Thompson Sampling vs Bayes Adaptive RL
Arms: 2
prior(q): 2, 6
prior(p): 1, 1