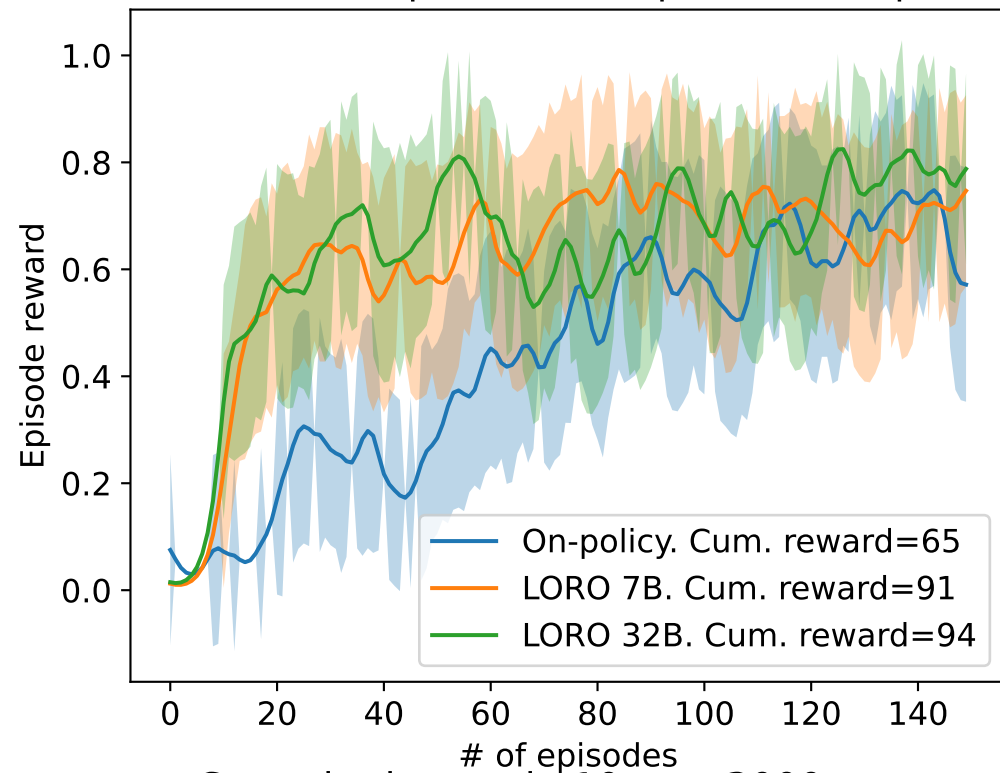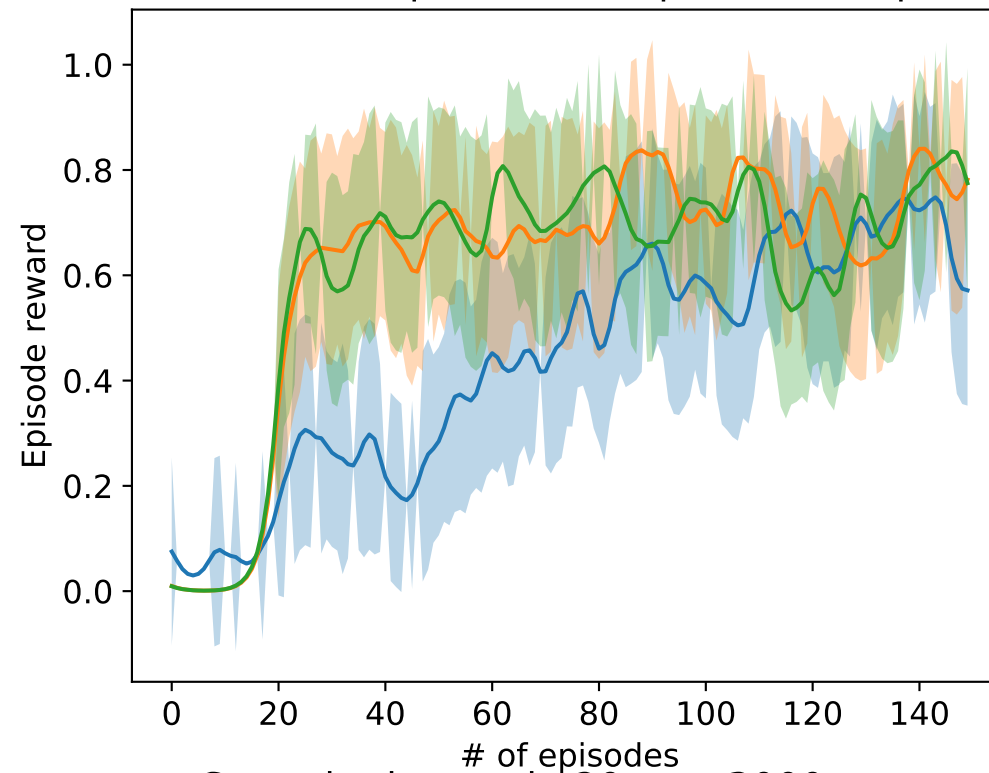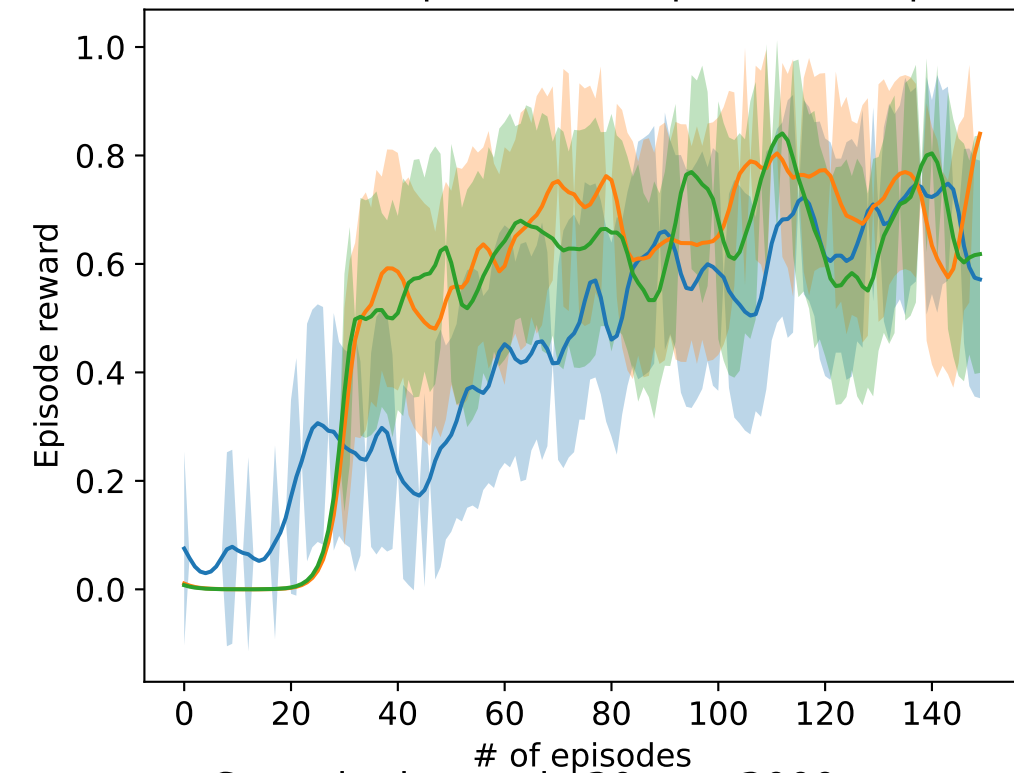**Smoothed pretrain 10 eps, 1000 steps**
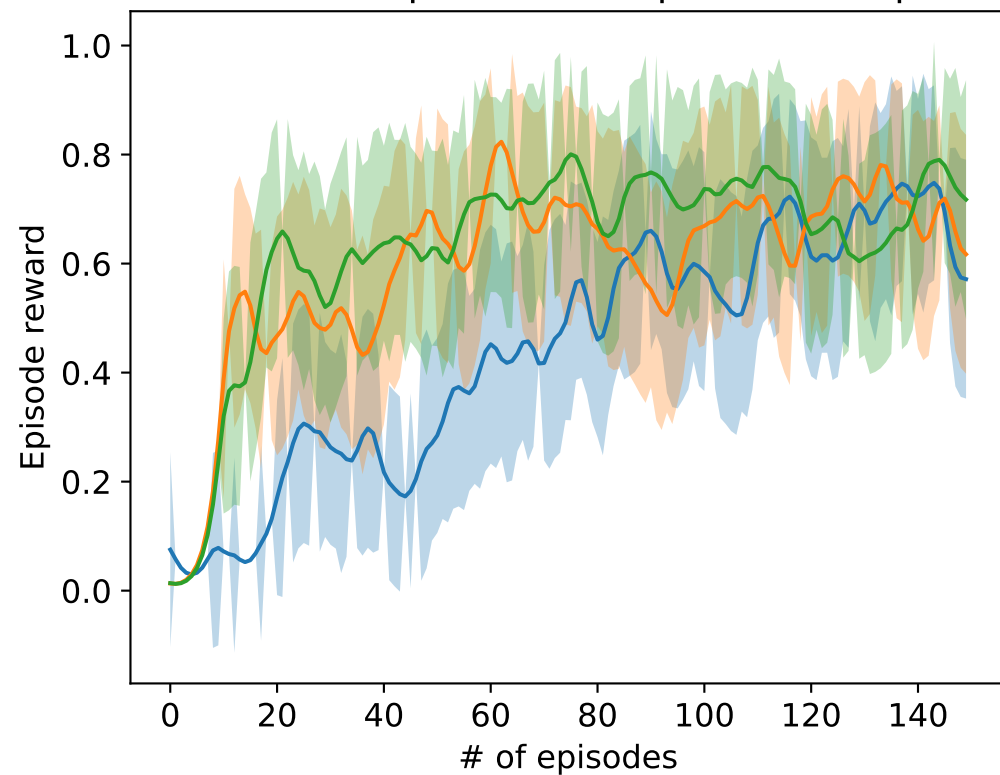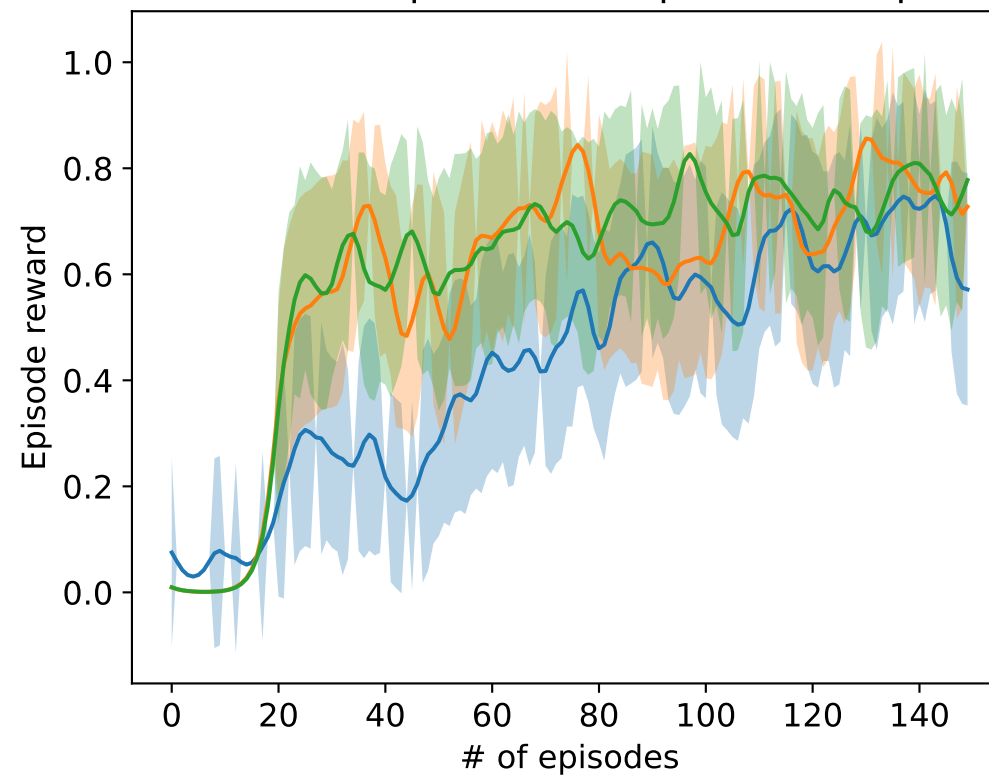
Smoothed pretrain 20 eps, 1000 steps

Smoothed pretrain 30 eps, 1000 steps

Smoothed pretrain 10 eps, 3000 steps

Smoothed pretrain 20 eps, 3000 steps

Smoothed pretrain 30 eps, 3000 steps

Legend:
- On-policy. Cum. reward=65
- LORO 7B. Cum. reward=91
- LORO 32B. Cum. reward=94

Axis labels: Episode reward (y-axis), # of episodes (x-axis)