Q2). We can see oscillations and spikes in the early part of the curve. It is due to the fact that $Q_1(a)$ is set to 5 for all a. Since q_* (True value) is selected from normal distribution with mean 0 and variance 1, an initial estimate of +5 is very optimistic and since the reward is also from a distribution with mean $q^*(a)$ and variance 1, it is less than the starting estimates. Since the desired reward is not found for a particular action, all the other actions are explored. This lowers the value of $Q(a)$. If still $Q(a)$ is higher than $q^*(a)$, the same cycle repeats and exploration takes place.

Since for such initial time steps exploration is taking place, we can see some spikes indicating the true action (best) must have taken place among all the actions.