REINFORCE P-T vs. DQN P-T
$\delta d = 0.4$ $\epsilon d = 0.2$

performance vs. context inference error

- REINFORCE P-T
- DQN P-T