Prisoners Dilemma 5 Seed Self-Play Episode Reward Max Config Size 250 Small Medium Large Obstacle 200 Policy Reward Max 00 01 50 0 1 0 2 3 4 5 1e6 Total steps