Averages over 100 runs Training error per episode Reward per episode Episode length 0.25 ┌ 300 1.0  $\operatorname{crm}$ 0.20 250 0.80.15 200 0.6 0.10 150 0.40.05 100 0.250 0.00 0.0 200 400 600 800 1000 200 400 600 800 1000 200 400 600 800 1000