



HAL
open science

Deciphering Algorithmic Collusion: Insights from Bandit Algorithms and Implications for Antitrust Enforcement

Frédéric Marty, Thierry Warin

► To cite this version:

Frédéric Marty, Thierry Warin. Deciphering Algorithmic Collusion: Insights from Bandit Algorithms and Implications for Antitrust Enforcement. *Journal of Economy and Technology*, 2024, 3, pp.34-43. <10.1016/j.ject.2024.10.001>. <halshs-04745409v1>

HAL Id: halshs-04745409

<https://shs.hal.science/halshs-04745409v1>

Submitted on 3 Jan 2025 (v1), last revised 3 Mar 2026 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Deciphering Algorithmic Collusion: Insights from Bandit Algorithms and Implications for Antitrust Enforcement

Frédéric Marty, Thierry Warin



PII: S2949-9488(24)00051-9

DOI: <https://doi.org/10.1016/j.ject.2024.10.001>

Reference: JET48

To appear in: *Journal of Economy and Technology*

Received date: 28 August 2024

Revised date: 16 October 2024

Accepted date: 17 October 2024

Please cite this article as: Frédéric Marty and Thierry Warin, Deciphering Algorithmic Collusion: Insights from Bandit Algorithms and Implications for Antitrust Enforcement, *Journal of Economy and Technology*, (2024) doi:<https://doi.org/10.1016/j.ject.2024.10.001>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 The Author(s). Published by Elsevier B.V. on behalf of KeAi Communications Co., Ltd.

Deciphering Algorithmic Collusion: Insights from Bandit Algorithms and Implications for Antitrust Enforcement

Frédéric Marty

CNRS – GREDEG – Université Côte d’Azur

CIRANO, Montréal

frederic.marty@gredeg.cnrs.fr

Thierry Warin

HEC Montréal

CIRANO, Montréal

thierry.warin@hec.ca

Abstract

This paper explores algorithmic collusion from both legal and economic perspectives, underscoring the increasing influence of algorithms in firms’ market decisions and their potential to facilitate anti-competitive behavior. By employing bandit algorithms as a model—typically used in uncertain decision-making scenarios—we shed light on the mechanisms of implicit collusion that occur without explicit communication. Legally, the primary challenge lies in detecting and categorizing possible algorithmic signals, particularly when they function as unilateral communications. Economically, the task of distinguishing between rational pricing strategies and collusive patterns becomes increasingly complex in the context of algorithm-driven decisions. The paper stresses the need for competition authorities to identify atypical market behaviors. Striking a balance between algorithmic transparency and the prevention of collusion is critical. While regulatory measures could mitigate collusive risks, they might also impede the development of algorithmic technologies. As this form of collusion gains prominence in competition law and economics discussions, understanding it through models like bandit algorithms becomes essential, especially since these algorithms have the potential to converge more rapidly toward supra-competitive prices equilibria.

Keywords:

Algorithmic Collusion, Machine-Learning, Bandit Algorithms, Antitrust Enforcement, Pricing Strategies

JEL codes: K21, L13, L41

I – Introduction

Algorithmic collusion, characterized by the confluence of traditional collusion dynamics and the innovative dimensions of algorithmic pricing in general and artificial intelligence in particular, has emerged as a pressing concern in the contemporary competition law enforcement conversation (Ezrachi and Stucke, 2017. Hansen et al., 2021). Within this landscape, this paper investigates the role of bandit algorithms in digital markets and their implications for collusive behavior.

Defining how self-learning algorithms could inadvertently (e.g. without explicit and coordinated instructions to do so) converge towards supra-competitive equilibria raises significant concerns for antitrust enforcement. Many papers have addressed the possibility of self-reinforcing machine learning algorithms converging towards supra-competitive prices (for a synthesis of economic literature see Marty and Warin, 2023). Notably, Calvano et al. (2020) demonstrated that Q-learning algorithms implemented independently by competing firms, without instructions to collude, can progressively (through a tit-for-tat mechanism) learn to cooperate, identify a stable equilibrium of tacit collusion, and eventually return to it in the event of an external shock. These results have been criticized from various angles. First, convergence towards supra-competitive prices does not necessarily mean the establishment of a collusive equilibrium: it can result from the gradual substitution of exploitation decisions for exploration decisions during machine learning (Abada et al., 2024). Second, the process is made more complex by the asynchronous nature of algorithm revisions (Asker et al., 2024). Third, collusive effects may be significantly more pronounced through training data than through data resulting from interactions (Eschenbaum et al., 2022). Fourth, achieving a supra-competitive price equilibrium is slow to emerge in the context of Q-learning algorithms (Cho and Williams, 2024).

Our contribution shows that bandit-type algorithms can partially address the last criticism by ensuring a faster learning speed than Q-learning. Additionally, our paper underlines that because of their intrinsic characteristics, Bandit-algorithms may also lead indirectly using-firms, to indirectly exchange information by making their pricing strategy more decipherable. By doing so, the interplay of Bandit-algorithms may be not only analysed under the lens of tacit collusion but also under the one of collusion through signaling. Such price signals may take the forms of unilateral communications or collusive offers (by adopting a conduct that can be interpreted as a cooperative one by the competitors) or the one of an artificial transparency. By making its algorithm decipherable to its competitors, a firm may drastically reduce the radical uncertainty associated to a perfect competition framework by making its future behavior predictable (see Salcedo, 2015). The case of the collusion through signals has been investigated by Marty and Warin (2024). Such a strategy makes the identification of a collusive focal point easier and faster to attain.

However, as soon as a price movement can be characterized as a unilateral act of communication or as soon as an artificial transparency can be spotted, the sanction through competition law provisions is highly probable. Such a behavior may be seen as pertaining to a facilitating practice, e.g. as an offer of a collusion contract. However, the sanction is not ensured as the signal may be hidden behind dynamic and personalized price patterns. The signal should be detected by competitors but must remain hidden to antitrust enforcers. Bandit algorithms may play a significant role here both for emitting and detecting even stealth signals. To this respect, the issue at stake echoes the ones of markets manipulations in the high frequency trading. The regulatory challenge consists in providing the market regulator efficient monitoring tools to detect abnormal price movements and eventually to sanction them.

In this contribution, we just provide some insights on this second aspect of the question (see Marty and Warin (2024) for a comprehensive view). We focus on the debate on the capacity of A.I. based pricing algorithms to reach supra-competitive prices equilibria without communications e.g. without relying on facilitating practices. We wonder to what extent the collusive capacities of Bandit algorithms may answer to some of the criticisms addressed to bot-led scenario based on Q-learning algorithms as described above.

Indeed, our research delves into the implications of bandit algorithms in this context. The bandit algorithm, rooted in probability theory and statistics, addresses the exploration-exploitation dilemma, determining optimal strategies under uncertainty (Agrawal, 1985; Auer, 2002). In the realm of algorithmic collusion, the bandit approach can elucidate how firms might balance between exploring new pricing strategies and exploiting known beneficial ones, especially when navigating the landscape of unilateral signals. The specificity of bandit algorithms compared to self-reinforcing machine learning lies in their unique design to optimize decision-making under uncertainty, balancing exploration and exploitation. This is a departure from self-reinforcing ML algorithms that predominantly learn from and reinforce patterns in historical data. Bandit algorithms, as exemplified in the work of Agrawal and Goyal (2012), excel in environments where they must make decisions with minimal prior data, adapting continuously based on immediate feedback. This adaptability, however, introduces significant collusive risks and supervisory challenges. As already demonstrated by Waltman et al. (2008) and later by Calvano et al. (2020) about Q-learning, algorithms might inadvertently learn collusive patterns (or more probably to supra competitive price equilibria), presenting difficulties for competition authorities to discern competitive behaviors from collusive ones due to their dynamic nature.

While bandit algorithms exhibit unique traits that facilitate faster convergence to supra-competitive equilibria, it is crucial to compare their performance and risks with other machine learning models, particularly reinforcement learning (RL) and deep learning (DL). Reinforcement learning models, such as Q-learning, are more complex and state-dependent, which makes them capable of handling more intricate market environments but at the cost of slower convergence and higher computational demand. In contrast, bandit algorithms thrive in simpler, stateless scenarios where they can make decisions based on partial information, quickly adapting their strategies. However, RL algorithms may have a higher propensity for discovering innovative, albeit more complex, collusive strategies in static or offline settings, making them more difficult to regulate.

Deep learning models, especially in the context of multi-agent environments, introduce additional layers of complexity by learning intricate patterns through large datasets. While DL models excel in handling vast and high-dimensional data, their reliance on historical data and slower adaptability can make them less prone to immediate collusive behavior compared to bandit algorithms. However, the opacity and complexity

of DL models pose significant challenges for regulatory oversight, as understanding the decision-making process becomes increasingly difficult for regulators.

In terms of susceptibility to collusion, bandit algorithms are particularly prone to tacit collusion due to their real-time adaptability and ability to signal pricing strategies indirectly through trial and error. RL models, while capable of reaching collusive outcomes, generally take longer to converge and require more structured environments. Comparatively, DL models, although powerful, are less likely to engage in real-time collusion but could create more persistent and hard-to-detect anti-competitive strategies over time. From a regulatory perspective, the ease of detecting collusive behavior diminishes as the algorithmic complexity increases, with bandit algorithms being easier to monitor due to their more transparent decision-making processes, whereas DL and RL models pose greater challenges for detection and interpretation.

It is technically plausible that bandit algorithms can converge faster than self-reinforcing machine learning models in certain contexts. This can be attributed to the fundamental operational differences between these two types of algorithms. Bandit algorithms are designed for optimal decision-making in environments with uncertain and variable outcomes. Their strength lies in their ability to quickly adapt to new information, adjusting their strategies based on immediate feedback. This adaptability is particularly evident in environments where real-time decision-making is crucial and where available data may be sparse or non-stationary. In contrast, self-reinforcing machine learning models typically rely on large datasets and often seek to identify and reinforce patterns based on historical data. While powerful in scenarios with abundant data and well-defined patterns, these models might require more time to converge in environments where the data is continually changing or where immediate feedback is integral to the decision-making process.

Numerical simulations can indeed be used to demonstrate this difference in convergence rates. By simulating environments where conditions rapidly change or where only limited data is available, one can observe how bandit algorithms adjust their strategies more swiftly to optimize outcomes. In contrast, self-reinforcing ML models might take longer to identify and adapt to these changes due to their reliance on historical data patterns. This faster convergence of bandit algorithms can be particularly relevant in dynamic market environments, such as digital markets, where conditions frequently change, and immediate decision-making is crucial. However, it is important to note that the speed of convergence does not necessarily imply superiority in all aspects. While bandit algorithms may excel in rapidly changing environments, self-reinforcing ML models might be more effective in stable environments with vast historical data. In this context, it is necessary to consider the broader implications of algorithmic collusion, including how these automated processes can create challenges for traditional competition law frameworks, complicate the detection of anti-competitive behaviors, and potentially necessitate new regulatory approaches to address the unique risks posed by algorithm-driven markets.

The overall contribution and positioning of the paper revolve around demonstrating how bandit algorithms can amplify existing risks of collusion in digital markets. The paper delves into the formal model underlying these algorithms to illustrate how such risks are heightened compared to standard ML algorithms. It argues that bandit algorithms not only perform better in real-time decision-making but also present greater challenges in interpretability and regulatory oversight, making them more difficult to decipher for competition authorities and potentially exacerbating undetected collusion risks. In terms of positioning in relation to existing literature, the paper distinguishes itself by investigating whether other studies have addressed similar questions using bandit algorithms, particularly in the context of antitrust implications. Additionally, examining literature from other fields, such as finance, where bandit algorithms are applied

for signal detection, provides further insights. This cross-disciplinary review helps in comprehending the functioning of bandit algorithms in different contexts, shedding light on potential implications for collusive risks and regulatory challenges in digital markets.

This paper contributes significantly by exploring the unique characteristics of bandit algorithms, their potential to heighten collusive risks in digital markets, and the challenges they pose for regulatory oversight, thus addressing a critical gap in the existing academic discourse. Subsequent sections will delve into the intricacies of how unilateral collusive contract offers might be orchestrated in oligopolistic settings and the potential countermeasures competition regulators might deploy. Section 2 suggests that the ability to achieve collusive equilibria can be enhanced using bandit algorithms. Section 3 outlines the challenges in terms of enforcing competition rules to characterize these trajectories as anti-competitive and to sanction them. Section 4 concludes.

II – Augmented collusive capacities through algorithms? Investigating the case of bandit-algorithms between collusive signals and tacit collusion scenarios

Recent contributions, such as those by Xu, Lee, et Tan (2023) and Dorner (2021), have provided deeper insights into the dynamics of algorithmic collusion. They have emphasized the role of recommendation algorithms, the potential of self-learning algorithms in real markets, and the credible threat posed by such collusion, respectively. The literature on tacit collusion, particularly when mediated by algorithms, is vast and multifaceted. A significant portion of this literature has delved into the role of complex machine learning algorithms in facilitating such collusion. Notably, Ezrachi and Stucke (2017), Mehra (2016), and Calvano et al. (2019) have posited that self-learning algorithms can spontaneously converge to a tacit collusive equilibrium, even if their initial parameters differ. This convergence occurs without any direct or indirect communication between the firms, thereby challenging traditional methods of detecting and prosecuting anti-competitive behaviour. Recent studies, such as those by Harrington (2018), Schwalbe (2018), Calvano et al. (2019, 2020), and Assad et al., (2023), have further emphasized the competitive risks posed by pricing algorithms. The consensus is growing around the potential of these algorithms to autonomously reach a tacit collusive equilibrium without explicit communication. This blurs the distinction between parallel behaviour and concerted practices (Calvano et al. 2020).

The integration of algorithms models in market collusion presents a complex landscape. Algorithms, particularly those akin to bandit algorithms, can serve as communication conduits, emitting signals discernible to competitors but potentially elusive to competition authorities. The pivotal question centers on the ability of algorithmic signals to act as unilateral collusion invitations and acceptance indicators, all while circumventing the appearance of mutual agreement information exchange. Moreover, the capacity of algorithms, especially those based on the multi-armed bandit framework, to amplify signal clarity and filter out disruptive "noise" is of paramount importance. Collusion can either emerge from sheer unilateral transparency without explicit information exchange or from bilateral transparency where information is reciprocally shared. Such maneuvers, from a competition law lens, might be construed as anti-competitive. Introducing artificial transparency might be perceived as a collusion intent signal, especially in parallel behavior contexts.

In terms of collusion through signals, several scenarios can be envisioned (Marty and Warin, 2024). One involves a professional association mitigating uncertainty regarding competitors' future actions. Another sees a benchmark firm signaling industry median firm preferences. A third scenario involves a Stackelberg leader, a dominant entity seeking alignment with peripheral competitive fringe firms. These fringe firms, as Gal (2019) illustrates, can adopt cooperative stances via pricing algorithms. A leader might employ an advanced pricing algorithm, while competitors opt for simpler, leader-adaptive algorithms using identical databases, expediting convergence. In this milieu, algorithms, especially those resembling the bandit algorithms, can extend cooperation invitations to competitors. This doesn't necessarily culminate in formal cartel formation. Instead, the algorithmic signals aim to pinpoint coordination focal points through iterative processes, hinging on competitor feedback. This can be analogized to a unilateral contract offer, with competitors potentially reciprocating by adjusting prices. Absent such reciprocation, the initiating firm might foresee a unilateral price hike resulting in sales plummeting.

Economically, this can be dissected into two scenarios. In a first one, a Stackelberg leader signals future prices, indicating that undercutting would be perceived as a price war provocation. This strategy hinges on aggressive behavior threats. In another, the leader proposes a universally profitable price, emphasizing collective over individual profit. The leader's strategic intent dictates the scenario choice. The leader can either announce immediate prices or hint at future adjustments. Immediate changes should occur during low-sales periods, allowing competitor algorithms to swiftly detect and adapt. Algorithmic price adjustments streamline this process. Alternatively, price shifts can be postponed, with unilateral price hikes contingent on unanimous future price direction changes by all firms. However, volatile price movements can introduce noise, complicating signal interpretation. Algorithmic pricing can lead to complex price dynamics and the difficulty for both competitors and regulators in discerning whether a particular pricing pattern is the result of competitive behavior or a signal of collusion (Calvano et al. 2019).

Bandit algorithms, commonly employed in decision-making scenarios that involve balancing exploration and exploitation, offer a valuable framework for understanding the dynamics of algorithmic collusion. These algorithms, particularly in their ability to encourage market participants to converge toward supra-competitive prices, depend on how effectively they manage this trade-off. As self-learning mechanisms, bandit algorithms enhance their exploration strategies by associating rewards with specific decisions, dynamically minimizing statistical regret through the balance between exploitation (decisions yielding immediate gains) and exploration (decisions that may involve a temporary loss but could reveal higher long-term gains). This process operates within a framework of partial information, meaning that there is no requirement to observe competitors' actions directly, potentially mitigating concerns over coordinated anti-competitive practices.

Several variants of bandit algorithms—such as epsilon-greedy, Thompson Sampling (TS), and Upper Confidence Bound (UCB)—share these key characteristics. When tested through numerical simulations, these algorithms have demonstrated a higher convergence to supra-competitive prices compared to Q-learning algorithms (Sadoune et al., 2024). In particular, UCB algorithms have shown a greater propensity to lead firms toward supra-competitive price equilibria, especially in homogeneous market settings with fewer participants (Sadoune et al., 2024). Specifically, UCB's ability to foster collusion emerges from its efficient balancing of exploration and exploitation, which allows firms to sustain high collusive prices while minimizing deviations from Nash equilibrium.

Interestingly, the study found that UCB agents consistently managed to coordinate toward collusion with a 60% frequency of collusive price strategies, especially in simpler settings such as two-player homogeneous markets. In contrast, epsilon-greedy and TS algorithms demonstrated a stronger tendency to converge to Nash equilibria rather than to collusion, with UCB agents achieving collusion 100% of the time in simpler market configurations, whereas epsilon-greedy and TS agents did not (Sadoune et al., 2024). As the market becomes more heterogeneous or the number of players increases, the likelihood of achieving collusion decreases, with market heterogeneity impeding the collusive potential of UCB more significantly than the number of players.

These findings align with previous literature confirming the tendency of bandit algorithms, and UCB in particular, to lead to supra-competitive price equilibria in online auction settings (Aoud & van den Boer, 2021; Hansen et al., 2021). All these algorithms adaptively learn and make decisions based on partial feedback, offering insights into how firms adjust pricing strategies in response to competitors' actions. This adaptability is especially significant in real-time decision-making environments where data availability may be limited or non-stationary (Agrawal & Goyal, 2012). However, the potential for bandit algorithms to foster collusive equilibria is not merely theoretical. Their capacity to systematically explore and exploit pricing strategies can inadvertently result in tacit collusion. This occurs as algorithms identify mutually beneficial pricing strategies through iterative learning processes, mirroring the decision-making dynamics firms undergo in competitive markets (Bubeck & Cesa-Bianchi, 2012).

In the realm of competition law, understanding the role of algorithms in fostering and maintaining collusive equilibria is crucial. Collusion can manifest through explicit signals that foster mutual understanding among competitors or through AI algorithms that spur spontaneous tacit collusion. Cooper et al. (1989) posited that unilateral communication, particularly when aligned with collective interests, simplifies the path to collusive equilibrium compared to bilateral communication, thereby enhancing the likelihood of mutually beneficial outcomes. The iterative nature of such interactions strengthens the credibility of unilateral declarations by a leading entity. If the leader's commitments do not align with subsequent actions or send misleading signals, their reputation could suffer, potentially undermining future credibility. This ongoing interaction builds trust, and the willingness of a leader to take unilateral risks underscores their commitment to cooperation.

In the context of signaling models, algorithmic pricing can foster collusive tendencies by creating a form of artificial transparency. This is notable in the case of pricing algorithms based on self-reinforcing machine learning. While parallel algorithms may facilitate collusive equilibria through tit-for-tat strategies as observed by Axelrod and Hamilton (1981), self-learning algorithms might offer even more advantageous outcomes. This aspect is elaborated by Hingston and Kendall (2004), who demonstrate how these algorithms can adapt and optimize strategies for competitive advantage. Crandall et al. (2018) empirically show that autonomously learning algorithms could develop enduring cooperative relationships with both machines and humans. However, the debate about the genuine cooperative capacity of AI algorithms persists. Notably, tacit collusion often relies on intuition, cultural norms, and the ability to decipher subtle signals from competitors. In this setting, non-binding signals, or "cheap talks," become crucial.

Yet, algorithm-driven collusion introduces significant complexities in competition law. Strategies rooted in creating unilateral transparency for competitors, without direct exchanges or overt agreements, challenge traditional competition law enforcement standards. This aspect demands careful legal scrutiny. In essence, algorithms, regardless of their nature, have the potential to catalyze the formation of collusive equilibria. Algorithms that signal, in particular, are intriguing from a competition law perspective, raising questions

about strategies rooted in unilateral transparency without overt exchanges. Thorough research is essential to unravel the intricacies of algorithmic collusion in the context of competition law.

Shifting focus to the bandit algorithm, this technique, often associated with multi-armed bandit problems, addresses decision-making amid uncertainty. Originating from a hypothetical scenario where one must choose the most rewarding slot machine arm to pull, the bandit algorithm confronts the exploration-exploitation dilemma: whether to explore all options to understand their reward distribution or exploit the most rewarding option based on current knowledge. This dilemma is addressed in various strategies like ϵ -greedy, Upper Confidence Bound (UCB), and Thompson Sampling, as detailed by Sutton and Barto (2018).

Bandit algorithms find diverse applications, from guiding patient treatments in healthcare to managing resource allocation in web services. In online advertising, as Li et al. (2010) highlight, these algorithms determine which ads to display to maximize click-through rates. In algorithmic pricing, a domain where bandit algorithms particularly excel, each pricing strategy is an 'arm', with the 'reward' being the resultant profit. Babaioff et al. (2015) demonstrate how these algorithms navigate consumer response uncertainties, balancing exploration of new pricing strategies with exploitation of known profitable ones.

In the formal model of the multi-armed bandit problem, an agent interacts with an environment over time, choosing actions to maximize cumulative rewards. The goal is to minimize regret, defined as the difference between potential and actual rewards. Various algorithms, including ϵ -greedy, UCB, and Thompson Sampling, tackle this problem by balancing exploration and exploitation. UCB1, for instance, is a heuristic approach that achieves logarithmic total regret, indicative of its efficiency, as shown by Auer, Cesa-Bianchi, and Fischer (2002).

Bandit algorithms differ significantly from traditional ML algorithms in real-time adaptability, efficiency in balancing exploration and exploitation, and less reliance on extensive datasets. These attributes, explored by researchers like Agrawal and Goyal (2012) and Bubeck and Cesa-Bianchi (2012), make them uniquely suited for dynamic environments. However, their evolving nature also complicates the interpretation of their strategies, especially in regulatory and antitrust contexts, as discussed by Harrington (2018). This complexity necessitates advanced monitoring techniques to effectively detect collusion and interpret the underlying intentions of these algorithms.

Q-learning, a model-free reinforcement learning algorithm, is designed to determine the best action to take in a given state. It functions without a model of the environment, hence its designation as 'model-free', and it is capable of handling environments with stochastic transitions and rewards. The principle behind Q-learning is the estimation of the value of a state-action pair, or Q-value, which reflects the expected utility of taking a specific action in a given state and then adhering to the optimal policy thereafter (Watkins & Dayan, 1992). This methodology enables Q-learning to effectively address complex decision-making tasks in environments with multiple states and actions.

Bandit algorithms, in contrast, are more straightforward and are primarily employed in multi-armed bandit problems. These problems involve an agent selecting from several choices, each with unknown and potentially varied reward distributions. The core challenge in these scenarios is balancing exploration (experimenting with different choices to discover their rewards) and exploitation (opting for the choice that has historically offered the best rewards). Bandit algorithms are typically used in simpler, stateless contexts, making them distinct from Q-learning, which factors in the state of the environment in its decision-making process. Their applications include areas like A/B testing, adaptive routing, and personalized

recommendations, where the primary objective is to maximize rewards in situations of uncertainty (Auer, Cesa-Bianchi, & Fischer, 2002). The contrast between Q-learning and bandit algorithms can be seen in several dimensions. Q-learning is more complex and appropriate for state-dependent decision-making environments, while bandit algorithms are simpler and designed for stateless decision-making contexts. Q-learning algorithms consider the state of the environment when making decisions, whereas bandit algorithms do not. Both approaches address the exploration-exploitation trade-off, but they do so in different contexts and with varying degrees of complexity. Q-learning is part of a broader suite of reinforcement learning algorithms and is suited for more complex tasks in structured environments. Bandit algorithms, on the other hand, focus on simpler reward maximization problems.

Bandit algorithms offer a significant advantage over reinforcement learning models in that they can learn effectively in environments with partial information. Unlike reinforcement learning models, bandit algorithms only require knowledge of the outcome associated with the selected choice, rather than the outcomes of all possible choices. This allows them to operate without the need for constructing counterfactuals (Hartline et al., 2024). Additionally, bandit algorithms, which are easier to implement and more widely adopted than Q-learning, are particularly suited for contexts of partial information due to their higher rate of exploration decisions (Aouad and van den Boer, 2021). In contrast, Q-learning algorithms tend to be used less frequently in practical operational settings, as their requirements for fuller information and more complex state representations make them less adaptable to certain real-world applications. This discussion aligns with the critiques directed at the model proposed by Calvano et al. (2020), particularly regarding several key issues. Among these are the asynchronous versus synchronous nature of algorithm updates, the relative importance of training data compared to interaction data, and the inherent tendency of Q-learning algorithms to progressively shift from exploration to exploitation decisions. Furthermore, the speed of convergence in these algorithms has also been a point of contention.

Bandit algorithms address several of the critiques leveled against Q-learning based collusive scenarios, particularly those highlighted in the ongoing academic debate:

- **Asynchronous vs. Synchronous Updates:** Bandit algorithms inherently operate in an environment where decisions are made in a stateless context, meaning that each decision is made independently of the previous state. This characteristic allows bandit algorithms to function effectively in environments where data is asynchronously updated, as they do not rely on a comprehensive model of the environment. In contrast, Q-learning, which operates within a state-dependent framework, often requires synchronous updates to ensure consistency across states, which can lead to challenges in dynamic environments where data changes asynchronously (Biancho and Mantegazza, 2023).
- **Training Data vs. Interaction Data:** Bandit algorithms are designed to optimize decision-making based on real-time feedback, which means they rely heavily on interaction data rather than pre-existing training data. This allows them to adapt quickly to changes in the environment, making them particularly effective in settings where the environment is dynamic, and the relevance of historical data is limited. In contrast, Q-learning algorithms typically rely more on extensive training data, which can slow their adaptation to new conditions if the data does not reflect current dynamics.
- **Exploration vs. Exploitation Balance:** Bandit algorithms are specifically designed to manage the exploration-exploitation trade-off efficiently. They employ strategies such as epsilon-greedy, Upper Confidence Bound (UCB), or Thompson Sampling to continuously balance the need to explore new

strategies against the need to exploit known profitable strategies. This allows for a more fluid transition between exploration and exploitation, unlike Q-learning algorithms, which can become increasingly exploitation-focused over time as they converge, potentially missing out on better strategies that might have been found through further exploration.

- **Speed of Convergence:** One of the primary advantages of bandit algorithms is their ability to converge quickly to an optimal strategy, particularly in environments with limited data or rapidly changing conditions. The simplicity of their decision-making process, which does not require modeling the entire environment, enables them to adjust their strategies in real-time, often leading to faster convergence compared to Q-learning algorithms. Q-learning, due to its dependence on state-action pairs and the need to update Q-values across potentially large state spaces, can experience slower convergence, particularly in complex or non-stationary environments.

These characteristics make bandit algorithms a robust alternative to Q-learning in scenarios where the limitations of Q-learning—such as the requirement for synchronous updates, heavy dependence on extensive training data, and slower convergence—pose significant challenges. Q-learning may, however, be more suited to offline training environments, where the goal is to uncover innovative strategies in highly complex market conditions (Werner, 2023). Nonetheless, the legal implications of algorithmic coordination are far from straightforward. Unilateral algorithmic behavior must carefully avoid crossing into overt collusion or anti-competitive practices that violate competition laws. It remains crucial for competition authorities to rigorously examine market signals and actions to detect potential anti-competitive behaviour. Bandit algorithms are particularly suited to the study of algorithmic collusion due to their capacity to navigate uncertainty and limited information, hallmarks of market environments. These algorithms adeptly balance the exploration of new strategies with the exploitation of known profitable ones, mirroring the decision-making process firms undergo when setting prices amidst opaque competitor actions and market responses. In the realm of adaptive strategies, bandit algorithms stand out for their ability to iteratively learn which actions yield the highest rewards. This attribute is crucial in market settings, allowing pricing algorithms to evolve towards profit-maximizing behaviors that could inadvertently result in tacit collusion, all without any explicit coordination between competing entities. The potential of bandit algorithms to decode signals from market competitors is particularly pertinent. They are capable of interpreting market fluctuations as indicators of the strategic shifts of rivals, enabling them to adjust their pricing strategies in a responsive manner. This capability is central to understanding how firms might employ subtle pricing cues to align on more profitable outcomes without direct communication, a concern for regulatory bodies. Moreover, bandit algorithms can systematically explore and potentially exploit collusive equilibria. Through a process of trial and error, they can identify pricing strategies that, when mirrored by competitors, lead to mutually beneficial, albeit collusive, outcomes. This exploration is key to understanding the pathways through which algorithms may autonomously converge on anti-competitive practices.

The ability of bandit algorithms to adapt in real-time further complicates regulatory oversight, as these algorithms can dynamically align pricing strategies without explicit communication, thereby evading traditional antitrust detection methods (Calvano et al., 2019). The dynamic nature of markets, characterized by shifting consumer preferences and varying external conditions, is where bandit algorithms truly excel. They are designed to adapt in real-time, constantly refining strategies to align with the current state of the market, ensuring efficiency and responsiveness in pricing decisions. The implications of bandit algorithms

extend into policymaking. A thorough grasp of how these algorithms function and the conditions under which they may lead to collusion is invaluable for regulators. Such knowledge is essential for developing measures that can preempt and mitigate the risk of anti-competitive behavior. For instance, regulatory bodies might need to implement continuous monitoring systems capable of detecting abnormal pricing patterns indicative of collusion, leveraging advanced data analytics and machine learning techniques.

Lastly, the deployment of bandit algorithms in setting prices also raises ethical questions regarding the responsibility of firms to prevent consumer harm. Insights gleaned from the study of these algorithms can contribute to the establishment of ethical standards for the application of artificial intelligence in commercial settings. Firms must balance the benefits of algorithmic efficiency with the need to maintain competitive integrity and consumer trust. Bandit algorithms offer a robust framework for probing the emergence and dynamics of algorithmic collusion. Their inherent learning and adaptive capabilities provide a lens through which one can examine the potential for autonomous systems to engage in market behaviors that undermine competitive integrity. Understanding these mechanisms is crucial for shaping business strategies and regulatory frameworks that uphold the principles of market competition. So, while algorithms, especially those akin to bandit algorithms, can potentiate coordination and cooperative equilibrium, signal interpretation challenges and competition law adherence are paramount. Algorithmic tools can augment coordination efficiency, but their deployment must respect legal and ethical boundaries to preserve market competition integrity.

In short, the deployment of bandit algorithms in pricing strategies can have significant economic implications, particularly concerning consumer welfare and market efficiency. By converging towards supra-competitive prices, bandit algorithms may reduce consumer surplus by elevating prices above competitive levels, leading to welfare losses. This effect is exacerbated in markets where price volatility—driven by the exploratory behavior of these algorithms—creates uncertainty for consumers, further diminishing their ability to make optimal purchasing decisions. Additionally, the potential for reduced competition and barriers to market entry could lead to inefficiencies in resource allocation, as fewer firms may find it viable to compete in markets dominated by algorithmic pricing strategies. The resulting deadweight loss, coupled with the erosion of competitive pressures that drive innovation, suggests a decrease in overall market efficiency. These factors underline the need for regulatory attention, as the economic harm facilitated by bandit algorithms can substantially distort market outcomes, resulting in both immediate and long-term consumer harm.

III. Addressing signaling collusions risks: assessing the legal challenges and augmenting competition authorities' capacities

The legal ramifications of algorithmic coordination are significant. Algorithms that signal, particularly those akin to bandit algorithms, pose challenges in discerning anti-competitive behaviour from legitimate market strategies. The iterative learning processes inherent in these algorithms can create conditions conducive to tacit collusion, where firms indirectly coordinate on pricing strategies without explicit communication (Harrington, 2018). Competition authorities must develop robust methodologies to detect and scrutinize these signals. Drawing parallels with financial market oversight, it is imperative for regulators to identify subtle signals and anomalous market patterns indicative of collusive behaviour. This requires continuous

surveillance and advanced analytical capabilities to evaluate the impact of algorithmic pricing on market dynamics (Evans, Tucker, & Fels, 2020).

From a legal perspective, discerning and evaluating collusive signals presents formidable challenges. Algorithm-driven collusion might encompass unilateral actions fostering transparency amongst competitors without manifest evidence of explicit agreements. This engenders queries regarding the categorization of such behaviour and its alignment with anti-competitive practices (Harrington 2018). Algorithms can enter a collusive state and charge supra-competitive prices without explicit communication. Tacit collusion, unlike explicit collusion, does not involve direct communication between parties to set prices or output. Instead, it occurs when businesses engage in practices that lead to a mutual understanding of shared market behaviour without explicit agreement, Antitrust authorities may, therefore, scrutinize such behaviours closely, despite the lack of explicit communication between firms. Understanding the subtleties of such market dynamics is crucial for both regulators and businesses to navigate the legal landscape effectively. There is a risk of a regulatory gap. If there is a convergence towards supra-competitive prices without any facilitating practice, without any exchange of information showing a concurrence of wills, it is hardly certain that the practice will be found to be anticompetitive. A sanction under US antitrust law would be highly unlikely, given the Supreme Court's decision-making practice (*Brooke Group Ltd. v. Brown & Williamson Tobacco Corp.*, 509 U.S. 209 (1993)). Even in other jurisdictions, the use of bandit algorithms, which rely on a mechanism for making decisions on the basis of partial information, may be difficult to sanction on the basis of a concerted practice. It is quintessential to differentiate between profit-maximising individual behaviour and anti-competitive intent, as competition regulations necessitate evidence of an anti-competitive motive rather than mere behavioural parallelism. In certain instances, bilateral transparency or unilateral declarations might be perceived as evidence of facilitating practices, potentially impeding effective market competition. Artificial transparency can be interpreted through industrial economics and game theory prisms. In narrowly oligopolistic markets, where a tacit collusion equilibrium might spontaneously materialize, curtailing uncertainty regarding market player future strategies can markedly truncate the trial-and-error phase and fortify the equilibrium by facilitating the identification of the coordination focal point (OECD 2012).

It's paramount to recognize that evaluating these factors necessitates meticulous assessment of specific circumstances, market conditions, and the underlying intentions of market participant actions. Competition authorities and courts play an instrumental role in analysing information exchange, transparency, and algorithmic behaviour impacts to ascertain if they constitute anti-competitive practices contravening competition law. A particularly dynamic field of economic literature has been developing in recent years, extending the scope and effectiveness of econometric tools for detecting abnormal market configurations using machine learning tools. This is the case for the detection of bid rigging in public procurement (Walliman et al., 2023), favoritism in public procurement (Gallego et al., 2021; Decarolis and Giorgiantonio, 2022), or also manipulations in electricity wholesale markets (Brown et al., 2023; Sun et al., 2022). In the context of enhancing market regulation in the era of complex algorithms, it is essential to focus on equipping authorities with effective market supervision tools. This includes providing them with the necessary technical means and imposing ex ante obligations akin to the practices in the financial sector, such as the registration of prices and conservation of orders. The utility of such practices in the financial world, where detailed records of transactions enable thorough ex-post investigations, suggests their potential effectiveness in monitoring algorithm-driven markets. Furthermore, the imposition of specific obligations on firms

regarding the risks associated with algorithms is crucial. Firms should be required to adopt a comprehensive compliance strategy. This strategy would entail not only ex ante certification of their algorithms but also a commitment to ongoing supervision of their effects, possibly through regular audits. This ensures continuous responsibility and awareness of the impacts of their algorithms.

Additionally, contemplating a shift in the burden of proof, adopting a 'comply or explain' logic, is significant even it raises significant legal concerns (De Cooman, 2023). This approach has parallels in the financial sector, particularly in regulatory compliance, where firms must adhere to regulations or explain their deviations. Translating this logic to the realm of algorithm-driven market practices could prompt firms to proactively ensure that their algorithms do not inadvertently facilitate anti-competitive behaviours. Adopting such measures would represent a substantial advancement in market regulation, addressing the unique challenges posed by algorithmic decision-making in the digital age.

The proposition of whether a unilateral algorithmic signal can be construed as constituting competitor communication is intricate. While there's no direct bilateral competitor communication in the conventional sense, one firm's algorithmic actions can still transmit a signal to other market competitors. Evans, Tucker, et Fels (2020) examine the problems posed by algorithmic collusion and discuss potential countermeasures, including the interpretation of unilateral algorithmic signals as a form of communication. In the context of tacit collusion, an operator's unilateral decision that contravenes its individual profit maximization and lacks economic rationale might be perceived as a collusion-facilitating factor. This premise is anchored in the notion that if the decision aligns with all market participants' collective self-interest if they concurred to act similarly but is antithetical to their self-interest if acted upon individually, it might indicate a coordinated strategy

However, the signal's efficacy is contingent on whether competitors' algorithms can capture and interpret it. If competitors possess advanced algorithms capable of real-time subtle signal detection and analysis, the unilateral signal might not remain undetected. In such scenarios, the signal can be as impactful as a financial market order containing concealed information. The deployment of intricate algorithms and data analysis techniques introduces challenges in determining communication extent and competition law implications. The ability of algorithms to capture and interpret signals might necessitate meticulous scrutiny by competition authorities to evaluate their potential anti-competitive effects and ascertain if they align with the communication scope that abets collusion.

While unilateral algorithmic signals don't encompass direct competitor communication, their influence on market behaviour and the capacity of competitors to capture and interpret them are factors warranting consideration in collusion analysis and competition law enforcement endeavors. Upon meticulous examination of the extant literature pertaining to algorithmic collusion and integrating insights from the field of computer science, it appears premature to recalibrate antitrust legislation to address the complexities associated with self-learning algorithms that may engage in collusion within actual market environments. Nonetheless, there exists a subset of algorithmic collusion—specifically, hub-and-spoke configurations enabled by centralized pricing algorithms—which may indeed necessitate prompt legislative intervention. Regulatory intervention in such cases may be considered necessary to prevent anti-competitive outcomes that can arise from these algorithmic arrangements. Adapting competition rules and guidelines that can

effectively address the nuanced and technical nature of such collusion without stifling innovation and the legitimate use of algorithms to improve market efficiency.

The phenomenon of algorithmic collusion presents intricate challenges for antitrust enforcement, necessitating a nuanced understanding of its multifarious dimensions. The implications of algorithmic signals and the potential perils of anti-competitive conduct warrant meticulous consideration. One of the pivotal tools in this exploration has been the utilization of bandit algorithms. Bandit algorithms, traditionally employed in decision-making scenarios where there's a trade-off between exploration (trying out new strategies) and exploitation (sticking with known strategies), offer a pertinent model for understanding the dynamics of algorithmic collusion. Their ability to adaptively learn and make decisions based on partial feedback makes them a representative model for how firms might adjust their pricing strategies in response to the actions of competitors. Directly, they provide insights into how algorithms can iteratively adjust to market signals, and indirectly, they shed light on the broader challenges of decision-making in uncertain environments.

To begin with, competition regulators must cultivate robust methodologies to discern and scrutinize price signals emanating from algorithms. Drawing parallels with financial market oversight, it becomes imperative for authorities to pinpoint subtle signals and anomalous market price patterns. This mandates relentless surveillance and the capability to identify departures from anticipated market dynamics. Subsequently, the onus of justification might be transitioned to enterprises employing algorithmic pricing paradigms. In scenarios where market conduct diverges from a predetermined model or manifests atypical price configurations, enterprises might be compelled to elucidate their maneuvers. Through rigorous data analysis from these firms, competition regulators can identify anomalies, subsequently placing the onus on enterprises to validate that their algorithmic operations don't culminate in anti-competitive consequences. Nevertheless, a mere inversion of the burden of proof might not be adequate to ascertain anti-competitive conduct, as the motivations and repercussions of algorithmic determinations warrant thorough assessment.

Furthermore, delineating between logical parallel pricing conduct and collusive configurations becomes increasingly intricate, especially within the ambit of unilateral communications. Unilateral signals, irrespective of their digital or traditional nature, engender complexities for both competition enforcers and corporate entities. Firms procuring insights regarding prospective competitor prices might grapple with distancing themselves from the tacit collusion proposition. This complexity is accentuated in the realm of AI-driven automated online pricing technologies, where pricing determinations are predicated on perpetual competitor price surveillance and automated recalibrations. The lucidity of pricing algorithms oscillates between being a proponent of competition and a potential collusion enabler, necessitating a judicious equilibrium.

Moreover, the facet of price transparency emerges as pivotal. Digital pricing strategies, frequently targeting price differentiation, are occasionally perceived as deliberately obfuscated. While competition regulators might advocate for transparency, unveiling pricing algorithms could inadvertently abet collusion. Striking an optimal balance between transparency and collusion deterrence is of paramount importance. In addressing the intricacies of algorithmic collusion, preemptively thwarting it without compromising algorithmic efficacy becomes a daunting endeavor. Instituting proactive regulations that curtail information assimilation and signal generation capabilities might inadvertently impede pricing algorithms, thereby

stifling innovation. Hence, the emphasis should be on specific attributes of information dissemination and assimilation that might bolster collusive equilibria, all while safeguarding the merits of algorithmic precision. The ascendancy of algorithmic collusion introduces profound challenges for antitrust enforcers. The tasks of detecting and interpreting algorithmic signals, discerning anti-competitive intentions and ramifications, and calibrating an equilibrium between transparency and collusion deterrence emerge as pivotal considerations in navigating these challenges. The insights derived from bandit algorithms further underscore the complexities and potential pathways in understanding and addressing algorithmic collusion.

The deployment of bandit algorithms in pricing strategies presents significant challenges for regulatory authorities, particularly due to their ability to quickly adapt and facilitate tacit collusion without explicit communication between firms. Current competition law frameworks, which focus on human-facilitated collusion, are often inadequate for capturing the complexities of algorithm-driven coordination. To address this, regulatory frameworks must evolve to recognize algorithmic decision-making as a form of implicit communication, even in the absence of direct agreements between firms. Expanding the legal definition of collusion to include algorithmic coordination—where autonomous systems align pricing strategies without overt interaction—would help regulators address the anti-competitive outcomes of these algorithms.

In practice, regulators could enhance their capabilities by implementing real-time data monitoring systems to track suspicious pricing behavior, including abnormal price patterns or synchronized pricing strategies that may indicate tacit collusion. Tools such as anomaly detection algorithms would be critical for identifying these patterns, particularly in fast-moving digital markets. Another effective measure would involve ex-ante algorithm certification, where firms are required to submit their algorithms for review before deploying them. Such reviews would help ensure that these algorithms do not foster collusion. Post-deployment audits could further provide regulators with detailed logs to assess whether pricing algorithms facilitated coordinated behavior.

To mitigate risks, a "comply-or-explain" mechanism could be introduced, shifting the burden of proof onto firms. This would require firms exhibiting suspicious pricing behavior to justify their actions, ensuring that their algorithms do not inadvertently promote collusion. Transparency requirements, where firms disclose the operational principles behind their algorithms and share real-time data on pricing decisions, would offer regulators the necessary insights to monitor and assess the potential for anti-competitive behavior.

Additionally, regulators could benefit from enhanced international coordination to oversee cross-border digital markets, where global firms might exploit jurisdictional gaps. Finally, leveraging artificial intelligence and machine learning tools could significantly enhance enforcement capabilities by enabling regulators to process large datasets, detect pricing anomalies, and respond more effectively to algorithmic strategies that distort competition.

IV. Conclusion

This paper advances the understanding of algorithmic collusion by focusing on the unique characteristics and implications of bandit algorithms in digital markets. While previous research, particularly by Calvano et al. (2020), has highlighted the risks of Q-learning algorithms leading to tacit collusion, our study demonstrates that bandit algorithms, with their faster convergence and real-time adaptability, may present even more significant challenges for antitrust enforcement. The speed at which bandit algorithms can adjust

to dynamic environments, combined with their ability to balance exploration and exploitation, enables them to quickly identify and exploit collusive equilibria, especially in markets characterized by rapid changes and limited data availability.

However, this increased efficiency and adaptability come at a cost. Bandit algorithms' ability to operate with minimal historical data and continuously optimize decisions based on immediate feedback introduces substantial risks of undetected collusion. These risks are further compounded by the difficulty in interpreting the decision-making processes of bandit algorithms, making regulatory oversight more challenging. The paper suggests that competition authorities need to develop more sophisticated monitoring and analytical tools to effectively detect and counteract the collusive behaviors facilitated by these algorithms.

Ultimately, while bandit algorithms may offer significant advantages in certain applications, their potential to amplify collusive risks in digital markets necessitates a reevaluation of current antitrust frameworks. As algorithms continue to play an increasingly central role in market dynamics, regulators must adapt their strategies to address the unique challenges posed by these advanced decision-making systems. This paper contributes to the ongoing discourse by providing a critical analysis of bandit algorithms' role in algorithmic collusion and proposing pathways for more effective regulatory intervention in the era of digital markets.

References

- Abada, I., Lambin, X., Tchakarov, N., 2024. Collusion by Mistake: Does Algorithmic Sophistication Drive Supra-Competitive Profits?, *European Journal of Operation Research*. 318(1), 927-953.
- Agrawal, R., 1995. Sample Mean Based Index Policies with $O(\log n)$ Regret for Multi-Armed Bandit Problem, *Advances in Applied Probability*, 27(4) 1054-1078.
- Agrawal, S., Goyal, N., 2012. Analysis of Thompson Sampling for the Multi-armed Bandit Problem. *Proceedings of the 25th Annual Conference on Learning Theory*, in *Proceedings of Machine Learning Research* 23:39.1-39.26 Available from <https://proceedings.mlr.press/v23/agrawal12.html>.
- Aoud, A., den Boer A.V., 2021. Algorithmic Collusion in Assortment Games. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3930364
- Asker J., Fershtman C., Pakes A., 2024. The impact of artificial intelligence design on pricing, *Journal of Economics and Management Strategy*, 33(2), 276-304
- Assad, S., Calvano, E., Calzolari, G., Clark, R., Denicolò, V., Ershov, D., Wildenbeest, M., 2021, Autonomous algorithmic collusion: economic research and policy implications. *Oxford Review of Economic Policy*, 37(3), 459-478. <https://doi.org/10.1093/oxrep/grab011>
- Auer, P. 2002. Using Confidence Bounds for Exploitation-Exploration Trade-offs. *Journal of Machine Learning Research*. 3 397-422.
- Auer, P., Cesa-Bianchi, N., Fischer, P., 2002, Finite-Time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2), 235-256. <https://doi.org/10.1023/A:1013689704352>
- Axelrod, R., Hamilton, W. D., 1981. The Evolution of Cooperation. *Science*. 211,1390 1396DOI:10.1126/science.7466396

Babaioff, M., Dughmi, S., Kleinberg, R., Slivkins, A., 2015. Dynamic Pricing with Limited Supply. *ACM Transactions on Economics and Computation*, 3(1), 1-26. <https://doi.org/10.1145/2559152>

Biancho, M., Mantegazza, G. 2023. Adaptive Algorithms and Collusion by Coupling. EC '23: Proceedings of the 24th ACM Conference on Economics and Computation. <https://doi.org/10.1145/3580507.3597726>

Brown, D. P., Eckert, A., Silveira, D., 2023, Screening for Collusion in Wholesale Electricity Markets: A Review of the Literature. *Utilities Policy*. 85, art. 101671

Bubeck, S., Cesa-Bianchi, N., 2012. Regret Analysis of Stochastic and Non-stochastic Multi-Armed Bandit Problems. *Foundations and Trends® in Machine Learning*, 5(1), 1-122. <https://doi.org/10.1561/22000000024>

Calvano, E., Calzolari, G., Denicolo, V., Pastorello, S., 2019. Algorithmic Pricing: What Implications for Competition Policy? *Review of Industrial Organization*, 55(1), 155-171. <https://www.jstor.org/stable/48702985>

Calvano, E., Calzolari, G., Denicolò, V., Pastorello, S., 2020. Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review*, 110(10), 3267-3297. <https://doi.org/10.1257/aer.20190623>

Cho I.-K., Williams N., 2024. Collusive Outcomes Without Collusion, <https://arxiv.org/abs/2403.07177>

Cooper, R., DeJong, D. V., Forsythe, R., Ross, T. W., 1989. Communication in the Battle of the Sexes Game: Some Experimental Results. *The RAND Journal of Economics*, 20(4), 568-587. <https://doi.org/10.2307/2555734>

Crandall, J. W., Oudah, M., Tennom et al., 2018. Cooperating with machines. *Nature Communications*, 9, 233. <https://doi.org/10.1038/s41467-017-02597-8>

De Cooman J., 2023. Outsmarting Pac-Man with Artificial Intelligence, or Why AI-Driven Cartel Screening Is Not a Silver Bullet, *Journal of European Competition Law & Practice*, lpad017, <https://doi.org/10.1093/jeclap/lpad017>

Decarolis, F., Giorgiantonio, C., 2022. Corruption red flags in public procurement: new evidence from Italian calls for tenders. *EPJ Data Science*, 11(16), 1-38. <https://doi.org/10.1140/epjds/s13688-022-00325-x>

Dorner, F. E., 2021 Algorithmic collusion: A critical review. *arXiv*. <https://doi.org/10.48550/arXiv.2110.04740>

Eschenbaum N., Mellgren F., Zahn P., 2022. Robust Algorithmic Collusion, <https://arxiv.org/abs/2201.00345>

Evans, D. S., Tucker, C. E., Fels, A., 2020. The Evolution of Antitrust in the Digital Era: Essays on Competition Policy. *Competition Policy International*. Retrieved from <https://www.gurkaynak.av.tr/docs/11288-volume-1-the-evolution-of-antitrust-in-the-digital-era-essays-on-competition-policy.pdf>

Ezrachi, A., Stucke, M. E., 2017. Artificial Intelligence & Collusion : When Computers Inhibit Competition. *Illinois Law Review*, *October*. <https://doi.org/10.2139/ssrn.2591874>

Gal, M. S., 2019. Algorithms as Illegal Agreements. *Berkeley Technology Law Journal*, 34(1), 67-118.

Gallego, J., Rivero, G., Martinez, J., 2021. Preventing rather than Punishing: An Early Warning Model of Malfeasance in Public Procurement. *International Journal of Forecasting*, 37(1), 360-377. <https://doi.org/10.1016/j.ijforecast.2020.06.006>

Hansen, K. T., Misra, K., Pai, M. M., 2021. Algorithmic Collusion : Supra-competitive Prices via Independent Algorithms. *Marketing Science*, 40(1), 1-12. <https://doi.org/10.1287/mksc.2020.1276>

Harrington, J. E., 2018. Developing Competition Law for Collusion by Autonomous Artificial Agents. *Journal of Competition Law & Economics*, 14(3), 331-363. <https://doi.org/10.1093/joclec/nhy016>

Hartline, J.D., Long, S., Zhang C., 2024. Regulation of Algorithmic Collusion. CSLAW'24, March 12-13, Boston. <https://arxiv.org/abs/2401.15794>

Hingston, P., Kendall, G., 2004. Learning versus evolution in iterated prisoner's dilemma. In *Proceedings of the 2004 Congress on Evolutionary Computation*, 1, 364-372 <https://doi.org/10.1109/CEC.2004.1330880>

Li, L., Chu, W., Langford, J., Schapire, R. E., 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, 661-670. <https://doi.org/10.1145/1772690.1772758>

Marty, F., Warin, T., 2023. Coordinated Anticompetitive Practices Based on Algorithms”, in Marciano A., Ramello G.B., eds, *Encyclopaedia of Law and Economics*, Springer, New York, https://doi.org/10.1007/978-1-4614-7883-6_818-1

Marty, F., & Warin, T., (2024). Algorithmes de prix et signaux collusifs : approches économique et concurrentielle. *Revue Française d'Economie*. XXXIX, 2, 81-127..

Mehra, S. K., 2016. Antitrust and the Robo-Seller: Competition in the Time of Algorithms. *Minnesota Law Review*, 100, 1323-1375.

OECD, 2012. Unilateral Disclosure of Information with Anticompetitive Effects. Retrieved from <https://www.oecd.org/daf/competition/Unilateraldisclosureofinformation2012.pdf>

Sadoune, I., Joanis, M., Lodi, A. 2024. Algorithmic Collusion and The Minimum Price Markov Game. <https://arxiv.org/abs/2407.03521>

Salcedo, B., 2015. Pricing algorithms and tacit collusion. Manuscript, Pennsylvania State University. Retrieved from <http://brunosalcedo.com/docs/collusion-slides.pdf>

Schwalbe, U., 2018. Algorithms, Machine Learning, and Collusion. *Journal of Competition Law & Economics*, 14(4), 568-607. <https://doi.org/10.1093/joclec/nhz004>

Sun, B., Deng, R., Ren, B., Teng, M., Cheng, S., Wang, F., 2022. Identification method of market power abuse of generators based on lasso-logit model in spot market. *Energy*. 238. <https://doi.org/10.1016/j.energy.2021.121634>.

Sutton, R. S., Barto, A. G., 2018. *Reinforcement Learning: An Introduction*. MIT Press..

Wallimann, H., Imhof, D., Huber, M., 2023. A Machine Learning Approach for Flagging Incomplete Bid-Rigging Cartels. *Computational Economics*, 62, 1169-1720. <https://doi.org/10.48550/arXiv.2004.05629>

Waltman, L., Kaymak, U., 2008. Q-Learning Agents in a Cournot Oligopoly Model. *Journal of Economic Dynamics and Control*, 32(10), 3275-3293. https://econpapers.repec.org/article/eedyncon/v_3a32_3ay_3a2008_3ai_3a10_3ap_3a3275-3293.htm

Watkins, C. J. C. H., Dayan, P., 1992. Q-learning. *Machine Learning*, 8, 279-292. <https://doi.org/10.1007/BF00992698>

Werner, T., 2023. Algorithmic and Human Collusion. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3960738

Xu, X., Lee, S., Tan, Y., 2023. Algorithmic Collusion or Competition: The Role of Platforms' Recommender Systems. <https://doi.org/10.48550/arXiv.2309.14548>

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Journal Pre-proof