

RL non-batched experiments -  $q_a^\pi$ 's: near one,  $\alpha$ : 0.001

