REINFORCE P-T vs. DQN P-T
$\delta d = 0.1 \quad \epsilon d = 0.1$

performance

REINFORCE P-T
DQN P-T

context inference error