Collusion by mistake: does algorithmic sophistication drive supra-competitive profits?

NOVEMBER 21, 2024

Ibrahim Abada
Grenoble Ecole de Management

In collaboration with Xavier Lambin and Nikolay Tchakarov (ESSEC Business School)

A Disclaimer

The opinions expressed in this presentation are those of the authors alone and might not represent the views of GEM or ESSEC.

Context

- The literature consistently reports that simple reinforcement learning algorithms systematically reach seemingly collusive outcomes.

- The drivers of cooperation are being investigated: sophisticated punishment strategies to sustain the cartel (Calvano et al. [2002b]), numerical biases (cooperation bias Banchio and Mantegazza [2023]), correlated learning (Lambin [2024]), etc.

- Often simple Q-learning algorithms are tested with an implicit asusmption: "*The enhanced sophistication of learning algorithms makes it more likely that AI systems will discover profit-enhancing collusive pricing rules*" in Calvano et al. [2020a].

The research questions

- Is algorithmic collusion always the aftermath of sophisticated punishment schemes deployed by the algorithms?
  - **We develop a simple theoretical illustration of competing Q-learning algorithms in a basic social dilemma and show that (seeming) collusion can be an aftermath of imperfect exploration.**
  - We validate our results via simulations in a market environment.

- Does algorithmic sophistication make seeming collusion easier?
  - We simulate the competition between more sophisticated algos (Deep Learning Actor-Critic networks, Reinforce, and Exp3) and demonstrate that seeming collusion disappears.
  - When agents are endowed with the possibility to choose the level of sophistication of the algorithms they use to operate, seeming collusion is not the unique equilibrium.
  - This result shows that the very choice of overly simple algorithms by market agents might be a sign of tacit collusion.

Literature overview

General issues related to algorithms:

- Algorithmic trading: Chaboud et al. [2014], Hendershott et al. [2011]
- Biased recommendations: Bourreau and Gaudin [2018], Fleder and Hosanagar [2009], Calvano et al. [2022]

Algorithmic cooperation:

- Simulations in synthetic environments: Waltman and Kaymak [2008], Klein [2020], Calvano et al. [2020a & b], Hettich [2021], Abada and Lambin [2023], etc.
- Empirical work: Brown and Mackay [2020], Assad et al. [2020]
- Drivers of cooperation are debated: Banchio and Mantegazza [2023], den Boer et al. [2022], Lambin and Epivent [2022], Asker et al. [2022], etc.

Grey literature actively looks for regulatory solutions:

- OECD [2017], ACB [2019], EC [2017]…

The theoretical illustration and collusion by mistake

A numerical application with Q-learning

AI over-sophistication can reduce seeming collusion

The game of the technological choice

Conclusion

The setting

- **Objective**: develop a (basic) theoretical illustration to highlight that imperfect learning can drive seeming collusion.
- **Environment**: A prisoner dilemma framework. Two possible actions: Cooperate (C) or Compete/Defect (D).
- **AI**: Two stylized stateless Q-learning (cannot deploy reward/punishment).
- **Exploration**: The general case where exploration decreases with learning.
- **Technical assumptions**:
  - A mean-field approach
  - Algorithms find it at some point that cooperation outperforms competition in their Q-matrices
  - + reasonable technical assumptions on the learning rates

		Agent 2	
		C	D
Agent 1	C	(α,α)	(σ,ϕ)
	D	(ϕ,σ)	$\boxed{(\beta,\beta)}$

Figure 1: Normal-form representation of the game

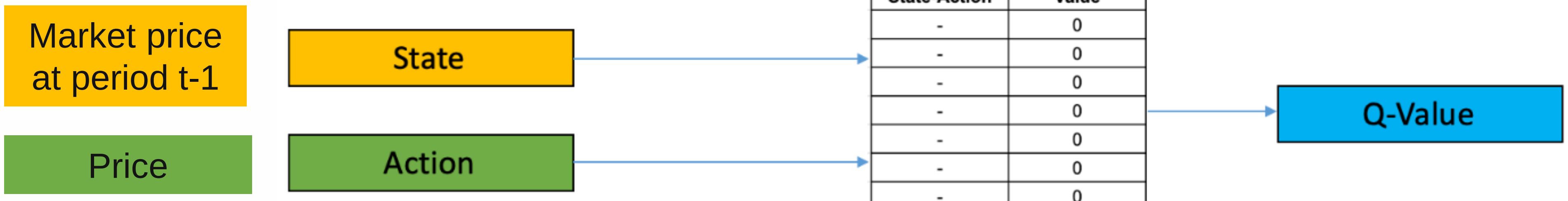
The static Nash equilibrium

Q-learning in a nutshell

Reinforcement learning:

- Interaction with environment generates penalties/rewards
- Model-free
- Balance between exploration (of uncharted territory) and exploitation (of current knowledge)

Q-Learning : value-based **reinforcement learning** algorithm used to find the optimal action-selection policy using a **Q** matrix



Q Table	
State-Action	Value
-	0
-	0
-	0
-	0
-	0
-	0
-	0
-	0
-	0

Q-value : maximum future expected discounted payoff of the agent starting from state s

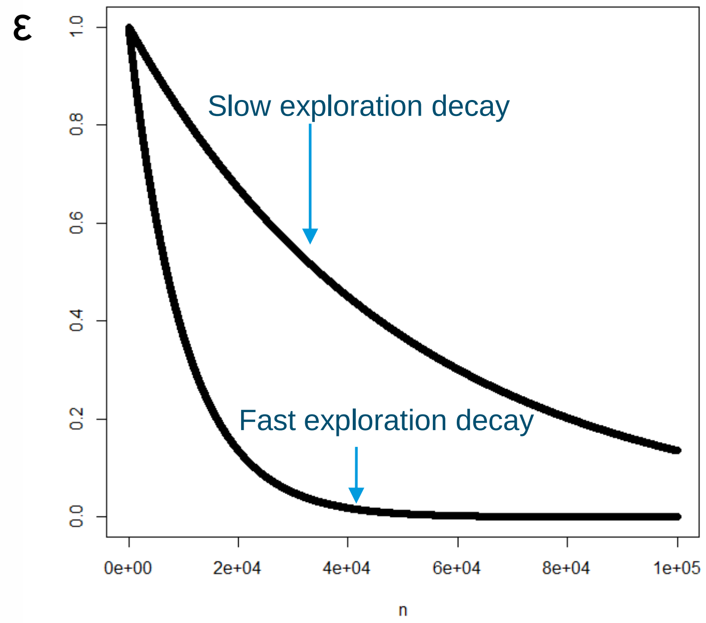
$$Q(s,a) = \pi(s,a) + \delta \max_{a' \in A} \mathbb{E} Q(s'(s,a), a')$$

Q-matrix updating

Q-matrix updating:

Updating, the learning rate

$$
\begin{aligned}
\text{if } s = s_n \text{ and } a = a_n:\quad Q_{n+1}(s_n, a_n) &= (1-\alpha)Q_n(s_n, a_n) + \alpha\Pi(s_n, a_n) \\
&\quad + \delta \max_{a' \in A} Q_n(s_{n+1}, a') \\
\text{otherwise:}\quad Q_{n+1}(s, a) &= Q_n(s, a)
\end{aligned}
$$



Exploration:

- The choice of the action a_n to play at each iteration is the result of a tradeoff between exploration and exploitation.
- Various exploration strategies can be implemented: Boltzmann, **epsilon-greedy**, etc.

The main theoretical results for Q-learning

- If the exploration rate is constant and the learning horizon if infinite, algorithms do not learn to cooperate at convergence.

- Cooperation as an equilibrium can be driven by mistake: *if the exploration rate of the algorithms decreases too rapidly*, the algorithms will never lean to compete.
  - The intuition is that algorithms may be trapped at some point into believing that cooperation yields higher payoffs and as exploration decreases, this belief will be reinforced.

- The latter is a sufficient but not necessary condition!

The numerical setting: from stylized to more realistic algorithms

- A Cournot competition with a linear (and elastic) demand function

- A one period memory (as in Calvano et al. [2020]) with price monitoring

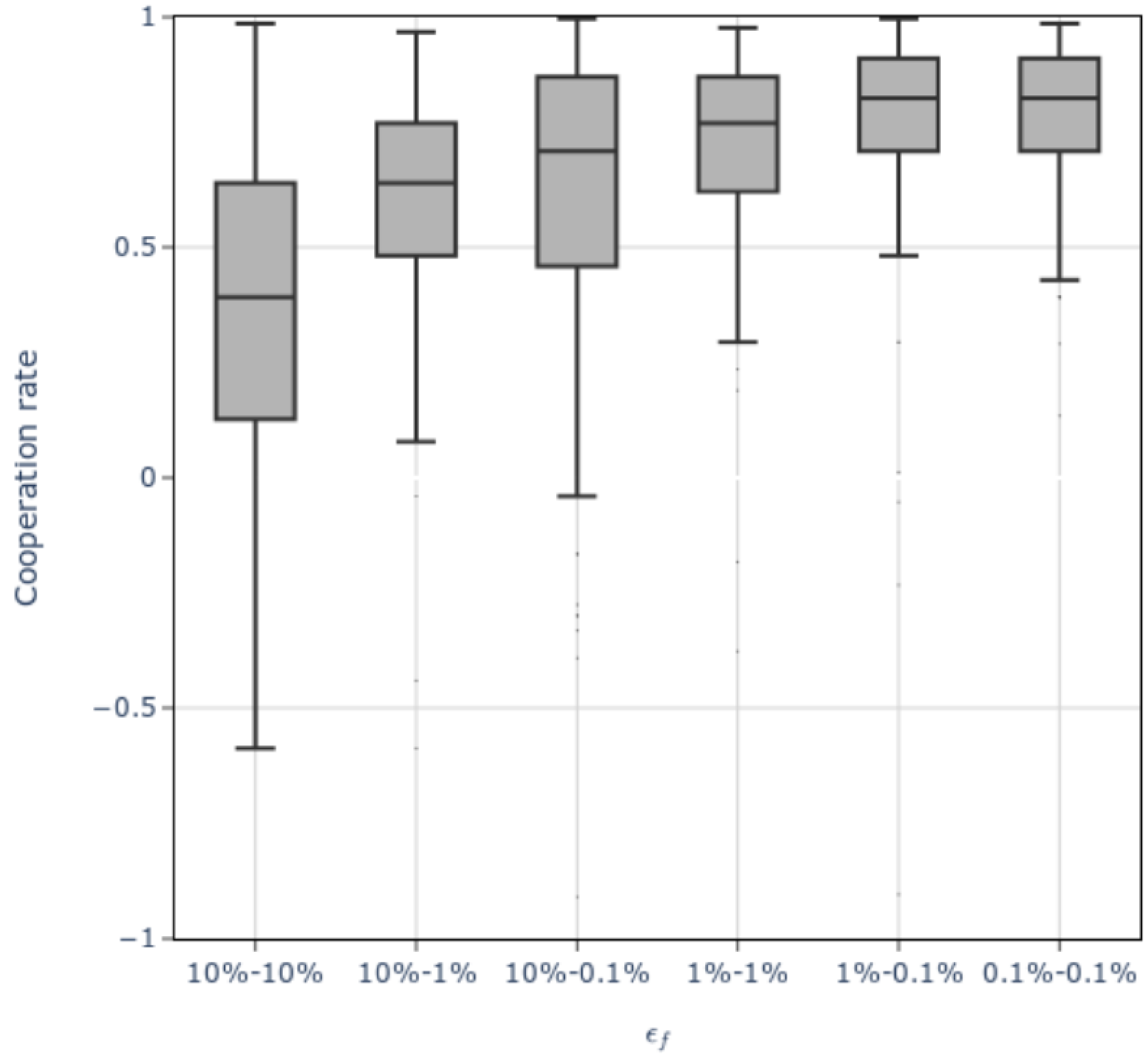
- A measure of the level of seeming collusion: the cooperation rate at convergence

$$\upsilon = \frac{\Pi^{Cartel} - \Pi^{AI}}{\Pi^{Cartel} - \Pi^{Cournot}}$$

- A varying exploration rate of the algos tuned by the final epsilon value (epsilon-greedy).

ϵ_f = 0,1% or 1% or 10%

A more thorough exploration decreases the cooperation rate



Cooperation rate after learning for various duels with Q-learning endowed with either

- parsimonious (ε_f = 0.1%),
- medium (ε_f = 1%),
- or expansive (ε_f = 10%) exploration policy during learning.

The theoretical illustration and collusion by mistake

A numerical application with Q-learning

AI over-sophistication can reduce seeming collusion

The game of the technological choice

Conclusion

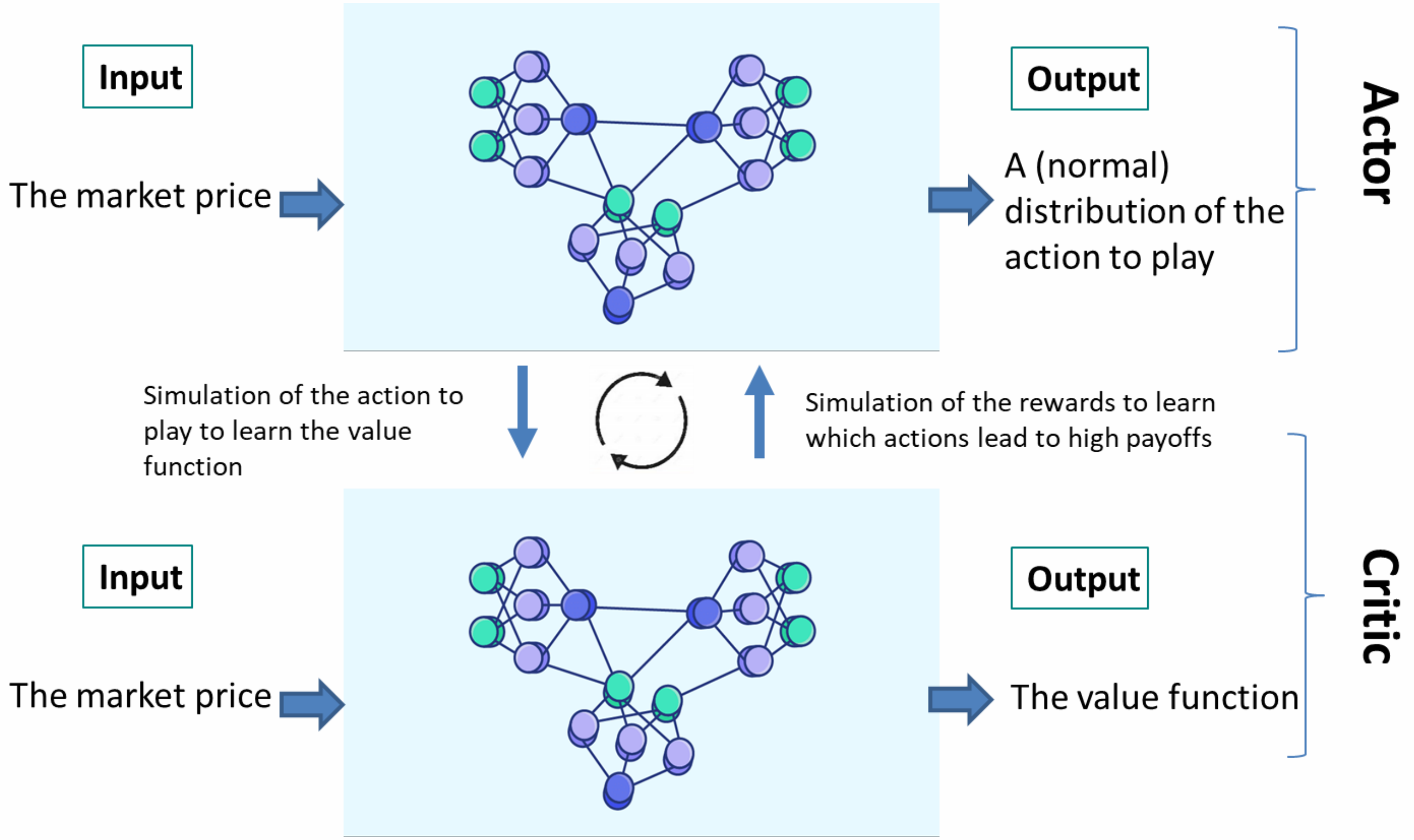
Three other basic Reinforcement learning algorithms

- The Reinforce algorithm (Williams [1992]): a policy-based reinforcement learning with memory.

- Exp3 (Lattimore and Szepesvári [2020]): a policy-based reinforcement learning without memory (stateless). Recently used in den Boer et al. [2022] to investigate the impact on cooperation.

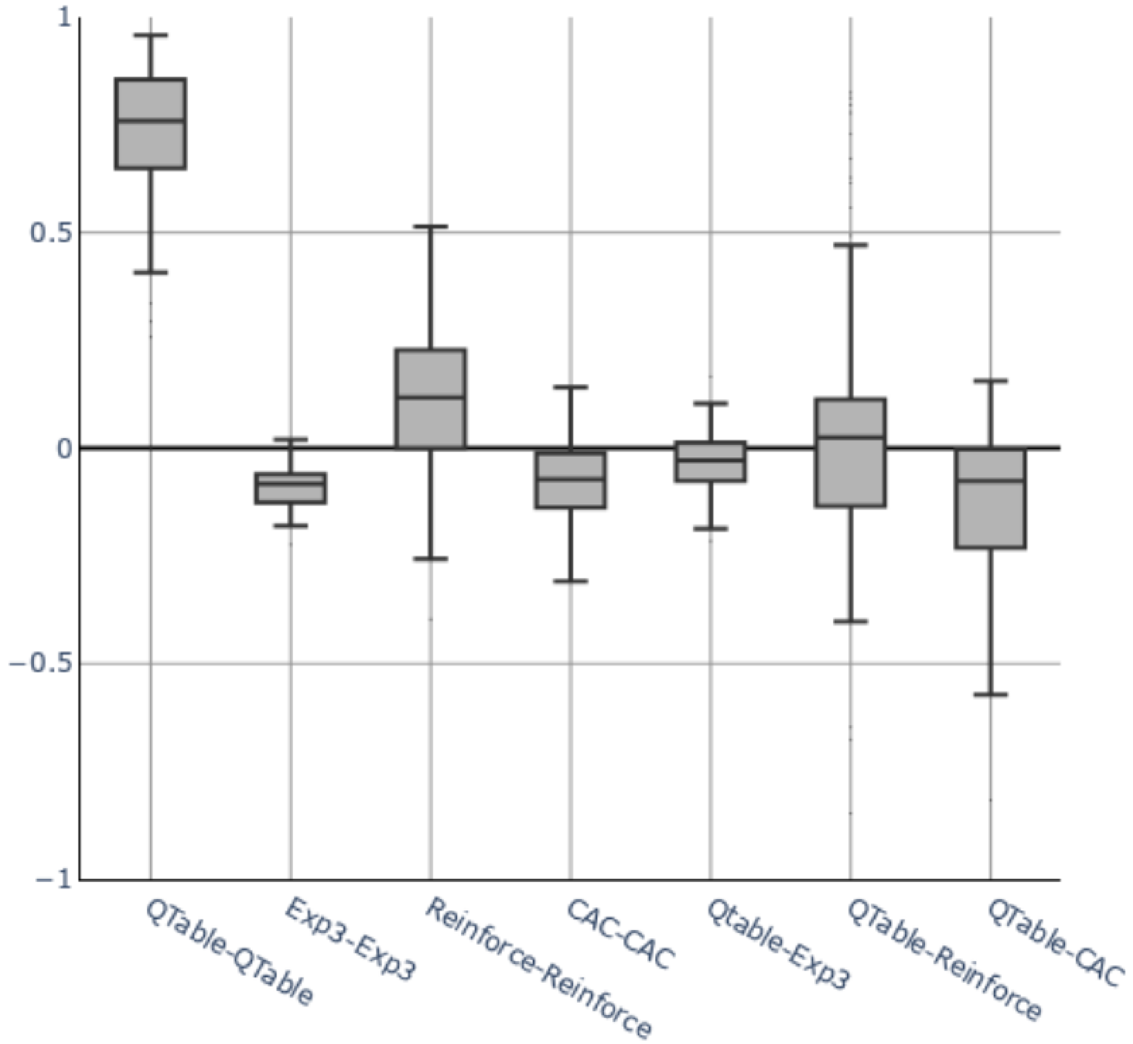
- More sophisticated Actor-Critic algorithms.

Continuous actor critic networks (CAC): a model-free RL setup with two interwined neural networks



- Unlike Q-learning, CAC are policy-based algorithms
- Both networks have three layers with 256 neurons in the hidden one.
- The exploration is endogenous to learning and can be tuned via an entropy parameter.
- CAC algos are routinely used in many fields: computer vision, robotics, autonomous driving, antilock braking system (ABS), etc.

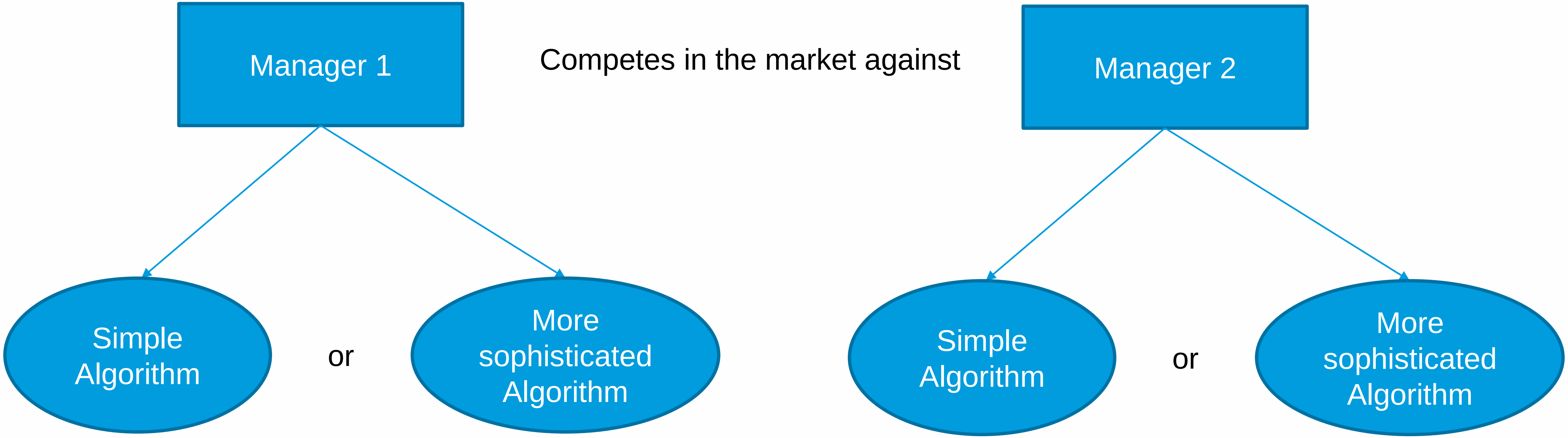
More sophisticated algorithms may not cooperate



Cooperation rate after learning for various algorithmic interactions.

The result has already been proven for Exp3 in den Boer et al. [2022].

The choice of AI technology



What would prevent agents from choosing simple seemingly colluding algorithms?

		Manager 2	
		Q-learning	**CAC**
Manager 1	**Q-learning**	**(12.13, 12.13)** (0.29, 0.29)	**(10.41, 11.42)** (0.50, 0.24)
	CAC	**(11.42, 10.41)** (0.24, 0.50)	**(11.00, 11.00)** (0.38, 0.38)

Table 1: Normal-form representation of the supergame when managers can choose Q-learning or CAC (bold characters show average limit payoffs, standard font shows the limit standard deviation).

- The (sophisticated) CAC algorithm consistently outperforms Q-learning.
- The choice of the colluding Q-learning algorithm is not individually rational.
- The equilibrium of the game of the algorithmic choice can lead to a competitive outcome.
- Results are qualitatively similar with Reinforce and Exp3.

The theoretical illustration and collusion by mistake
A numerical application with Q-learning
AI over-sophistication can reduce seeming collusion
The game of the technological choice
Conclusion

When algorithms collude by mistake

- The degree of exploration of Q-learning algorithms seems to have an impact on their propensity to cooperate at equilibrium.
  - **We encourage to verify that algorithmic cooperation is not due to insufficient exploration before investigating whether it is due to genuine collusion.**

- Sophistication limits cooperation (at least in our economic environment):
  - The reason might lie in the fact that the alternative algos we studied are policy-based.
  - **We encourage the use of algorithms other than Q-learning to study algorithmic collusion.**

- The game of algorithmic choice is complex, and selecting basic cooperative algorithms is not the only possible equilibrium for managers.
  - **This might be an indication of genuine collusion.**

- Extension:
  - Other competing environments.
  - Other sophisticated algorithms.
  - Other exploration strategies.