

Average reward over first
1000 steps

Epsilon-Greedy Gradient Bandit Greedy (optimistic initialization) UCB

