Prisoners Dilemma 3 Seeds Population-Play Episode Reward Max Config Size Small 250 Medium Large Obstacle 200 Policy Reward Max 50 0 1 2 0 3 4 5 1e6 Total steps