Reward Trajectory for Needle In Haystack Low inst 0.55 -Instantaneous Mean Reward Algorithm ThompsonSampling 0.53 -0 2000 6000 4000 time