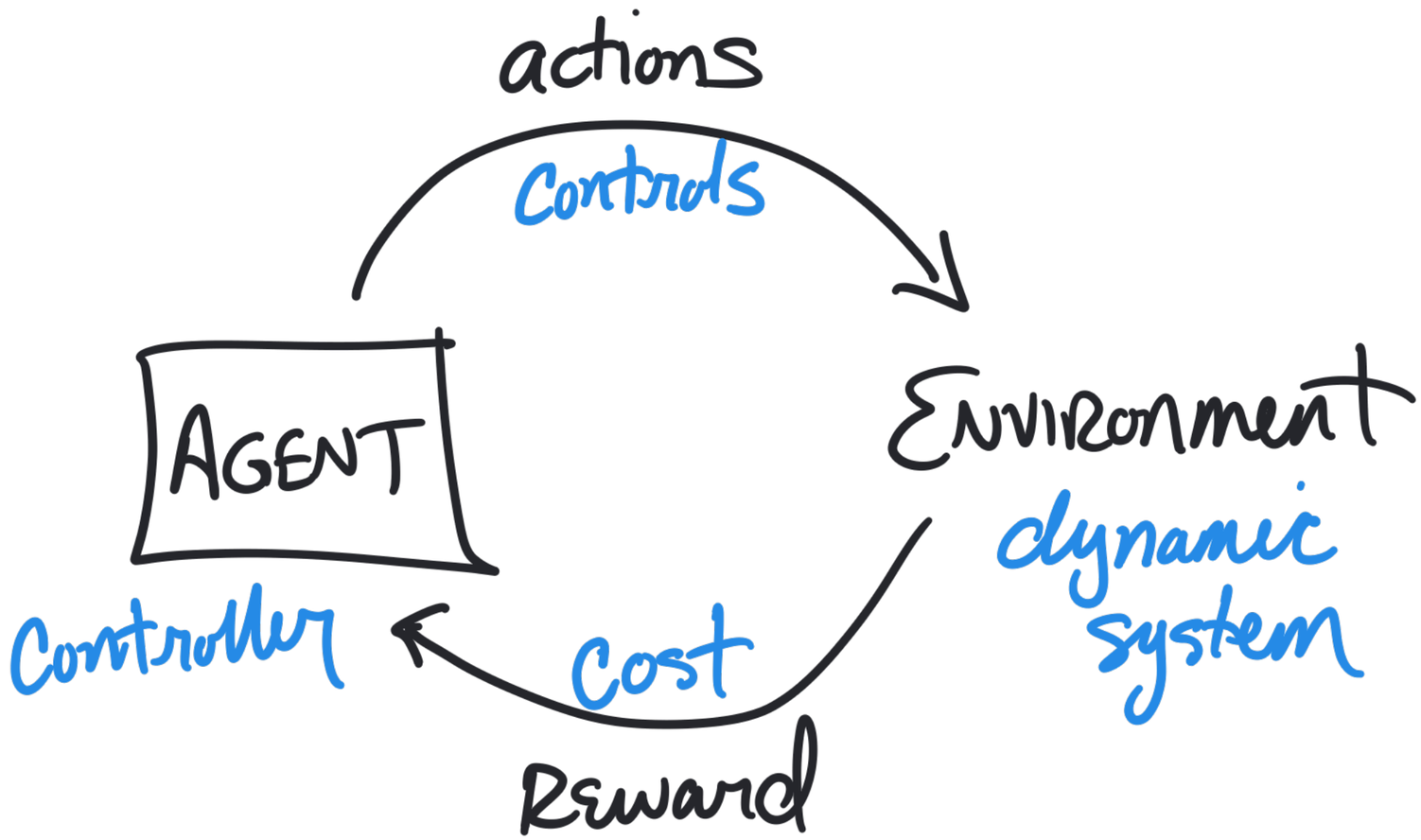
Introduction to Reinforcement Learning

BIOE 498/598 PJ

RL

Computer sci
AI

optimal control



OBJECTIVE: max $\sum$ Rewards

min $\sum$ costs

Objectives

1. Exploration vs. Exploitation
2. Policies
3. Deep RL

Multi-arm (k-arm) Bandit

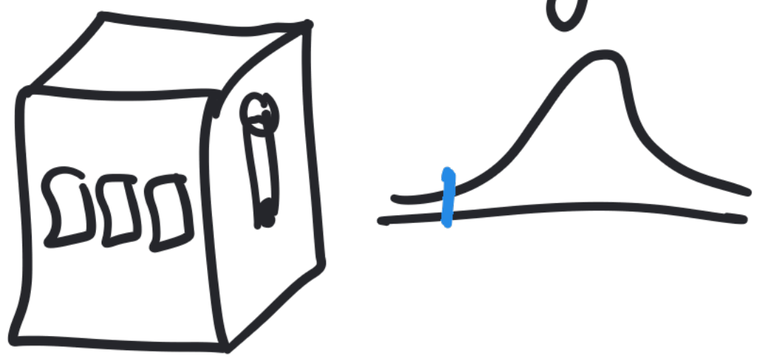
In: 25¢

Payout

Exploration

1. Determine the payout of all machines

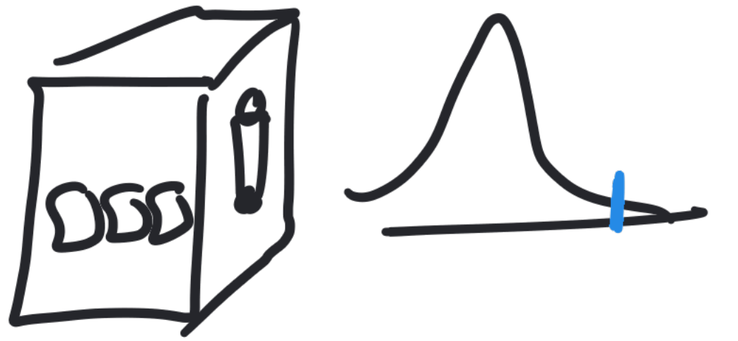
0



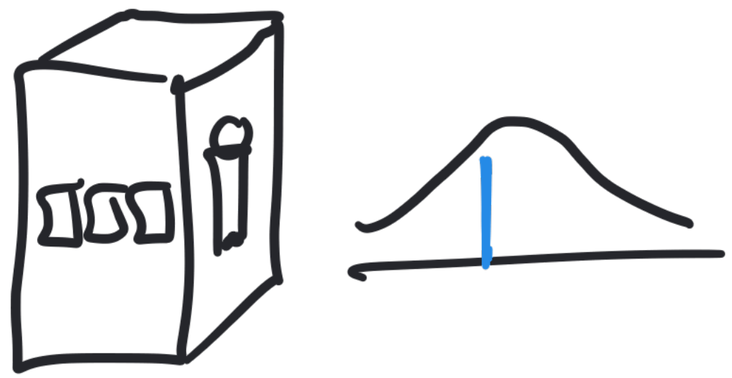
2. Play the best machine.

Exploitation

50¢



25¢



What is ~~the~~ a solution to Exploration vs. Exploitation?

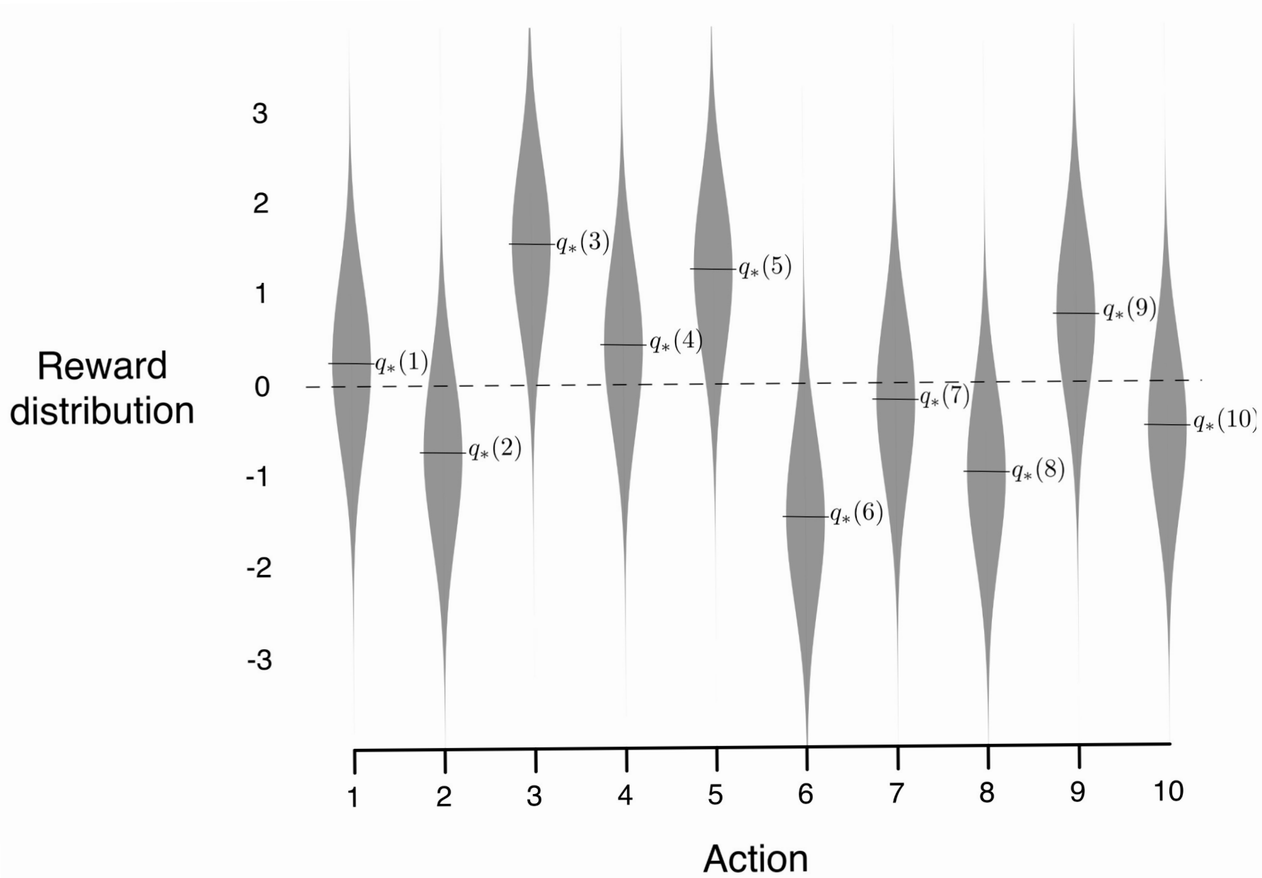
1. Do the optimal action most of the time. Exploit
2. Some of the time, do random actions. Explore

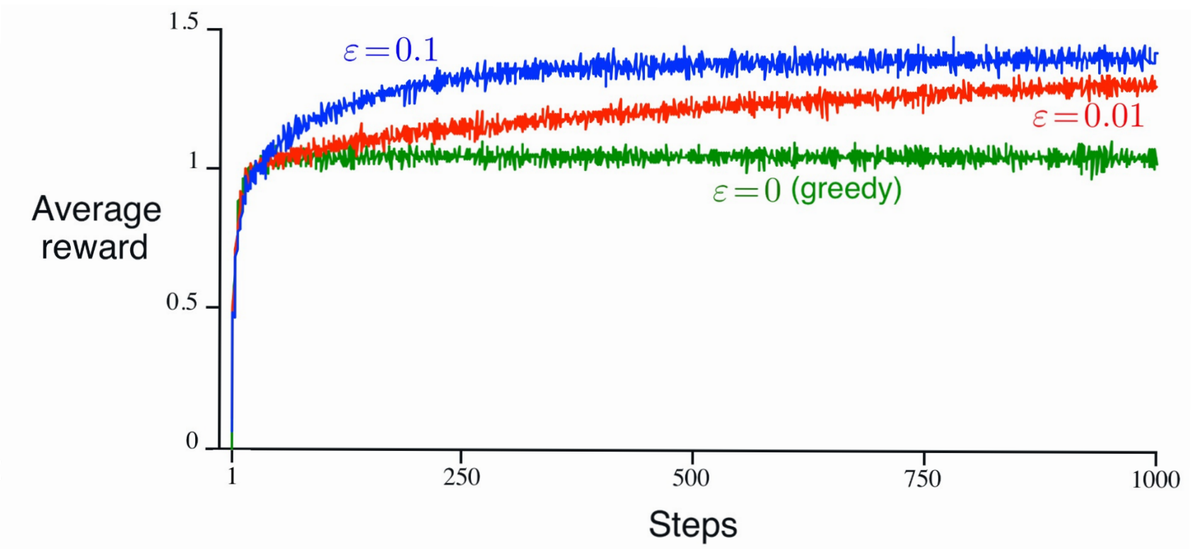
ε-greedy policy

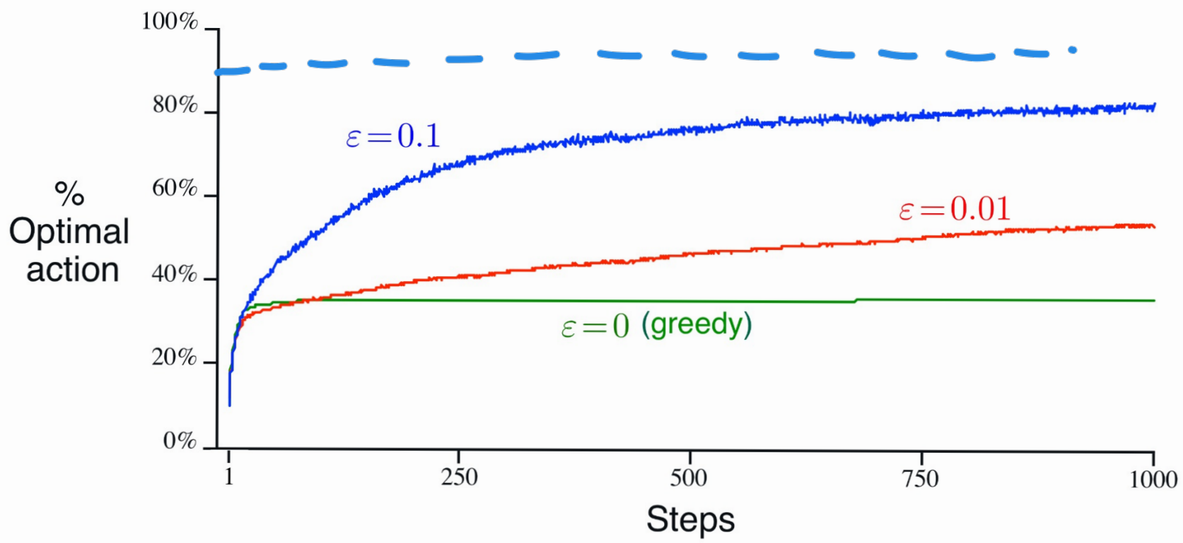
↑ fraction of times where we explore

90% exploit, 10% explore $\longrightarrow$ $\varepsilon = 0.1$

ε-greedy policies for 10-arm bandits







Why does this look familiar?

Explore — Resolution III : Screening, lots of factors
Confounded effects

Exploit — RSM / CCD : Very intensive, few factors
clean, predictive