* For implementing a multi-armed bandit algorithm, we will use the ε-greedy action-value method. This method stabilises the exploration and exploitation by choosing the action with the highest estimated value which is exploitation which has a probability of 1 - ε and a random action which is exploitation has been chosen which has a probability of ε.
* The implementation of the ε-greedy action-value method in Python has been detailed in the code file Implementation.py
* Using the MultiArmedBandit class, we can simulate a multi-armed bandit problem with 10 arms and ε = 0.1, as in Figure 2.1 of the Sutton & Barto book. This has been illustrated in the code file Illustration.py
* This will output a list of estimated action values for each arm, which can be visualised as a histogram.
* For analysing the results, we will plot the estimated action values for each arm according to the time. The average reward per play can also be calculated and compared to the optimal reward. In this case, it is 0.5. The results of the visualisation for a 10-armed bandit with ε = 0.1 and 2000 plays have been explained in Visualisation\_example file.