ALGORITHMIC COLLUSION OR ALGORITHMIC CONFUSION? RETHINKING THE RISKS OF AI PRICING FOR COMPANIES AND POLICYMAKERS

BY AI DENG[1]

1 Ai Deng, PhD, Managing Director, BRG. adeng@thinkbrg.com. Disclaimer: The views and opinions expressed in this article are solely those of the author and do not necessarily reflect the views of BRG or any of its affiliates.

CPI ANTITRUST CHRONICLE
September 2026

Visit www.competitionpolicyinternational.com for access to these articles and more!

ALGORITHMIC COLLUSION OR ALGORITHMIC CONFUSION? RETHINKING THE RISKS OF AI PRICING FOR COMPANIES AND POLICYMAKERS

By Ai Deng

Concerns about algorithmic collusion are often driven by the belief that increasingly sophisticated AI systems are more likely to autonomously learn to collude. This article reviews a growing body of academic literature suggesting a more nuanced insight. Across studies of AI pricing algorithms, supracompetitive prices frequently arise not from advanced strategic intelligence, but from limitations in algorithm design. More sophisticated algorithms, by contrast, often learn to compete more effectively. The article examines this distinction between pricing outcomes and the mechanisms that generate them. It concludes by considering emerging regulatory proposals and their implications for firms, policymakers, and antitrust enforcement.

I. INTRODUCTION

Regulators and the public often view the threat of algorithmic collusion through the lens of sophisticated, autonomous agents. One prevailing view is that as Artificial Intelligence ("AI") becomes more advanced, it is more likely to *learn* to coordinate with rivals to fix prices and extract monopoly rents without human intervention. This sentiment is made clear in a CMA whitepaper titled *Algorithms: How They Can Reduce Competition and Harm Consumers* which states "[a]s algorithmic systems become more sophisticated, they are often less transparent, and it is more challenging to identify when they cause harm…. collusion appears an increasingly significant risk if the use of more complex pricing algorithms becomes widespread."[2] A former FTC commissioner echoed: "[a]s algorithms and the software running them become more sophisticated, however, coordinated behavior may become more common without explicit 'instruction' by humans."[3] An article published in *Science* by a group of academic scholars expressed similar concerns: "[t]he enhanced sophistication of learning algorithms makes it more likely that AI systems will discover profit-enhancing collusive pricing rules, just as they have succeeded in discovering winning strategies in complex board games such as chess and Go."[4] This perspective implies that the risk to market competition scales with the complexity of the technology.

Indeed, as Deng (2024) discussed extensively, in the context of *first-party* algorithms, academic literature has shown that they *can* be designed to be collusive and sophisticatedly so *when the design objective of an algorithm is collusion*.[5] In this article, I synthesize a recent line of academic literature on first-party algorithms suggesting an important nuance. When an algorithm is not explicitly designed to collude, this emerging literature suggests that the danger to competition does not necessarily stem from AI becoming too smart, but *potentially* from it remaining too simple. For example, a recent article coined the term "artificial stupidity" and argues that rather than *artificial intelligence*, it likely explains the seemingly collusive outcomes observed in experimental studies in the academic literature.[6]

This line of research indicates that simple, "asynchronous," or otherwise less sophisticated AI pricing algorithms—those lacking the sophistication to understand the market environment or to model counterfactuals—may be the ones more prone to supracompetitive pricing. Conversely, algorithms capable of complex reasoning and robust exploration tend to learn to compete aggressively, driving prices down to more competitive levels. Although this research remains in its infancy and is based almost exclusively on stylized models and simulations, these early findings nonetheless offer valuable lessons for policymakers, regulators, and companies alike. Policymakers and regulators may wish to consider carefully whether policies or regulations that discourage the development of advanced algorithms or reduce diversity in such technologies could have unintended consequences. Policy may therefore need to account for two distinct risks: *first*, that poorly specified or insufficiently sophisticated algorithms may drift toward supracompetitive outcomes (the focus of the article); and *second*, that sufficiently capable systems could, in principle, be designed or configured to facilitate intentional collusion (see the broader discussion in Deng, 2024). These considerations also matter for companies. How supracompetitive outcomes are interpreted, whether as evidence of coordination or as the byproduct of lack of sophistication in algorithmic design, can shape how companies think about the design, deployment, and oversight of pricing algorithms. Central to disentangling these risks is a distinction that runs throughout this article: the difference between *outcome*, i.e. whether a pricing process yields sustained supracompetitive prices, and *mechanism*, i.e. what causes those prices to arise. As the literature surveyed below will show, collusive-looking outcomes may be explained more by limitations in algorithm design than by superior strategic intelligence.

II. THE FOUNDATIONAL FINDINGS ON AUTONOMOUS ALGORITHMIC COLLUSION

To understand the nature of autonomous algorithmic collusion, one must first visit the foundational findings by Calvano et al. (2020). Their work demonstrates that standard Reinforcement Learning ("RL") agents—specifically those using the so-called Q-learning—could autonomously

2 CMA *Algorithms: How They Can Reduce Competition and Harm Consumers* 19 (Jan. 19, 2021), https://www.gov.uk/government/publications/algorithms-how-they-can-reduce-competition-and-harm-consumers/algorithms-how-they-can-reduce-competition-and-harm-consumers.

3 Terrell McSweeny & Brian O'Dea, *The Implications of Algorithmic Pricing for Coordinated Effects Analysis and Price Discrimination Markets in Antitrust Enforcement*, 32 *Antitrust* 1 (Fall 2017).

4 Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, Joseph E. Harrington, Jr. & Sergio Pastorello, *Protecting Consumers from Collusive Prices Due to AI*, 370 *Science* (2020).

5 Ai Deng, *What Do We Know About Algorithmic Collusion Now?* (Working Paper, 2024). Separately, the broader literature has bifurcated along another dimension: the study of *first-party* algorithms, in which competing firms each develop and deploy their own pricing algorithms, and *third-party* algorithms, in which competitors rely on a common algorithm provided by an outside vendor. These settings raise distinct questions. This article addresses the first-party setting. For a discussion of third-party algorithms, see Deng (2024).

6 Winston Wei Dou, Itay Goldstein & Yan Ji, *AI-Powered Trading, Algorithmic Collusion, and Price Efficiency* (Nat'l Bureau of Econ. Rsch., Working Paper No. 34054, 2025), http://www.nber.org/papers/w34054.

sustain high prices without being programmed or otherwise instructed to do so.[7] This finding established the baseline concern: that AI could generate pricing behavior that could be interpreted as resembling cartel conduct without explicit instruction.

As an RL algorithm, the Q-learning pricing agents in the Calvano et al. (2020) operate on a principle of "trial and error" with no prior knowledge of the demand for their products. Other than rival's prices, they do not know rival's costs or profits. The learning process involves a trade-off between exploration and exploitation.[8] In the beginning, the algorithms "explore" by choosing random prices to gather data on what works.[9] As they accumulate experience, they shift toward "exploitation," increasingly choosing the price that their memory suggests will yield the highest reward, i.e. profit.[10] In the Calvano et al. model, this shift is hard-coded: the rate of exploration is programmed to decay over time, eventually reaching zero.[11] Practically speaking, this means the agents are effectively programmed to "freeze" their behavior that they "believe" is optimal given their experiences. We will return to this critical assumption and explain why more recent literature has identified it as a major limitation in the study of algorithmic collusion using RL pricing algorithms.

Calvano et al. (2020) find that this process consistently leads the pricing agents to settle on supracompetitive prices. Recognizing that supracompetitive prices are not synonymous with collusion, they further report experimental evidence suggesting that their AI pricing agents not only "learn" to charge supracompetitive prices but also to implement a form of punishment for price deviation by the competitor: If one agent deviates from the initial supracompetitive price by cutting its price, the other agent would immediately cut prices significantly, in some situations, even more than the initial cut by the competitor.[12] Such a phenomenon mimics the type of reward-punishment scheme that, in standard economic theory, can sustain supracompetitive prices in a cartel despite the incentive to cheat. Interestingly, their simulation evidence shows that over time, both agents gradually raise prices back to the supracompetitive level after the initial unilateral price cut, all without ever communicating.[13] Calvano et al (2020) ultimately conclude that the AI pricing agents in their experiments "consistently learn to charge supracompetitive prices, without communicating with one another. The high prices are sustained by collusive strategies with a finite phase of punishment followed by a gradual return to cooperation."[14] Since then, reinforcement Q-learning has become the workhorse AI pricing algorithm in academic literature. Calvano et al. (2020) has since become highly influence and, not surprisingly, captured attention from both academic researchers and antitrust regulators.[15]

---

7 Calvano, E., et al. (2020), "Artificial Intelligence, Algorithmic Pricing, and Collusion," American Economic Review, 110(10), p. 3267: "We find that the algorithms consistently learn to charge supracompetitive prices, without communicating with one another. The high prices are sustained by collusive strategies with a finite phase of punishment followed by a gradual return to cooperation."

8 Calvano et al (2020), supra note 7, p. 3271, ("the algorithm has to be instructed to experiment, i.e. to gather new information by selecting actions that may appear suboptimal in the light of the knowledge acquired in the past. Plainly, such exploration is costly and thus entails a trade-off between continuing to learn and exploiting the stock of knowledge already acquired. Finding the optimal resolution to this trade-off may be problematic, but Q-learning algorithms do not even try to optimize in this respect: the mode and intensity of the exploration are specified exogenously.")

9 Calvano et al (2020), *supra* note 7, p. 3274 ("This means that initially the algorithms choose in purely random fashion, but as time passes, they make the greedy choice more and more frequently.")

10 Technically speaking, each algorithm maintains a memory bank (a "Q-matrix") that estimates the total long-term profit expected from choosing a specific price, given the current state of the market. In the context of this study, a "state" is simply the memory of the prices charged by all firms in the preceding periods. See, for example, Calvano et al (2020), *supra* note 7.

11 Calvano et al (2020), *supra* note 7, p. 3274, ("We use the -greedy model with a time-declining exploration rate. Specifically, we set , where is a parameter. This means that initially the algorithms choose in purely random fashion, but as time passes, they make the greedy choice more and more frequently. The greater , the faster the exploration diminishes.") In this formulation, as .

12 Calvano et al (2020), *supra* note 7, p. 3282, "Clearly, the deviation gets punished. As Table 3 shows, in more than 95 percent of the cases the punishment makes the deviation unprofitable; that is, 'incentive compatibility' is verified." And p. 3285 "For small price cuts, the pattern just described represents a form of overshooting: that is, both algorithms cut their prices further in period $\tau = 2$, below the exogenous initial reduction of period $\tau = 1$. This is illustrated in Figure 6, which shows the average impulse-response corresponding to one of these smaller deviations. The overshooting would be difficult to rationalize if what we had here was simply a stable dynamic system that mechanically returns to its rest point after being perturbed. But it makes perfect sense as part of a punishment."

13 Calvano et al (2020), *supra* note 7, p. 3282, "The dynamic structure of the punishment is very interesting. After an initial price war, the algorithms gradually return to their predeviation behavior," and p. 3283, "What is systematic is the return to the initial prices; in most of the cases, the punishment ends after 5–7 periods."

14 Calvano et al (2020), *supra* note 7, p. 3267.

15 As of March 8, 2026, this article has been cited by nearly 1,000 times according to Google Scholar. The article has also been cited by CMA (see, e.g., *Algorithms: How they can reduce competition and harm consumers*, Jan 2021), OECD (see, e.g., OECD Competition Policy Roundtable Background Note on Algorithmic Competition, 2023), the US FTC (see, e.g., Issue Spotlight: The Rise of Surveillance Pricing, 2025).

III. KEY INSIGHTS FROM LITERATURE: DEEPENING THE ANALYSIS

A number of subsequent academic studies form a cohesive body of work that systematically reevaluates the "smart AI cartel" narrative. This emerging literature provides evidence that the observed seemingly collusive pricing behavior such as those reported by Calvano et al. (2020) is not a result of superior intelligence.

A. The Illusion of Strategy?

The most direct challenge to the "smart AI cartel" narrative targets the behavioral evidence that Calvano et al. (2020) offered in support of a collusive interpretation: the apparent reward–punishment patterns. Abada & Lambin (2022), Epivent & Lambin (2023), and more recently Grondin et al. (2025) and den Boer et al. (forthcoming) cast further doubt on the "smart AI cartel" narrative by closely examining this question.[16]

The first three studies show that, rather curiously, while unilateral price cuts are met with a substantial price reduction by the competing algorithm — often interpreted as evidence of punishment — Q-learning agents respond to price increases in exactly the same way: when a competitor unilaterally raises its already supracompetitive price closer to the monopoly level, the competing algorithm does not respond by maintaining or increasing its own price. Instead, it nevertheless goes on to start a price war.[17],[18] Epivent & Lambin (2023) further document instances of *self*-punishment, in which the deviating algorithm itself initiates aggressive price cuts following a price increase.[19] The symmetric punishment of both pro-competitive and pro-collusive deviations as well as self-punishment is difficult to reconcile with a rational reward–punishment strategy for sustaining supracompetitive prices.[20] As Abada & Lambin (2022) argue, this behavior "contradicts the nature of targeted punishment," noting that a rational agent "should reward or at least be indifferent to pro-collusive moves." Based on these findings, Epivent & Lambin (2023) argue that it may be premature to conclude that AI algorithms such as Q-learning have truly learned to collude. Instead, they point out that "what has been observed so far may not be collusion but is instead a failure to learn about the environment and competitive setting." (emphasis added). As such, these researchers call for "further research … to rule out alternate justifications to supracompetitive profits and conclusively prove algorithmic 'collusion.'" and pending more conclusive research, "courts should prioritize the intentions and competence with which algorithms are designed and tested over observations of what the algorithms seem to be doing."[21]

Grondin et al (2025) report two additional results that they argue "would [further] indicate that the strategies are more dependent on changes of the other agent's behaviour, rather than an actual understanding of the market dynamics."[22] First, they observe that, after reaching the initial supracompetitive prices, if one pricing agent *increases* its price further but the other is forced to hold the (supracom-

16 Ibrahim Abada & Xavier Lambin, *Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided?* (SSRN Working Paper, 2022). Andrea Epivent & Xavier Lambin, *On Algorithmic Collusion and RewardPunishment Schemes* (Working Paper, 2023). Suzie Grondin, Arthur Charpentier & Philipp Ratz, *Beyond Human Intervention: Algorithmic Collusion through MultiAgent Learning Strategies* (Working Paper 2025), arXiv:2501.16935. Arnoud V. den Boer, Janusz M. Meylahn & Maarten Pieter Schinkel, *Artificial Collusion: Examining Supracompetitive Pricing by Q-Learning Algorithms, Mgmt. Sci.* (forthcoming).

17 Epivent and Lambin (2023), *supra* note 16 ("The table shows that whatever the magnitude of the price increment, invitation to collude are systematically punished by aggressive price wars." Grondin et al (2025), *supra* note 16, report "The centre panes depicts a slightly different result when we manually force a higher price on one of the agents. Here the results are somewhat surprising, as the second agents again lowers its price, although one would expect it to keep it steady instead (as it would result in higher market share, at the same price)."

18 Calvano et al (2023) commented on this finding, stating that "…, Epivent and Lambin (2023) point out that the algorithms of CCDP learn to punish not only deviations to prices lower than the collusive prices, but also higher. *Naturally, this is still genuine collusion, not spurious*: even grim trigger strategies, for instance, entail punishments of higher prices. However, the algorithms do not have an analytical comprehension of the strategic environment but learn purely by trial and error. That by doing so they may learn to punish also higher prices seems therefore puzzling." (emphasis added). Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò & Sergio Pastorello, *Algorithmic Collusion: Genuine or Spurious?* 90 Int'l J. Indus. Org. 102973–102993 (2023). Epivent and Lambin (2023), *supra* note 16, however, while acknowledging the theoretical point by Calvano et al (2023), offer a more practical perspective: "to the best of the author's knowledge, such a phenomenon has never been observed in real-life cases of collusion. Further, the punishment of price increments can be viewed as the enforcement of a price ceiling which is often used as a remedy to, rather than a source of, collusive concerns (see e.g. United States Department of Justice [1982], European Commission [2016], Japan Fair Trade Commission [1997] for examples)." Epivent and Lambin (2023) go on to say that it is "hardly conceivable that rational agents would collude on strategies that involve even more cut-throat competition than Nash competition, and that sanction profit-increasing deviations."

19 Epivent and Lambin (2023), *supra* note 16. ("Another striking observation is that the deviating agent appears to self punish with even lower prices in the period that follows a deviation: it always does so following upward deviations (price increments) or mild downward deviations (price cuts), but not for larger downward deviations.")

20 Epivent and Lambin (2023), *supra* note 16, also address and reject the interpretation that the self-punishing behavior may be due to the deviating AI's anticipating retaliation from the competitor. ("To counter this claim, one could argue that the apparent self-punishment we mention in the previous section can be justified through the fact that the deviating algorithm anticipates the punishment of the nondeviating algorithm (Calvano et al. [2020b]). In this section, we show with a simple experiment that such self-punishment is difficult to rationalize with genuinely collusive behavior. In particular, we reject the interpretation that AIAs anticipate rivals' reactions with preemptive self-punishments.")

21 Epivent and Lambin (2023), *supra* note 16.

22 Grondin et al (2025), *supra* note 16.

petitive) price, the first pricing agent again immediately lowers its price well below the supracompetitive level reached prior to the unilateral price increase. Second, when one of the pricing agents' prices is permanently reset to just above the competitive level, the competing pricing algorithm drops its price in response but "not enough [to] counteract the effects of the lower price of the competition, in turn losing most of the market share." They conclude that "the observed behaviour might put into question some results from the literature, as it appears that the algorithms simply learn to converge back to the equilibrium price after a simple shock rather than having a truly collusive strategy. Further, the learned strategy appears to imply a limited understanding of the mechanisms, where lowering prices is understood, but not raising prices.… In turn, it appears that the algorithms are easy to fool and would present a significant risk for any firm employing them without close supervision."

den Boer et al. (forthcoming) provide another comprehensive analysis of the Q-learning pricing algorithm. Among others, they find that Q-learning's learning process is "intrinsically slow, without hope for being sped up," formalizing an observation made by Deng (2020).[23] More importantly, the researchers find that the Q-learning algorithms "generate negligible extra-profits for the firms committing to their use and are easily outperformed by reasonable alternative pricing rules available."[24] As a result, in the researchers' words, it is "difficult to reconcile with the idea that Q-learning would be implemented in practice by rational agents."

B. Causes of Supracompetitive Prices

If the behavioral patterns do not reflect genuine collusive strategy, the natural next question is: what mechanisms do account for the supracompetitive prices reported by Calvano et al (2020) and other studies? The literature identifies several distinct but complementary explanations, each pointing to limitations in algorithm design rather than strategic sophistication.

1. The Failure of Counterfactual Reasoning

Asker et al. (2024) highlight that one explanation for why the Q-learning algorithm of the type studied by Calvano et al (2020) drift toward high prices may be the lack of an understanding of basic economics of how competition and markets work.[25] To that end, they compare and contrast two types of AIs, known as asynchronous and synchronous learning.

An asynchronous algorithm, the one that Calvano et al. (2020) used, updates its valuation of a strategy based solely on the action it actually took.[26] If the asynchronous algorithm sets a high price and the competing algorithm also happens to set a high price, say, by pure price experimentation, both AIs record a healthy profit, which in turn reinforces the choice of high prices. Asynchronous AI does not calculate what it would have earned had it undercut its competitor — a counterfactual scenario. They are simply not equipped with that basic economic understanding. Consequently, these AI agents often stagnate in a situation where prices remain high because they lack the ability to assess the value of deviating.

By contrast, a more sophisticated, synchronous AI acts more like a deliberate economic agent. Equipped with basic economic knowledge — such as a downward-sloping demand curve — it uses a model to estimate counterfactuals ("If I had charged $5 instead of $10, given my competitor's prices, what would have happened?"). Asker et al. (2024) show that when agents possess this sophistication, they drive prices down to (more) competitive level.[27]

To understand the intuition, consider a duopoly where both firms are currently charging a high price of $10.

- **The Asynchronous Agent**: This agent looks at its memory. It sees that when it charged $10 in the past (and say, the rival happened to match), it earned a high profit. It does not ask what happens if it cuts the price to $9 when the rival charges $10 so it effectively fails to recognize a profitable price change. Over time, both agents may be led to believe that $10 is the best price to set.

23 Ai Deng, *Algorithmic Collusion and Algorithmic Compliance: Risks and Opportunities*, in *The GAI Report on the Digital Economy* (2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3733743.

24 Specifically, the researchers show that Q-learning underperforms against other commonly used and familiar AI algorithms such as Exp3, see, den Boer et al, *supra* note 16, § 4.5.

25 John Asker, Chaim Fershtman & Ariel Pakes, *The Impact of AI Design on Pricing*, 33 *J. Econ. & Mgmt. Strategy* 276 (2024).

26 Asker et al. (2024), *supra* note 25 ("Asynchronous learning occurs when the AI only learns about the return from the action it took... [it] can lead to pricing close to monopoly levels.")

27 Asker et al. (2024), *supra* note 25 ("Synchronous learning occurs when the AI conducts counterfactuals to learn about the returns it would have earned had it taken an alternative action... (perfect) synchronous updating leads to competitive pricing.")

- **The Synchronous Agent (The "Smart" Learner)**: This agent does not rely solely on memory and past actions. It observes the rival's price of $10 and uses its internal demand model to simulate a counterfactual: "If I cut my price to $9 right now, I would capture the entire market." It calculates that the immediate profit from undercutting is greater than "sharing" the market at $10.

Because the synchronous agent can "see" the profit opportunity in undercutting, it is more ready to deviate from the high price before it can stabilize. Thus, the crucial insight is that supracompetitive prices may persist not because the asynchronous algorithm has cleverly solved the dynamic game, but precisely because it lacks the intelligence to calculate the immediate benefits of competition.[28]

2. Imperfect Exploration and the Mirage Effect

Like Asker et al. (2024), Abada & Lambin (2022) and Abada et al. (2024) also did not view the supracompetitive prices as evidence that Q-learning algorithms have learned to tacitly collude through sophisticated strategic thinking.[29] Instead, their analysis suggests that the observed outcomes (supracompetitive prices as well as the seeming reward and punishment scheme discussed above) may in fact be better understood as a failure to learn to compete rather than a successful attempt to collude.[30]

The core mechanism they identify is imperfect exploration. As alluded to above, simple reinforcement learning algorithms are typically implemented with exploration rates that decay deterministically and rapidly. As exploration diminishes, AI agents may stop experimenting with alternative actions before they have adequately learned the true profitability of competitive deviations. This gives rise to what Abada et al. (2024) term a mirage effect: algorithms may converge to high prices not because they have discovered a stable collusive equilibrium supported by credible punishments, but because the algorithms never fully learn that aggressive competition is individually optimal. In this sense, the algorithms become "stuck" in supracompetitive outcomes by mistake. As Abada & Lambin (2022) put it, "seeming collusion could originate in imperfect exploration, rather than excessive algorithmic sophistication."[31]

Abada et al. (2024) further show that when exploration is made sufficiently thorough, even the same type of AI algorithms set prices toward competitive levels.[32] Moreover, when they consider more sophisticated reinforcement learning protocols, they find that these algorithms also systematically converge to competitive outcomes and, in direct competition, outperform the seemingly collusive Q-learning agents.[33] Based on these findings, Abada et al. call for caution "considering the claims of some of the extant literature where it is sometimes conjectured that more sophisticated algorithms would, if anything, collude more effectively than coarser algorithms. Instead, sophistication appears to allow algorithms to learn to compete more effectively, at least for the particular set of algorithms we analyze and for the particular economic environment in which we make them compete."

den Boer et al. (forthcoming) make a related point through a detailed look at Q learning, emphasizing that while supracompetitive prices may arise in simulations, their persistence as a stable long-run equilibrium outcome is fragile. In particular, they show that sustaining supracompetitive prices as a stable equilibrium of the learning dynamics requires the rate of exploration to fade neither too quickly nor too slowly, but to fall within a narrow range. Too quickly, the algorithms may freeze before learning strategies that can sustain supracompetitive prices as a stable

28 Asker et al (2024), *supra* note 25, consider both the case where the pricing agents are completely short-sighted so they do not care about the potential retaliatory responses from competitors in the future and the case where they do care and therefore may hesitate to price undercut. In both cases, the synchronous agents are found to set much lower prices than asynchronous agents. In a subsequent article, Calvano et al (2023) points out a difference between their original study and Asker et al (2024) in how they specify the exploration rate.

29 Ibrahim Abada & Xavier Lambin, *supra* note 16. Ibrahim Abada, Xavier Lambin & Nikolay Tchakarov, *Collusion by Mistake: Does Algorithmic Sophistication Drive Supra-Competitive Profits?* 319 *Eur. J. Operational Res.* 927 (2024).

30 See, e.g., Abada & Lambin (2022), *supra* note 16 ("Interestingly, the results reported in Section 4.2 show that algorithms fail to set competitive or fully collusive quantities which would be the two (economically) natural benchmarks. Instead, they converge to quantities strictly between these two benchmarks (which translates into a welfare loss parameter g always strictly ranging between 0 and 1). This raises questions about whether algorithms truly learn sophisticated collusive strategies, or simply fail to learn a rational reaction function to the competitors in a sufficiently large range of actions.")

31 Abada & Lambin (2022), *supra* note 16 ("We show that seeming collusion could originate in imperfect exploration, rather than excessive algorithmic sophistication.")

32 Abada et al. (2024), supra note 29 ("In particular, we show that allowing for more thorough exploration does lead otherwise seemingly-collusive Q-learning algorithms to play more competitively... [sophistication] may, in some situations, provide a solution to the challenge of algorithmic seeming collusion, rather than exacerbate it.")

33 Abada et al. (2024), supra note 29 ("Our results suggest that the occurrence of seeming collusion in our simple setting is inherent to the specificity of Q-learning and its exploration policy: we observe that the other algorithms, that embed different exploration policies, systematically converge to the Nash equilibrium…. In addition, by numerically analyzing the competition between (seemingly-collusive) Q-learning and the other algorithms, we observe that these consistently outperform Q-learning, when trained in a duel.")

longrun equilibrium. Too slowly, persistent experimentation may prevent convergence to any stable equilibrium that sustains supracompetitive prices.[34]

A recent study provides additional affirmative evidence that more sophisticated algorithms are less likely to lead to supracompetitive prices.[35] The researchers simulated market interactions using both simple Q-learning a la Calvano et al (2020) and several more advanced deep learning algorithms. They found that one particular advanced deep learning algorithm, known as Proximal Policy Optimization ("PPO"), when competing against itself, most frequently reaches competitive price levels, especially compared to the simple Q-learning algorithm. Even in the scenarios where it produced supracompetitive outcomes, those prices were consistently lower than those produced by other algorithms. The researchers acknowledge that it may be counterintuitive that an "algorithm that is usually seen as 'data hungry' and sensitive to a huge amount of hyperparameters can effectively mitigate collusion." Their explanation is that, "once properly trained and parameterized," this sophisticated deep learning algorithm "allows robust exploration implicitly by directly parameterizing and learning from a gradually improving policy," making it easier to "escape local optima compared to the other algorithms analyzed and thus leads more often to a competitive consensus or decreased degrees of collusion."[36]

3. Artificial Stupidity: Negativity Bias

The mirage effect attributes premature convergence to insufficient exploration. A separate critique points to a different defect in how Q-learning processes the exploration it does undertake. In the context of financial markets, Dou et al (2025) identify another mechanism through which Q-learning algorithms may appear to soften competition.[37] The researchers term this mechanism "artificial stupidity" because it stems from a shared flaw in the AI learning process rather than sophisticated strategies.[38] This flaw manifests in presence of high market noise: when an algorithm attempts a competitive and aggressive trading strategy during a period of high market noise, it may suffer a loss simply due to bad luck. To the Q-learning algorithm, however, this loss is treated as evidence that aggressive (and hence competitive) action is inferior and should be avoided. In other words, Q-learning algorithm inherently possesses a "negativity bias": it tends to prune such aggressive action from its playbook immediately after suffering a significant loss. Consequently, the algorithms retreat to conservative, passive strategies, appearing to soften competition. In Dou et al.'s account, this behavior reflects market noise interacting with a shared learning flaw, rather than strategic coordination among the algorithms.[39]

4. Beyond Q Learning: The Naivety of Context-Free Bandit

If supracompetitive pricing were an artifact unique to Q-learning, one might dismiss it as a curiosity of a single class of algorithms. But similar patterns emerge from even simpler algorithms. One leading example is the so-called multi-armed bandit algorithm.[40]

34 den Boer et al. (2025), *supra* note 16, explain that "The speed with which the algorithms converge depends on the decreasing exploration rate in two ways: (1) there is an initial period in which the collusive strategy equilibria do not exist yet, and (2) after they exist, a specific sequence of events has to occur in order to transition to a collusive strategy equilibrium. Supra-competitive limit prices sustained by collusive strategy equilibria can only be learned after a sufficiently long exploration period because only then do they come into existence. To see this, consider the extreme case of a fixed value $\epsilon = 1$ [fixed exploration rate], for which we know there are no collusive equilibria—compare Figure 1…. Q-learning with time-dependent exploration probability $\epsilon_t = \exp(-\beta t)$ does not admit stationary equilibria in the sense defined in Section 4.2.1. … So considered, Figure 1 suggests that convergence to a collusive strategy equilibrium is not possible before ϵt [exploration rate] has dropped below a critical value $\epsilon(\delta)$,…"). In their experiments, they show that making the exploration rate decay faster "reduces the amount of collusion…" and "$\epsilon t \equiv 0.1$ [a fixed exploration rate] reduces collusive limit strategy profiles to less than 20 percent."

35 Shidi Deng, Maximilian Schiffer & Martin Bichler, *Algorithmic Collusion in Dynamic Pricing with Deep Reinforcement Learning* (Working Paper, 2024). Note that this study focuses on the pricing outcomes, rather than the mechanisms.

36 *Supra* note 35.

37 Dou et al (2025), *supra* note 6.

38 Dou et al (2025), *supra* note 6 ("Algorithmic collusion emerges from two distinct mechanisms. The first mechanism is... ('artificial intelligence'), while the second stems from homogenized learning biases ('artificial stupidity')."

39 Readers may question why a "lucky" outcome—where aggressive trading yields high profits due to a positive market shock—does not simply cancel out an "unlucky" one. As explained by Dou et al (2025), *supra* note 6, the answer lies in the inherent asymmetry of Reinforcement Learning (RL). As noted, RL balances "exploration" with "exploitation" (maximizing known rewards). If an aggressive action generates a substantial profit due to a positive shock, the algorithm flags it as a high-value action. Consequently, the algorithm's "exploitation" function ensures this strategy is revisited frequently. Over time, this repetition allows the random market noise to average out, leading the algorithm to self-correct and learn the "fair" value of the action. Conversely, when an aggressive action results in a significant loss due to a negative shock, the algorithm downgrades it and avoids using it. By effectively "pruning" the action from its playbook to avoid further losses (as it believes), the algorithm deprives itself of the opportunity to discover that the culprit was random noise rather than a flawed learning protocol.

40 For additional examples, see A. V. den Boer & J. M. Meylahn, *A (Mathematical) Definition of Algorithmic Collusion* 1 (Working Paper, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5012923. ("There are many algorithms that, if used by both players in a pricing duopoly, may generate supracompetitive outcomes: e.g., the algorithm by Cooper et al. (2015, cf. their Figure 6 and 7) that suffers from incomplete learning, an algorithm that always chooses a price uniformly at random from the set of feasible prices (cf. den Boer et al., 2022, Section 4.2), an algorithm that learns to price 1 percent above the Nash equilibrium price (assuming this is uniquely defined), or an algorithm that learns to price at the monopoly price in a symmetric market. These algorithms might generate supracompetitive outcomes if used by both players, but they can easily be 'punished' or outperformed if the other player uses a different and more sensible algorithm.")

To understand a multi-armed bandit, imagine a gambler at a row of slot machines (the classic "one-armed bandits"). The gambler has no idea how the machines work or what anyone else in the casino is doing; they simply pull a lever, record the payout, and, over time, learn to pull the lever that pays the most. In a market context, a "context-free" pricing bandit is similarly oblivious: it tests a price, sees the resulting profit, and updates its strategy without ever "seeing" its competitor or understanding the broader market dynamics. Notice that this simple algorithm—like the more sophisticated Q-learning—still "learns" through pricing experimentation, meaning repeated trial-and-error adjustments of price based on the profit feedback from prior price choices.

Hansen et al. (2021) study competing firms that use such bandit algorithms independently (i.e. with no coordination) but with a key form of ignorance: each firm's algorithm behaves as if it is a monopolist, even though the firms operate in an oligopolistic market.[41] They show that algorithmic prices can become supracompetitive when market feedback is sufficiently stable and informative. Supracompetitive prices emerge because the firms' independent price experiments tend to become correlated over time, that is, their price choices synchronize. Once prices synchronize—so that the firms are more likely to post high prices or low prices at the same time—each firm is more likely to observe high profits in precisely those periods when both firms are pricing high. The resulting profit feedback can then reinforce the mistaken conclusion that high prices are the best choice. In other words, the algorithms set high prices not because they understand the difference between competition and collusion, but because mutual and independent ignorance can become self-reinforcing.

Douglas, Provost & Sundararajan (2024) provide a complementary and more general analytical explanation for when supracompetitive prices should be expected from context-free bandits.[42] They model pricing agents that do not observe competitors' actions or outcomes and have no model of the strategic interaction, yet can still converge to high price behavior; they coin the term "naïve collusion" for this phenomenon.

Their central distinction is whether the algorithm is deterministic: in their terms, a deterministic bandit is one that, given a realized history, selects a single action with probability 1. In plain language, these are algorithms that follow rigid, predictable rules: given the same history of sales/profit data, they will make the same pricing decision. Under symmetry (both firms using the same deterministic algorithm), "naïve collusion" emerges robustly in their framework. By contrast, with persistent random exploration, the agents do not settle on supracompetitive prices in the long run, because continued experimentation disrupts synchronization. These results highlight, once again, that supracompetitive outcomes can arise not because the algorithm is strategically sophisticated, but because it is too naïve—mis-specified, context-free, and rigid enough to synchronize.

More broadly, these results point to the role of similarity across competing algorithms as a contributing factor to supracompetitive outcomes. The next section considers this idea more directly, examining how standardization and symmetry in algorithmic design and deployment can give rise to such outcomes.

5. The Role of Orchestration and Standardization

Most of the academic studies surveyed so far find the core issues inside the *individual* algorithm—in how it reasons, explores, and updates. But as alluded to above, a separate strand of literature finds them one level up: in the *choice* of symmetric pricing algorithms by competitors.

Multiple studies argue that the seemingly reliable supracompetitive outcomes in the literature are tied to researcher-imposed design and evaluation choices that effectively *standardize* how learning unfolds across firms.

Carissimo, Falniowski, Rahimi & Nax (2025) emphasize that many well-known simulation results rely on similar learning protocols shared by competitors.[43] They argue that "clean" collusive looking outcomes highlighted in the literature are closely linked to such coordinated design choices. To that effect, they propose a different (called "meta game") perspective to clarify how those choices relate to the real world: instead of treating such an algorithmic design as neutral research settings, they ask whether those design choices would be made if

41 Karsten T. Hansen, Kanishka Misra & Mallesh M. Pai, *Frontiers: Algorithmic Collusion*, 40 *Mktg. Sci.* (2021). ("We find that misspecified learners, by correlating pricing experiments, drift toward monopoly outcomes . . . even when algorithms are independent.").

42 Cameron Douglas, John Provost & Arun Sundararajan, *Naive Algorithmic Collusion*, at 1 (Working Paper, 2024), https://arxiv.org/abs/2411.16574 ("We show that these context-free bandits, with no knowledge of opponents' choices or outcomes, still will consistently learn collusive behavior what we call 'naive collusion.'").

43 Cesare Carissimo, Fryderyk Falniowski, Siavash Rahimi & Heinrich Nax, *Algorithmic Collusion Is Algorithm Orchestration* (Working Paper, 2025), arXiv:2508.14766v1. ("Independent of exact implementation, these policies often handle the exploration exploitation tradeoff similarly: the learner's price setting is random in the initial learning phase, and the randomness decreases such that the learner's price setting eventually becomes deterministic and price fluctuations are minimized.")

algorithm developers were designing algorithms to act in *unilateral* interest such as maximizing unilateral profit.[44] Their main finding is that, in their study of Q-learning, collusive pricing patterns require non-competitive co-parametrization ("orchestration"), whereas competitively parametrized algorithms, i.e. those designed based on unilateral interest, yield pricing that is much closer to competitive benchmarks.[45]

den Boer et al. (2025) and Grondin et al. (2025) also emphasize that many Qlearning simulations yielding supracompetitive prices rely on a high degree of *standardization* and *synchronization* across competitors, including, e.g., identical algorithms, hyperparameters, action spaces, and deployment timing.[46] Both papers argue that this level of symmetry is difficult to reconcile with fully autonomous algorithmic discovery. As den Boer et al. (2025) put it, "[t]his level of synchronization suggests the need for an explicit cartel agreement."[47] Grondin et al. (2025) similarly note that "in general, there is no reason why this [symmetry of hyperparameters] should be the case in general applications," and conclude that these requirements "cast doubt on the feasibility of achieving the exact same deployment without any prior agreement."[48], [49]

IV. COLLUSION OR CONFUSION? REGULATORY IMPLICATIONS

Having surveyed the latest academic literature, it is now possible to revisit the two questions posed in the introduction. On outcome, the answer is clear: simple learning algorithms tend to result in supracompetitive prices across Q-learning and context-free bandit settings. On mechanism, these studies point away from the "smart AI cartel" narrative and toward a family of design-level limitations—lack of counterfactual reasoning, imperfect exploration, negativity bias, and orchestrated standardization—that cause algorithms to fail to learn to compete rather than succeed in learning to collude. This in turn raises a provocative question: if supracompetitive prices stem from algorithmic limitations rather than from strategic sophistication, should that simply be regarded as incompetence?

An emerging line of scholarship suggests the answer is not so straightforward, and that the apparent lack of sophistication of these algorithms may itself be analytically meaningful. Carissimo et al. (2025) and Hartline et al. (2024, 2025)[50] both ground their proposed regulatory frameworks in a core economic principle familiar to economists and legal practitioners alike: a competitive firm should act in its unilateral interest. In practical terms, this means a firm's pricing algorithm should behave consistently with unilateral best responses to its competitors, regardless of how those competitors behave. From this shared premise, the two sets of authors develop complementary proposals. Carissimo et al. (2025), building on their meta-game analysis discussed earlier, argue that "… regulators should start by out-ruling algorithm designer collusion, i.e. algorithms that are being checked must be shown to be parametrized to maximize payoffs unilaterally, or at least be tuned to that goal." Hartline et al. (2024, 2025) go

44 Carissimo et al (2025), *supra* note 43 ("We have learned from the literature that price collusion between two (and more) learning algorithms is possible. In particular, it is shown to be possible in certain ranges of the parameter spaces of the learning algorithms. What is not known from the literature is whether algorithmic collusion would naturally arise from algorithms being designed by competitors or by allies…. we study the incentives of algorithm designers by formulating a meta-game where the strategies of the players, that is, of the algorithm designers are the hyperparameters that they pick for their algorithms which then compete in an underlying repeated game.")

45 Carissimo et al (2025), *supra* note 43 ("Our main finding is that algorithmic collusion can only strive if algorithms are orchestrated, that is, co parametrized non competitively in the pursuit of supra competitive prices and joint payoff improvements. By contrast, algorithms set competitively will end up pricing close to competitively.")

46 den Boer et al. (2025), *supra* note 16. Grondin et al. (2025), *supra* note 16.

47 den Boer et al. (2025), *supra* note 16. ("Competitors are committed to use the same Q-learning algorithm, which start at the same moment, with the same hyper-parameters and action spaces, while it is outperformed by the first alternative pricing rule. This level of synchronization suggests the need for an explicit cartel agreement … More extensive experiments of this kind are conducted by Eschenbaum, Mellgren and Zahn (2022). However, these still do not entirely eliminate the need for coordination, as the algorithms are, for example, assumed to have the same hyperparameters, start at the same time and/or have the same state and action spaces.'").

48 Grondin et al. (2025), *supra* note 16. ("As compared to previous studies, this indicates that algorithmic collusion has a large dependence on the symmetry of the hyper-parameters. In general, there is no reason why this should be the case in general applications. A possible argument is that several companies could be using the same pricing algorithm, however, our results cast doubt on the feasibility having the exact same deployment without any prior agreement.")

49 Other studies also shift attention to coordinated algorithmic design itself. See, for example, Nicolas Eschenbaum, Filip Mellgren & Philipp Zahn, *Robust Algorithmic Collusion* (Working Paper, 2022), arXiv:2201.00345, https://arxiv.org/abs/2201.00345 ("Our findings suggest that policy-makers should focus on firm behavior aimed at coordinating algorithm design in order to make collusive policies robust."). See more discussion below.

50 Jason D. Hartline, Sheng Long & Chenhao Zhang, *Regulation of Algorithmic Collusion*, in *Symposium on Computer Science and Law (CSLAW '24)*, Mar. 12–13, 2024, Boston, Mass. (ACM 2024), https://doi.org/10.1145/3614407.3643706. Jason D. Hartline, Chang Wang & Chenhao Zhang, *Regulation of Algorithmic Collusion, Refined: Testing Pessimistic Calibrated Regret*, in *Proc. of the 2025 Symposium on Computer Science and Law (CSLAW '25)* (ACM 2025), https://doi.org/10.1145/3709025.3712217.

further, advocating a bright-line rule under which deploying a pricing algorithm that fails to best respond under their proposed definition[51] is legally sufficient for liability.[52]

The implications of such proposals are significant. If the threshold for non-collusive behavior requires an algorithm to satisfy a best-response benchmark of this kind, relatively little room would remain for technological imprecision or prolonged learning phases — and an "algorithmically confused" algorithm would, by construction, be more likely to fail the benchmark. Viewed in this light, the line between confusion and collusion may be less clear-cut than it first appears: the very features that make an algorithm look "dumb" — rigid exploration schedules, deterministic convergence, identical parametrization across competitors—are also the features that, under these emerging frameworks, would be hardest to defend as the product of independent, unilateral profit-seeking. This is a question that the literature is only beginning to address.

V. CONCLUSION

This article has surveyed a recent and rapidly growing body of academic literature that challenges the prevailing narrative about autonomous collusion by first-party algorithms. The conventional concern—that increasingly sophisticated AI will inevitably learn to collude — finds limited support in the studies reviewed here. The evidence so far instead suggests a counterintuitive pattern: it is simpler, less sophisticated algorithms that are more prone to producing supracompetitive prices, while more advanced algorithms tend to compete more aggressively.

Several limitations warrant emphasis. The studies surveyed are based almost exclusively on stylized models and simulations involving small numbers of firms, homogeneous products, and simple market structures, and they focus predominantly on Q-learning and a limited set of alternatives. Whether the findings generalize to richer market environments, to non-price dimensions of competition, or to newer algorithm classes remains an open question.

Even so, the literature — nascent as it is — carries important lessons for company executives, attorneys, and policymakers.

For companies and their counsel, the research suggests that algorithm design choices are not competitively neutral. Simpler algorithms might be more prone to supracompetitive pricing, while also underperforming more sophisticated alternatives. Companies may therefore wish to invest in algorithmic sophistication and to ensure that their algorithm's design reflects independent unilateral profit-seeking objectives. In addition, an emerging proposal by academic researchers is to treat algorithmic *symmetry* among competitors as an indicator of explicit coordination. Companies and their counsel would do well to at least be attentive to whether their algorithmic design choices mirror those of competitors.

For policymakers and regulators, the findings present a tension worth monitoring. The literature reviewed here suggests that the risk of supracompetitive pricing may be greater with simpler, less sophisticated algorithms than with more advanced ones. If regulatory or enforcement postures create an environment in which companies perceive the deployment of pricing algorithms — particularly sophisticated ones — as inherently inviting scrutiny, companies may gravitate toward simpler algorithmic tools that are less conspicuous but, based on the latest literature reviewed here, more prone to producing the very supracompetitive outcomes that regulators seek to prevent.

How to navigate these cross-cutting pressures is the central challenge the next wave of research and enforcement practice will need to address.

51 Hartline et al.'s notion of "best response" asks whether the observed pricing behavior can be plausibly explained as the outcome of a firm unilaterally attempting to maximize its own payoff, given what it observed. This is stronger than simply asking whether the algorithm avoided obvious mistakes in hindsight; it is meant to rule out pricing behavior that persistently leaves profitable unilateral adjustments on the table or can be predictably exploited by the market. Hartline et al (2024, 2025), *supra* note 50.

52 Hartline et al (2025), *supra* note 50. ("Being unable to pass the test is a violation of their [Hartline et al, 2024] suggested per se rule." See also Table 1, classifying "requiring passing data audit (Hartline et al. [23] and this paper)" as "per se.")

CPI COMPETITION POLICY® INTERNATIONAL

CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.