Prisoners Dilemma Evaluation Seed Episode Reward Mean Config Size Small Medium 80 Large Obstacle Policy Reward Mean 20 1 2 4 5 0 3 1e6 Total steps