Q1: Please regenerate Fig. 2.4 in the book, and explain why in the UCB curve there is a spike at the 11th time step.

In the first 10 steps, the algorithm traverses all 10 actions in a random order. The average reward is roughly the average reward of all the actions.
On the 11th step, the player has a rough estimation on each action value, and the uncertainty term is small (due to small t) -> if c is small enough, Q dominates in the selection criterion, and the player will pick the greedy action (the best among the first 10 actions). At the 12th and beyond, the uncertainty term corresponding to the unselected actions increases, and the player will take more explorative actions, such that the average reward may dramatically decrease. When c=2 (compared with c=1), more exploration is done, leading to a larger drop of the average reward from the 11th peak to the 12th value.