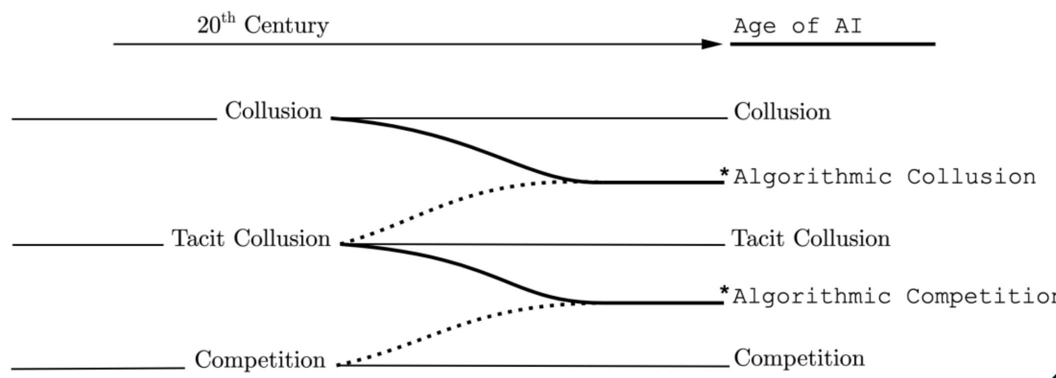
Multi-Agent Learning in the Age of Algorithmic Collusion, PhD Thesis, *Cesare Carissimo*

Learning Algorithms enter the landscape of agents in economies (figure on left). How will reinforcement learning behave in multi-agent economic settings?

This thesis is divided in three sections:

1. Methods: what simulation setup should we use to study the self-organization of multi-agent learners?
2. Application: how can we intervene to steer independent learning in a multi-agent system?
3. Meta-Analysis: what does it mean for an algorithm designer to compete with another algorithm designer?

Learning algorithms can spontaneously coordinate even when the underlying incentives are misaligned. In some contexts, this is desirable, but not in others.



1. Method

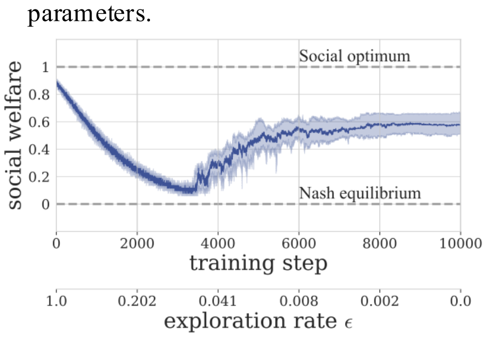
To investigate multi-agent learning dynamics, we use Q-learning, a foundational reinforcement learning model, as the model of learning behavior. We test Q-learning in settings from Game Theory. To develop the method, we started with Braess's Paradox.

The defining characteristics of our setup are:

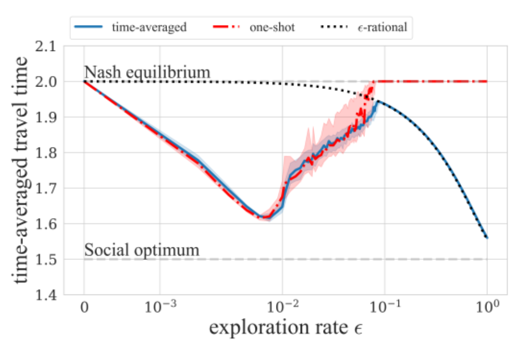
- Independent learning, where agents train in a decentralized manner.
- Continual learning, also known as online learning, where agents must balance exploration and exploitation dynamically.

We argue both of these characterstics are necessary to study the learning dynamics of agents in large multi-agent systems which are not centrally controlled, like traffic with independent drivers, and markets with independent firms and customers.

Our results show spontaneous coordination of independent Q-learners, dependent on learning parameters.



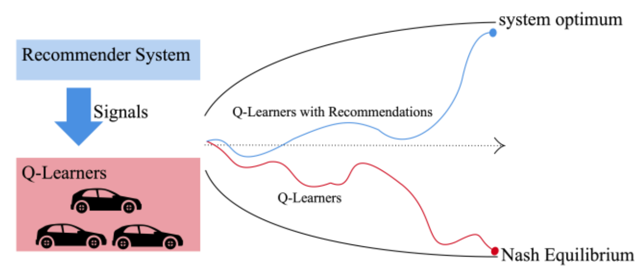
This figure shows the learning dynamics in time of 100 Q-learners in the Braess Paradox when their exploration rate is decayed to 0, and converge above the Pareto inefficient Nash Equilibrium.

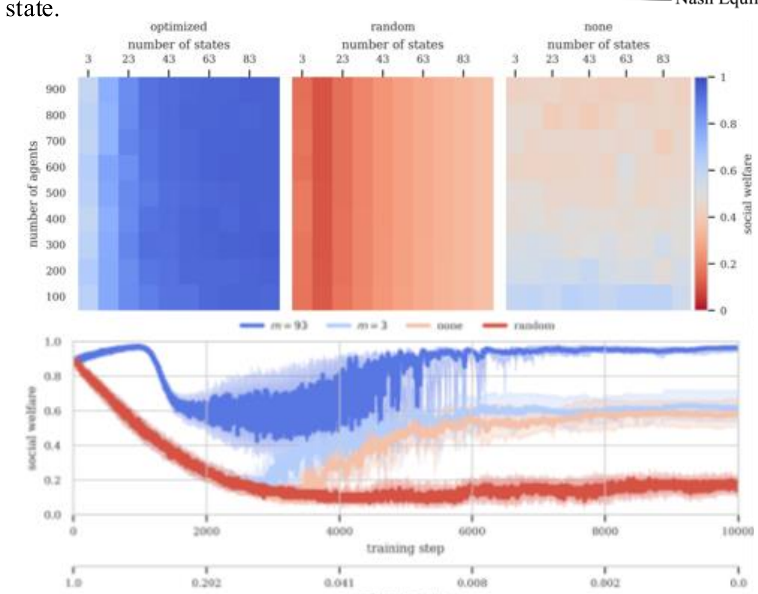


This figure shows the time averaged behaviour of 100 Q-learners, as the exploration rate is kept fixed. We see that small values of the exploration rate can lead to the most socially beneficial results, avoiding the Nash Equilibrium.

2. Application

We extend our model to include recommendations. Q-learners receive recommendations as signals, affecting their state.

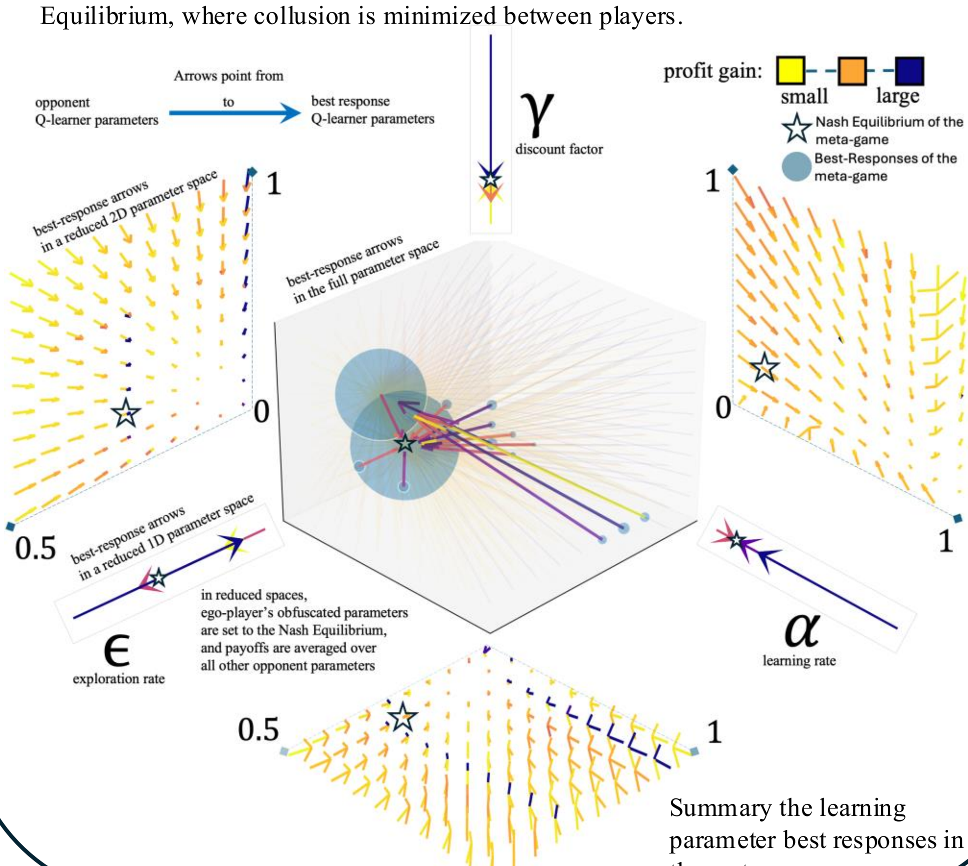




We designed a custom recommender algorithm that could steer the learning dynamics towards the social optimum, taking advantage of the coordinated dynamics we discovered in the previous work.

3. Meta-Analysis

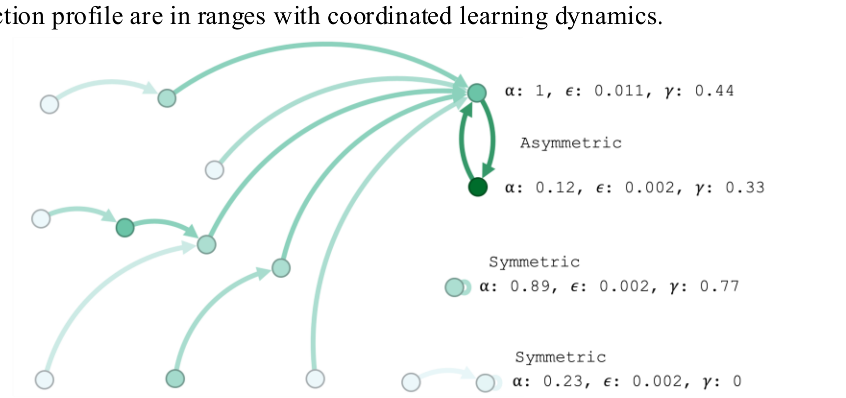
We define the Meta-Game played by Q-learning algorithm designers that select the learning parameters of an algorithm that competes with another algorithm. We analyse the Meta-Game of the Bertrand Duopoly, and find a unique pure symmetric Nash Equilibrium, where collusion is minimized between players.



Summary the learning parameter best responses in the meta-game.

Addendum

We apply our meta-analysis to 100 algorithm designers in Braess's Paradox, and find that the best response parameters for a unilateral deviation from a symmetric action profile are in ranges with coordinated learning dynamics.



Contrary to the Bertrand Duopoly, the meta-Nash equilibrium of the Braess Paradox meta-game leads to highly coordinated behaviour.

Conclusion

To correctly determine the possible outcomes of algorithms in markets we must look at the incentives of the designers of algorithms, the meta-game. Our analyses of meta-game equilibria suggest that desirable social outcomes may spontaneously arise from the incentives of learning algorithm designers: a) competition in a market duopoly and b) coordination in a congestion game.

Substantial open questions remain about: algorithms other than Q-learning, algorithms implemented by real world designers, other dilemma games, the ease of reaching the meta-game equilibrium.