- n-Armed Bandit Problem   (Chapter 2 in Sutton book)

  - n actions

  - a: action

  - $q(a)$: reward/value for action a. (true reward)

  - Problem: Find the sequence of actions (say over 10K tries) ~~the~~ value associated with this sequence. is maximized

  - $Q_t(a)$: estimate of $q(a)$ at time t.

  - $A_{t+1} = \arg\max\limits_{a} Q_t(a)$; greedy action / exploitation