MountainCar, ppo2, observation: mean reward