

Average  
Reward  
(30 runs)

R-Learning

Differential Q Learning

RVI Q-Learning

Timesteps

