REINFORCE L-T vs. DQN L-T
$\delta d = 0.2$ $\epsilon d = 0.2$

REINFORCE L-T
DQN L-T

performance

context inference error