

GAMES WITH ALGORITHMIC AGENTS

GIACOMO MANTEGAZZA

We develop a theoretical model to study strategic interactions between adaptive learning algorithms. Applying continuous-time techniques, we first define a tractable general class of algorithms that encompasses well-known reinforcement learning procedures. We then uncover the mechanism responsible for collusion between algorithms documented by recent experimental evidence. We show that inadvertent coupling between the algorithms' estimates leads to periodic coordination on actions that are more profitable than static Nash equilibria, and we provide a condition under which convergence to Nash equilibrium is guaranteed. Finally, we apply these insights to the design of strategy-proof mechanisms robust to the presence of learning agents: robustness relies on ex-post feedback provision, which enables counterfactual evaluations, thus eliminating the risk of inadvertent coordination