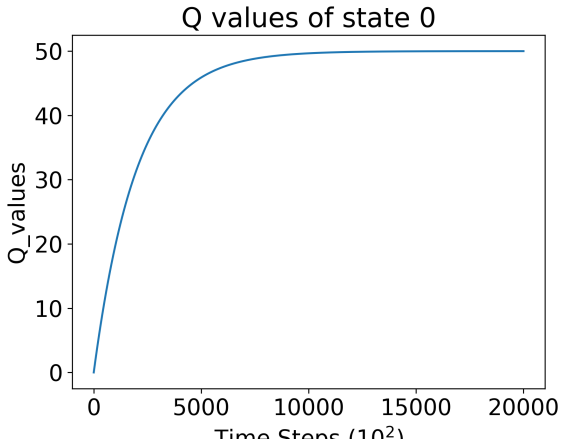
Bias 50 eps 0

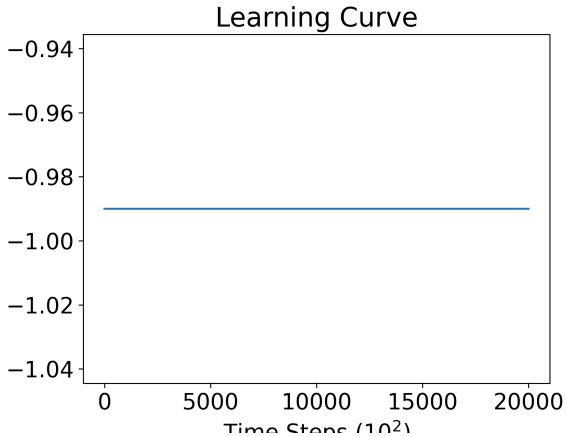





```
Qtable at the end: [[-9.59427927e-01 -9.59214699e-01 -9.59204186e-01 -9.59427927e-01]
 [-4.91706612e-01 -4.91366990e-01 -4.91367544e-01 -4.91568880e-01]
 [-2.72603678e-01 -2.72388714e-01 -2.72460740e-01 -2.72505382e-01]
 [-1.50937105e-01 -1.51018751e-01 -1.51021842e-01 -1.51021842e-01]
 [-4.91073792e-01 -4.91012543e-01 -4.91005588e-01 -4.91279966e-01]
 [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
 [-9.33078723e-02 -9.34070113e-02 -9.33078723e-02 -9.35305768e-02]
 [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
 [-2.34698027e-01 -2.34809206e-01 -2.34706387e-01 -2.34983480e-01]
 [-8.65640696e-02 -8.66163804e-02 -8.64910754e-02 -8.65400794e-02]
 [-4.59915656e-02 -4.58084491e-02 -4.57788970e-02 -4.58045816e-02]
 [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
 [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
 [-2.95955562e-02 -2.97412255e-02 -2.97904775e-02 -2.98243263e-02]
 [-2.50000000e-04 -2.50000000e-04  7.88179306e-02  0.00000000e+00]
 [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]]
```

Bias -50 eps 0

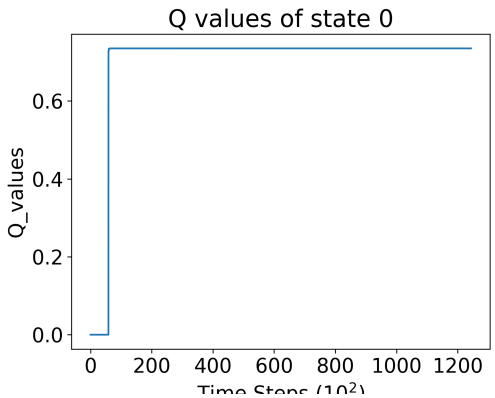


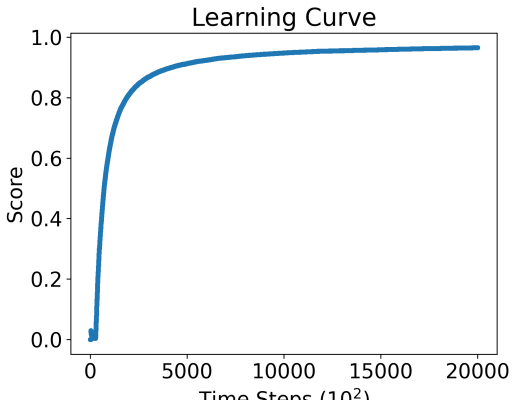


```
[0. 0. 0. 0.]]
Qtable at the end: [[49.99773006  0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]
 [ 0.          0.          0.          0.        ]]
```

Tried with different hyperparameters;
Lr = 0.8
Gamma = 0.95

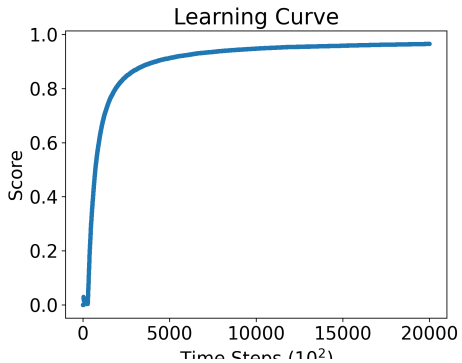
Qtable Initialization 0



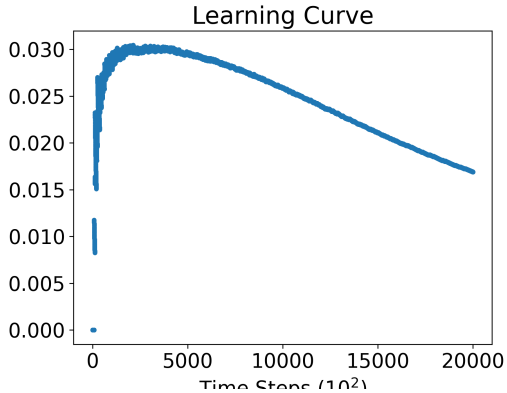


Plotted Learnig Curves using *plt.plot(np.cumsum(score_)/np.arange(1, 20000+1), '.')*
I.e. not added any negative reward

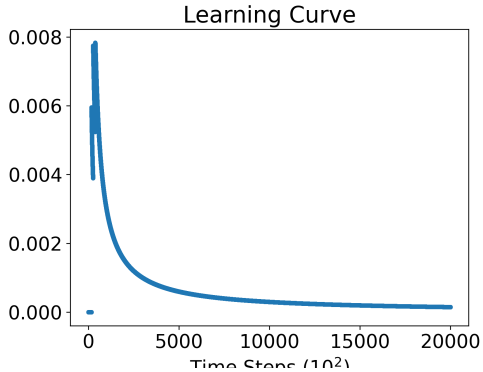
Qtable Initialization 0



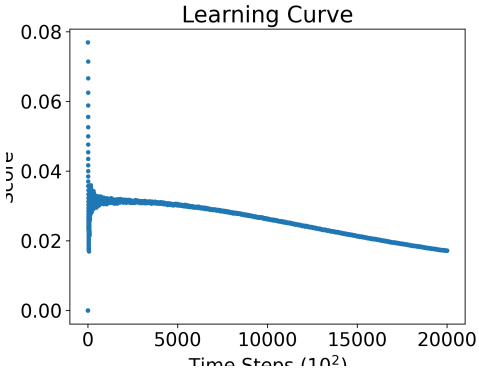
Qtable Initialization 50



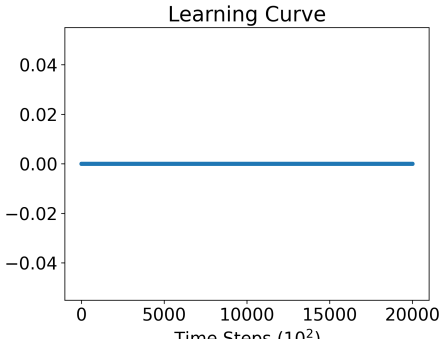
Qtable Initialization -50



Bias 50 eps 0



Bias -50 eps 0


Learning Curve
0.04
0.02
0.00
−0.02
−0.04
0
5000
10000
15000
20000