

**Mean  $\pm 1\sigma$  Cumulative Reward vs. Episode**  
(first 2 episodes skipped, averaged over seeds)

Env/Rewards

Target Env/Rewards

