

RL non-batched experiments -  $q_a^\pi$ 's: between zero and one,  $\alpha$ : 0.1

