Prisoners Dilemma Evaluation Seed Episode Reward Max Config Size Small Medium Large 200 Obstacle Policy Reward Max 50 1 5 0 2 3 4 1e6 Total steps