REINFORCE P-T vs. DQN P-T
$\delta d = 0.4$ $\epsilon d = 0.4$

- REINFORCE P-T
- DQN P-T

performance

context inference error