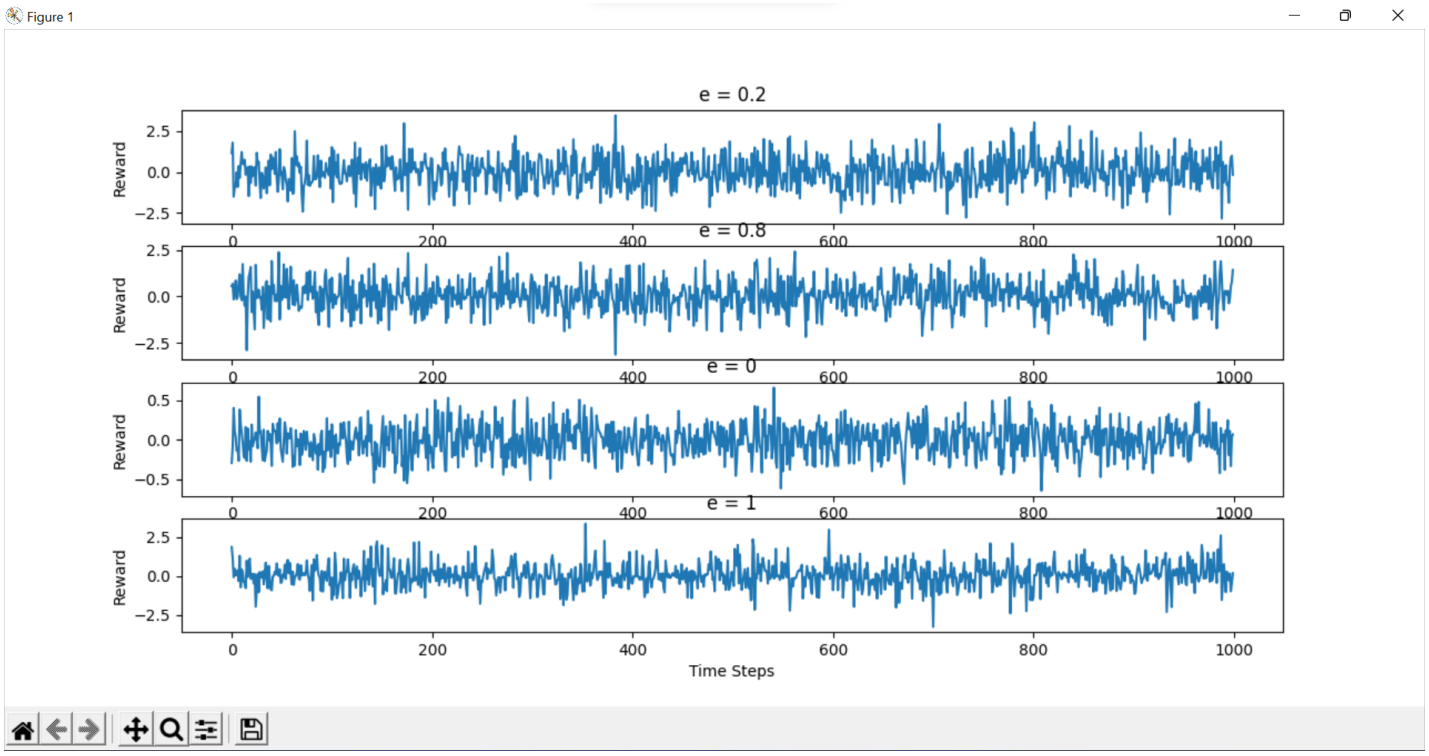
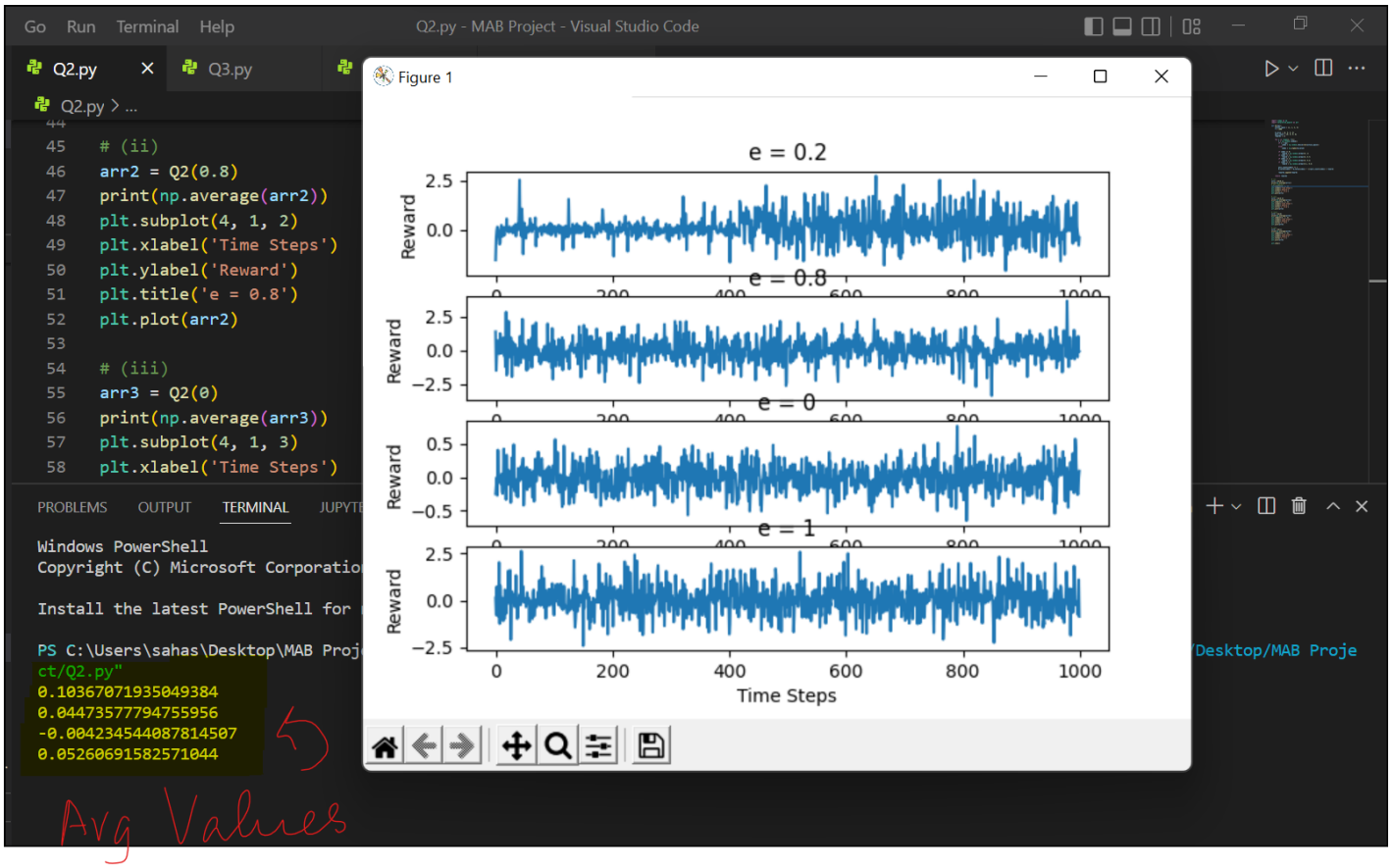
Assignment 1 REINFORCEMENT LEARNING

QUESTION 1:
Q1.pdf

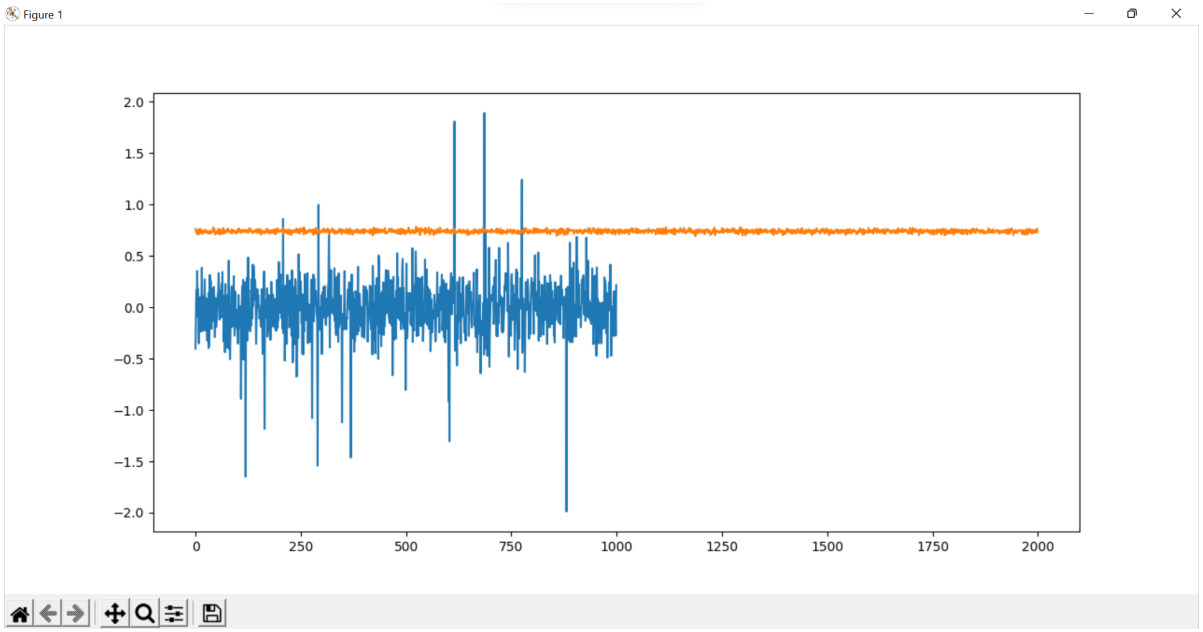
QUESTION 2:
Next page

QUESTION 2:
Q2.py





QUESTION 5:
Q5.py



Exercise 2.8:
Note: I have not got the exact graph as Fig. 2.4, but I can infer the following.

UCB Spikes
We can see a spike at the 11th step as the question is of 10 Arms.
So, after each arm is once iterated over, or each arm is once initialized, we now have the option to choose the best (most optimal) arm at the 11th episode.

Now at subsequent pulls, we see a fall because everything is kind of a reset.
By this, I mean that we now cannot choose the optimal reward and have to start exploring again like we did at the start.

Also, if c is more the spike will be more as we start to explore.