Pure Coordination 5 Seed Self-Play Episode Reward Max Config Size 70 Small Medium Large 60 Obstacle 50 Policy Reward Max 20 10 0 1 2 5 4 0 1e6 Total steps