Prisoners Dilemma 5 Seed Episode Reward Mean Config Size 100 Small Medium Large 80 Policy Reward Mean 60 40 20 0 1 2 3 5 4 1e6 Total steps