TI 2025-012/VII
Tinbergen Institute Discussion Paper

How the design of cartel fines affects prices: Evidence from the lab

Sindri Engilbertsson[1]
Sander Onderstal[2]
Leonard Treuren[3]

[1] University of Amsterdam, Tinbergen Institute

[2] University of Amsterdam, Tinbergen Institute

[3] KU Leuven

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

How the design of cartel fines affects prices: Evidence from the lab[*]

Sindri Engilbertsson[†], Sander Onderstal[‡], and Leonard Treuren[§]

January 2025

Abstract

Competition authorities impose substantial penalties on firms engaging in illegal price-fixing. We examine how basing cartel fines on either revenue, profit, or price overcharge influences cartel and market prices, as well as cartel incidence and stability. In an infinitely repeated Bertrand oligopoly game, we show that revenue-based fines incentivize firms to charge prices above the monopoly price, whereas only overcharge-based fines encourage prices below the monopoly price. Cartels are stable for a smaller range of discount factors when fines are based on overcharges rather than other bases. We test these predictions in a laboratory experiment where subjects can form cartels, which allows them to discuss pricing at the risk of being detected and fined. By equalizing expected fines across treatments, we isolate the effect of the fine's base. We find that market prices are lowest under overcharge-based fines and highest under revenue-based fines. Variation in market prices across treatments is fully driven by cartel prices. While these results align with the theoretical predictions, cartel incidence remains unchanged across regimes. Our results suggest competition authorities could improve enforcement by shifting from revenue-based fines to profit- or overcharge-based fines.

Keywords: Antitrust; Cartel; Collusion; Repeated game; Experiment
JEL Codes: C73; C92; D43; K21; L41

[*] We thank Lisa Bruttel, Joe Harrington, Florian Heine, João Montez, Evgenia Motchenkova, Catherine Roux, Maarten Pieter Schinkel, Yossi Spiegel, Christian Thöni, and participants at the University of Amsterdam, KU Leuven, LMU München, Sun Yat-Sen University Guangzhou, and the 2024 MaCCI (Mannheim), 2024 IIOC (Boston), 2024 BECCLE (Bergen), 2024 M-BEES (Maastricht), 2024 CRESSE (Crete), 2024 EEA (Rotterdam), 2024 EARIE (Amsterdam), and 2024 ESA (Helsinki) conferences for useful comments and discussions. Engilbertsson thanks the UvA Research Priority Area in Behavioral Economics (grant EB-1145). Treuren acknowledges support from ERC Consolidator grant 816638. This experiment was pre-registered with the AEA RCT Registry (RCT ID: AEARCTR-0012043). Opinions and errors remain ours.

[†] University of Amsterdam and Tinbergen Institute; s.engilbertsson@uva.nl.

[‡] University of Amsterdam and Tinbergen Institute; a.m.onderstal@uva.nl.

[§] KU Leuven; leonard.treuren@kuleuven.be.

1 Introduction

Competition authorities and courts regularly impose substantial penalties on firms that participate in illegal horizontal price-fixing agreements.[1] Across antitrust jurisdictions, such fines are based on cartel members' revenue.[2] However, theoretical literature has long established that revenue bases can increase cartel prices (e.g., Bageri et al. (2013); Katsoulacos and Ulph (2013)). This discrepancy between theory and practice raises the question of the relative performance of different fining regimes.

This paper experimentally investigates how different fining bases affect cartel and market prices. Specifically, we examine whether the theoretical concerns surrounding revenue-based fines–which are based on equilibrium selection assumptions —- are justified. We also compare the effectiveness of revenue-based fines to viable alternatives. Since these questions are inherently empirical, but observational data offer limited insights, we use a laboratory experiment to address them. We compare revenue-based fines to two alternatives for which a legal basis exists in the U.S. fining guidelines: fines based on cartel members' profits, and fines based on cartel members' price overcharge with respect to the competitive price.[3]

We build our hypotheses on the analysis of infinitely repeated Bertrand games where firms can use trigger strategies to realize above one-shot equilibrium prices as part of a subgame-perfect Nash equilibrium. Coordinating on such trigger strategies leaves a paper trail that is detected with a fixed probability each period by the antitrust authority. Members of detected cartels are fined, and undiscovered cartels from previous periods can be detected. The basis of a cartel member's fine is either her revenue, her profit, or the price overcharge she sets. After discovery, cartels are reformed with a fixed probability in every subsequent period. This theoretical model isolates key factors of the antitrust policy under consideration.

To deal with the multiplicity of equilibria in infinitely repeated oligopoly games, we follow the theoretical literature and make assumptions on equilibrium selection. In particular, we assume that firms coordinate on the joint-profit-maximizing price by using trigger strategies. Where our theoretical assumptions fail, a comparison of the fining regimes might deliver

[1] For instance, the European Commission imposed a €3.8 billion fine to truck manufacturers in the Trucks case (2016/2017). Other examples include the $2.5 dollars and €1.4 billion fines in the Foreign Exchange Market case by the United States Department of Justice (2015) and the European Commission (2019/2021), respectively, and the fine of ¥101 billion imposed by Japan's competition authority in 2023 on the electric power cartel.

[2] For instance, the guidelines on the method of setting fines by the European Commission (2006) state that: "In determining the basic amount of the fine to be imposed, the Commission will take the value of the undertaking's sales of goods or services to which the infringement directly or indirectly relates. . . " According to United States Sentencing Commission (2023), the base fine for bid-rigging, price-fixing, and market-allocation agreements is "20 percent of the volume of affected commerce" (p.311).

[3] For non-antitrust criminal organizations, fines are typically based on either the organization's pecuniary gain from the offense (profit) or the pecuniary loss inflicted by the offense (damages) (United States Sentencing Commission, 2023, p.526).

different results. In particular, revenue bases might not cause cartel prices to exceed the monopoly price.[4]

Our theoretical results on cartel prices align with the broader literature (e.g., Bageri et al. (2013); Katsoulacos et al. (2015)). Revenue bases incentivize cartel members to charge a price above the no-antitrust monopoly price. In contrast, basing the fine on a cartel member's overcharge reduces the optimal cartel price compared to the monopoly price, while a profit base leaves the optimal cartel price unaffected. Intuitively, fines based on revenue serve as a tax pushing prices up; a profit-based fine does not affect the profit-maximizing price because the cartel's expected profits are a fraction of its profits without a fine; in an overcharge regime, the fine is strictly increasing in price, which gives the cartel an incentive to mitigate the price.

In contrast to earlier findings, our model suggests that cartel stability is lowest when cartel fines are based on the price overcharge. The reason is that defection in the revenue and profit regimes increases the expected fine, while defection decreases the price overcharge, and hence the expected fine in the overcharge regime. Therefore, we point towards an additional theoretical benefit of an overcharge regime compared to the currently used fining regime. Central to this result is the assumption that defectors can be fined, which is in line with antitrust practice (Buccirossi and Spagnolo, 2007).

We test the above predictions using a laboratory experiment.[5] While no laboratory experiment can fully replicate real-world cartels, our approach allows us to overcome significant challenges posed by field data. First, observing cartels in the field is inherently difficult because of their illegal nature. Moreover, discovered cartels likely form a non-representative sub-sample of the entire population of cartels. Second, laboratory control allows the researcher to obtain an apples-to-apples comparison of different fining regimes, which is challenging in the field as it is difficult to establish exogenous variation and to measure variables of interest like marginal costs and demand. Finally, experiments enhance internal validity by ensuring that theoretical assumptions are met as closely as possible under controlled conditions. For these reasons, laboratory experiments are widely employed as wind-tunnel tests of theory-based policy recommendations (Falk and Heckman, 2009; List, 2020).

In the experiment, 279 participants compete in indefinitely repeated Bertrand markets that closely mirror our theoretical model. Subjects can opt into cartels by voting, which allows them to freely discuss pricing at the risk of being detected and fined. In the `REVENUE`,

[4] For instance, Bageri et al. (2013, p.F550) remark that "of course, it could be argued that the practical significance of this distortion is likely to be small because it requires managers of firms involved in cartels to be well-informed and forward-looking, and to formulate strategic decisions at a level that may not be easily met in reality."

[5] We are not the first to study the effect of fining regimes on cartel formation in a laboratory study. Fonseca et al. (2022) examine whether fining regimes that impose sanctions on managers involved in cartels have greater deterrence power than regimes that only levy corporate fines, allowing shareholders to determine the labor contracts of the firms' managers. They find less cartel formation when managers can be prosecuted.

PROFIT, and OVERCHARGE treatments, the fine of a discovered cartel member is based on that individual's revenue, profit, or price overcharge, respectively. We equalize excepted fines across treatments so that our results are not driven by behavioral responses to the size of the fine. Varying the treatments between participants allows us to identify the causal link between the three fining regimes and a host of outcomes of interest, including the market price, the price charged by cartels, the likelihood of cartel formation, cartel incidence, and cartel recidivism.

Our experimental findings on prices are in line with the theoretical predictions. While uncartelized markets yield prices close to the one-shot Nash equilibrium price in all three fining regimes, cartel prices are lowest when fines are based on the overcharge and highest when they are based on revenue. Indeed, when fines are based on revenue, both the price agreements that subjects form and the market prices that result from such agreements exceed the monopoly price. However, we find no significant differences in cartel formation, incidence, and recidivism across treatment. Therefore, market price differences across treatments are entirely determined by cartel prices. Our findings suggest benefits from antitrust authorities moving away from revenue bases towards profit or overcharge bases. We conclude by arguing that such a change is realistically implementable given current legal and institutional constraints–particularly for a profit-based regime.

We contribute to a strand of theoretical literature studying how cartel fining regimes influence market outcomes. In particular, revenue regimes have been shown to have the perverse effect of increasing cartel prices in Bageri et al. (2013) and Katsoulacos and Ulph (2013). Profit bases are studied in Block et al. (1981), Harrington Jr. (2004), and Harrington Jr. (2005), among others. The overcharge base is proposed as an attractive alternative to revenue and profit bases by Katsoulacos et al. (2015), the paper most closely related to our theoretical model as it compares the same fining regimes.[6] To generate unique equilibria, theoretical models of collusion based on infinitely repeated oligopoly games routinely make assumptions on equilibrium selection, for instance that firms coordinate on the joint-profit-maximizing price. Our main contribution to this literature is to test its theoretical predictions empirically.[7]

Oligopoly laboratory experiments studying corporate leniency programs often compare treatments with fines to treatments with fines and a leniency program. Fines are either independent of firm conduct (e.g., Bigoni et al. (2012, 2015)), or based on revenue (e.g., Apesteguia et al. (2007); Hinloopen and Soetevent (2008)). Across different treatments, the

[6] In contrast to our model, Katsoulacos et al. (2015) assume that defectors cannot be fined, that cartels immediately reform after detection, and that cartels can only be detected in the period in which they are formed. While these differences do not influence the ranking of cartel prices across fining regimes, deterrence is equal in all three regimes under the assumptions of Katsoulacos et al. (2015).

[7] As we compare commonly studied fining bases while holding fixed the level of the fine, we do not address the optimal level or design of fines, which have been studied by Buccirossi and Spagnolo (2007), Katsoulacos and Ulph (2013), and Houba et al. (2018).

size of the fine varies but the base of the fine is fixed. In contrast, we hold the size of the expected fine fixed and vary the fining base, which allows us to study the effect of fining structure on cartel behavior. We also have chosen to abstract from leniency programs for the same reason and to limit the demands on our experimental subjects. We view the inclusion of leniency as an avenue for future work.

Our paper also contributes to the broader literature analyzing the impact of various competition policy instruments on cartel behavior. This literature has studied a wide range of policy questions including the effectiveness of leniency programs (see Marvão and Spagnolo (2018) and Hinloopen et al. (2023) for overviews), spillovers from legal cooperation in some markets to tacit collusion in others (e.g., Duso et al. (2014); Sovinsky (2022)), the role of communication in collusion (e.g., Kandori and Matsushima (1998); Awaya and Krishna (2016)), and the effect of market transparency programs on collusion (e.g., Vega-Redondo (1997); Byrne and De Roos (2019)). Theoretical predictions in these domains are routinely tested in laboratory experiments.[8]

Finally, we contribute to the experimental literature on cooperation in indefinitely repeated games, surveyed by Dal Bó and Fréchette (2018). This literature finds the discount factor exceeding the critical discount factor to be a necessary, but insufficient, condition for cooperation to emerge in the absence of communication. Moreover, an important finding is that cooperation rates are increasing in the difference between the actual discount factor and the critical discount factor. While laboratory experiments on collusion typically equalize critical discount factors across treatments, this is impossible in our theoretical model as overcharge based fines always have higher critical discount factors than the other two regimes. However, we do not find differences across fining regimes in any of our measures of collusion, and do report a tendency towards complete cartelization over time, suggesting that the results surveyed in Dal Bó and Fréchette (2018) do not extend to a setting where subjects can freely communicate.

The structure of this paper is as follows. In Section 2, we present our model and the theoretical results on which we base our hypotheses. Section 3 contains our experimental design, experimental procedures, and hypotheses. Section 4 gives our experimental findings and their implications. Concluding remarks are in Section 5. Proofs of propositions are relegated to Appendix A.

[8] Lessons from the experimental literature include: leniency programs having the desired effects on cartel formation, cartel discovery, and the price (Hinloopen and Soetevent, 2008; Bigoni et al., 2012, 2015); spillovers emerging from legal cooperation in one experimental market to tacit collusion in another (Normann et al., 2015; Hinloopen et al., 2024); explicit communication about future conduct facilitating collusive prices (Fonseca and Normann, 2012; Freitag et al., 2021; Gomez-Martinez et al., 2016); transparency about competitors' actions increasing competition (Huck et al., 1999, 2000; Offerman et al., 2002).

2 Theoretical framework

Our theoretical framework is based on three main assumptions. First, collusive agreements must be self-enforcing due to the illegal nature of price-fixing. Second, communication is required to achieve collusion, and leaves a paper trail which can be detected by the antitrust authority. Third, firms internalize the possibility of price-fixing fines. To replicate key findings from the theoretical literature for the setting we implement in the laboratory, our model closely follows existing theory of antitrust penalties, where these assumptions are routinely made (e.g., Motta and Polo (2003); Aubert et al. (2006); Katsoulacos et al. (2015)).

2.1 The model

Consider an infinitely repeated homogenous-goods oligopoly game with $n \geq 2$ firms that maximize expected profit and have a common discount factor $\delta \in (0, 1)$. Each period t, each firm i sets a price $p_{it} \in [0, \bar{p}]$, with $\bar{p} > 0$ the highest possible price. Market demand in period t, $q_t \equiv q(p_t)$, depends on the market price p_t, which is the lowest price set that period, i.e., $p_t \equiv \min_i p_{it}$, with $q(\bar{p}) = 0$. Firms produce at constant marginal cost $c \in (0, \bar{p})$, and average and marginal market revenue are assumed strictly decreasing in market quantity. The m firms that set the lowest price in a given period share the resulting market demand equally: $q_{it} = \frac{q_t}{m}$ if $p_{it} = p_t$. Firms that do not set the lowest price face no demand: $q_{it} = 0$ if $p_{it} > p_t$.

Absent explicit communication, we assume that the one-shot Bertrand Nash-equilibrium always occurs: $p_{it} = c$, implying that profits are zero. To establish an experimental setting where this assumption holds roughly true, we opted for $n = 3$ in the experiment, as the literature finds that, absent communication, prices typically converge close to the static Nash-equilibrium for three or more players (e.g., Dufwenberg and Gneezy (2000); Fonseca and Normann (2012); Chowdhury and Crede (2020)). We denote the firm-period-level price and profits realized in this competitive benchmark by p^N and π^N.

Firms can choose to form a cartel and explicitly communicate, which allows market prices above p^N to emerge.[9] Such collusive prices are supported by a grim trigger strategy whereby firms coordinate on a price and set it as long as all firms set that price in all previous periods since the inception of the cartel. Otherwise, firms revert to setting price p^N forever. We assume that cartels coordinate on the joint-profit-maximizing price. Let $p^M(c)$ denote the monopoly price conditional on the firms' marginal costs c in the absence of antitrust. Note that our assumptions imply that $p^M(c)$ strictly increases with c.

Collusion leaves a paper trail that the antitrust authority can detect. In particular, once a cartel has been formed, it is detectable and remains so in later periods until the antitrust

[9] Explicit communication is typical of uncovered cartel cases–even duopolies such as vitamin A500 USP and beta-carotene cartels (Marshall and Marx, 2012). Models of collusion routinely assume that communication is required for collusion to emerge (e.g., McCutcheon (1997); Aubert et al. (2006)).

authority has discovered it, regardless of the behavior of the cartel members. This implies that firms that defect from the cartel agreement and those on the punishment path of the grim trigger strategy profile can be convicted and fined. This is in line with reality, and our experiment, as defectors do not face reduced fines in either the U.S. or the EU (Buccirossi and Spagnolo, 2007). In addition, the paper trail used to convict cartels typically originates years prior to detection and conviction (Kwoka and White, 2018). Cartels are reformed with fixed probability $\rho \in [0, 1]$ each period after detection, in line with evidence of limited recidivism (Marvão and Spagnolo, 2018). This nests the two assumptions commonly made in the literature, $\rho = 0$ and $\rho = 1$.

Each period, after prices are set and the market clears, the antitrust authority detects, prosecutes, and convicts all active cartels with probability $\alpha \in (0, 1)$. Upon conviction in period t, each cartel member i pays fine $F_{it} = rB_{it}$, where r is the penalty rate and B_{it} is the penalty base. We focus on how different choices for B_{it} affect cartel pricing and stability. In theory, cartels can be deterred entirely by ensuring that F_{it} is sufficiently large. For instance, in the spirit of Becker (1968), by imposing a penalty which ensures that the expected profit of forming a cartel is negative. The starting point of the literature on cartel fines is that complete deterrence is not feasible for several reasons. In particular, the legal principle of proportionality puts a general cap on fines, and bankruptcy concerns put downward pressure on fines in particular instances (Buccirossi and Spagnolo, 2007; Houba et al., 2018).[10]

In the absence of side payments, and due to the symmetric nature of firms, we focus on collusive agreements where all firms set the same price, and share demand. Let π^C denote a firm's single-period before-fine profit in a cartel whose members all set price p^C. F^C represents that firm's concomitant fine upon detection. A firm's expected present value of participating in the cartel and the competitive benchmark are, respectively, given by

$$V^C = \frac{(\pi^C - \alpha F^C) f(\alpha, \delta, \rho)}{1 - \delta} \qquad \text{and} \qquad V^N = \frac{\pi^N}{1 - \delta}, \tag{1}$$

where

$$f(\alpha, \delta, \rho) = \frac{1 - \delta(1 - \rho)}{1 - \delta(1 - \alpha)(1 - \rho)} \ \in (0, 1]. \tag{2}$$

With perfect recidivism (i.e., if $\rho = 1$), $f(\alpha, \delta, \rho) = 1$, so that $V^C = \frac{\pi^C - \alpha F^C}{1 - \delta}$.

Let p^D denote the optimal defection of a cartel member if all other firms set p^C, and its resulting profit and fine π^D and F^D, respectively. If the cartel is convicted immediately after defection, all firms revert to playing the one-shot-Nash price forever. However, if the

[10] To not further burden subjects in what is already a complicated experiment, we have opted to follow the literature and let fines depend on outcomes in the current period only. Note that we do introduce dynamic detection. In reality, fines are often based on the estimated duration of the cartel. Introducing a dynamic component to fines substantially complicates the analysis and is, therefore, typically ignored in the literature. Notable exceptions are in Harrington (2004; 2005).

cartel is not immediately convicted after defection, firms select the price that maximizes their unilateral profit in the one-shot game taking into account the possibility of being fined, denoted by p^{PD}, in the next period(s). We denote the concomitant per-firm profit and fine by π^{PD} and F^{PD} respectively. After defection *and* detection, all firms set p^N again. Therefore, the expected present value of defection is given by

$$\begin{aligned} V^D &= \pi^D - \alpha F^D + \alpha\left(\delta\pi^N + \delta^2\pi^N + \ldots\right) \\ &\quad + (1-\alpha)\delta\Big[\pi^{PD} - \alpha F^{PD} + \alpha\left(\delta\pi^N + \delta^2\pi^N + \ldots\right) \\ &\quad + (1-\alpha)\delta\Big\{\pi^{PD} - \alpha F^{PD} + \alpha\left(\delta\pi^N + \delta^2\pi^N + \ldots\right) + \ldots\Big\}\Big] \\ &= \pi^D - \alpha\underbrace{\left(F^D - \frac{\delta}{1-\delta}\pi^N\right)}_{\text{Immediate detection}} + \frac{\delta(1-\alpha)}{1-\delta(1-\alpha)}\underbrace{\left(\pi^{PD} - \alpha\left(F^{PD} - \frac{\delta}{1-\delta}\pi^N\right)\right)}_{\text{Potential future detection}}. \end{aligned}$$

For all fining regimes that we consider, $\pi^{PD} - \alpha F^{PD} = 0$, so that $V^D = \pi^D - \alpha F^D$.

For stable cartels to be part of a subgame-perfect Nash equilibrium, two conditions must be met. First, the participation condition requires that $V^C \geq V^N$. This condition is always satisfied in our setting as $V^N = 0$, and caps on the maximum fine–such as the legal principle of proportionality–ensure that $\pi^C - \alpha F^C > 0$. Second, the stability condition requires that $V^C \geq V^D$. That is, defecting from the collusive agreement should not increase the expected present value of a firm's payoff stream. Cartel members, therefore, solve

$$\max_{p^C} \quad \pi^C - \alpha F^C \quad \text{s.t.} \quad V^C \geq V^D. \tag{3}$$

We next analyze how basing F^C on either revenue, profit, or the price overcharge (relative to the competitive price) influences cartel pricing and stability.

2.2 Revenue-based fines

We implement revenue-based fines by setting the penalty base for a cartel member equal to that firm's revenue: $F_{it} = r_R p_{it} q_{it}$, where r_R is the exogenous penalty rate. We study a revenue base because it is the norm in practice (ICN, 2017). The European Commission, for instance, selects the most recent annual revenue of the product to which the infringement pertains as the fine base.[11] While U.S. guidelines base fines for organizations on the loss caused by the offense and the illegal gains, the guidelines mention that the volume of affected

[11] The relevant annual sales are multiplied by a factor up to 0.3 based on the gravity of the infringement and then adjusted upward, primarily based on the duration of the infringement. Finally, the amount can be increased or decreased based on aggravating factors, mitigating factors, leniency applications, bankruptcy concerns, and out-of-court settlements (European Commission, 2006).

commerce–revenue–should be used instead for price-fixing, bid-rigging, and market allocation agreements (U.S. Sentencing Commission, 2023, p.311).

When fines are based on revenue, let p_R^C denote the price set by a stable cartel, and let δ_R^* denote the critical discount factor above which cartels are stable.

Proposition 1. *If fines are based on revenue,*

i) *the price set by a stable cartel exceeds the monopoly price:* $p_R^C > p^M(c)$*;*

ii) *with imperfect recidivism* $(\rho \in [0,1))$*, the critical discount factor increases in the probability of discovery* α*:* $\frac{\partial \delta_R^*}{\partial \alpha} > 0$*;*

iii) *with perfect recidivism* $(\rho = 1)$*, the critical discount factor is independent of the probability of discovery* α*:* $\delta_R^* = \frac{n-1}{n}$*.*

Proposition 1 shows that revenue-based fines have a perverse price effect, which has also been established in the existing theoretical literature using similar models (e.g., Bageri et al. (2013); Katsoulacos and Ulph (2013)). The fine acts as a tax on revenue, reducing marginal revenue but leaving marginal cost unaffected. Specifically, as shown in the proof of Proposition 1, the cartel acts like a monopolist in the absence of antitrust, facing marginal cost $\frac{c}{1-\alpha r_R} > c$. Hence, cartel output decreases and the price increases above the monopoly price. While this reduces before-fine profit compared to the monopoly price, it increases expected profit by reducing the fine.

In our model, a firm's best response to all other firms setting p_R^C is not the monopoly price $p^M(c)$, as defectors can be detected and fined. Instead, the optimal defection is to slightly undercut p_R^C and capture the entire market. This increases the defector's before-fine profit and fine n-fold compared to the cartel case. Because defection scales up expected profit by n, this is the only relevant parameter for cartel stability in the case of perfect recidivism. With imperfect recidivism, increasing the rate of detection affects V^C more than V^D as, in addition to the higher expected fine, not all cartels immediately reform following detection.

An additional effect of the revenue base is that, following defection, prices remain above the competitive benchmark if previous cartel members can still be detected. Then, since setting p^N results in positive revenue but no profit, firms set price $p_R^{PD} = \frac{c}{1-\alpha r_R} > p^N$ until they are detected, after which they revert back to setting p^N.

2.3 Fines based on profit

We implement profit-based fines by setting the penalty base for a cartel member equal to that firm's profit: $F_{it} = r_\pi(p_{it} - c)q_{it}$, where r_π is the exogenous penalty rate. We study this base as it is arguably a relatively straightforward-to-implement alternative to a revenue base as it has a legal basis – the US guidelines for non-antitrust offenses mention the incremental profit due to the offense ('pecuniary gain') as a base for the fine (U.S. Sentencing Commission, 2023, p.526).

Let p^C_π denote the price set by a stable cartel, and let δ^*_π denote the critical discount factor above which cartels are stable if fines are based on profit.

Proposition 2. *If fines are based on profit,*

i) *the price set by a stable cartel equals the monopoly price:* $p^C_\pi = p^M(c)$*;*

ii) *with imperfect recidivism* $(\rho \in [0, 1))$ *the critical discount factor increases in the probability of discovery* α*:* $\frac{\partial \delta^*_\pi}{\partial \alpha} > 0$*;*

iii) *with perfect recidivism* $(\rho = 1)$ *the critical discount factor is independent of the probability of discovery* α*:* $\delta^*_\pi = \frac{n-1}{n}$*.*

As a profit-based fine acts as a tax on profit, it does not affect the profit-maximizing price, so that the cartel sets the monopoly price. Similar results on cartel pricing in different Bertrand games are in Bageri et al. (2013) and Katsoulacos et al. (2015). The optimal defection, like in the revenue case, is to slightly undercut the cartel. This increases expected profit by a factor n, so the critical discount factor is identical to the revenue-based critical discount factor–only affected by antitrust when there is imperfect recidivism.

Note that in our Bertrand setting, incremental profit and profit bases are equivalent as profits are zero in the absence of a cartel. To investigate the generality of Proposition 2, consider what happens if incremental profit is bench-marked against a but-for price p^{BF} that results in positive firm-level profit $\pi^{BF} > 0$. It can be shown that a cartel in an incremental profit-based regime still sets p^C_π, but that the critical discount factor is *below* δ^*_π as defecting scales up the fine more than the before-fine profit.[12] That is, when benchmark profits are positive, a profit base has more attractive properties than an incremental profit base as it results in the same cartel price and makes collusion sustainable for a smaller set of discount factors. This is reassuring, as a measure of total profit is relatively easy to obtain, while determining incremental profit requires and estimate of the counterfactual but-for quantity.

2.4 Overcharge-based fines

We implement overcharge-based fines by setting the penalty base for a cartel member equal to the difference between that firm's price and the competitive price: $F_{it} = r_O(p_{it} - p^N)q^N$, where r_O is the exogenous penalty rate and $q^N = \frac{q(p^N)}{n}$. We study this base as Katsoulacos et al. (2015) show that it leads to cartel prices strictly below the monopoly price, and estimates of price overcharges are routinely made for damage cases, suggesting that the base could be implemented. Following Katsoulacos et al. (2015), we multiply the overcharge by the competitive output of an individual firm. However, we could multiply the overcharge by

[12] The cartel price then solves $\arg\max_p (1 - \alpha r_\pi)(p - c)q(p) + \alpha r_\pi(p^{BF} - c)q(p^{BF})$, and the critical discount factor (for $\rho = 1$) is $\frac{(n-1)(1-\alpha r_\pi n \pi^{BF})}{n(1-\alpha r_\pi (n-1)\pi^{BF})}$.

any constant the cartel cannot control without losing any of the base's attractive properties, which is important as calculating the counterfactual quantity q^N is challenging in practice.[13]

We stress that overcharge-based fines do not correspond to damage claims. In the U.S., where damage claims are most prevalent, the standard formula for consumer damages in cartel cases is $(p^C - p^N)q^C$ (Harrington, 2014). That is, in practice, only damages for goods that were sold are taken into consideration. The overcharge base replaces the cartel output in the damage claim by some constant that the cartel cannot affect–in our case q^N. This distinction generates the attractive properties of an overcharge-based fine, as a price increase will put upward pressure on both overcharge-based fines and consumer damages on sold goods, but also put downward pressure on damages by reducing the collusive quantity.

Let p^C_O denote the price set by a stable cartel, π^C_O and F^C_O the corresponding firm-level profit and fine, and δ^*_O the critical discount factor above which cartels setting the joint-profit-maximizing price are stable if fines are based on the overcharge.

Proposition 3. *If fines are based on the overcharge,*

i) *the price set by a stable cartel lies below the monopoly price:* $c \leq p^C_O < p^M(c)$*;*

ii) *the critical discount factor increases with the probability of discovery* α*, regardless of the level of recidivism:* $\frac{\partial \delta^*_O}{\partial \alpha} > 0 \ \forall \rho \in [0, 1]$.

An overcharge-based fine reduces the cartel price compared to $p^M(c)$ as it directly targets the distortion created by the cartel: a price above p^N. The only possible way for a cartel to reduce the fine is to lower the cartel price. At $p^M(c)$, a price reduction decreases the expected fine by more than it decreases before-fine profit, thereby increasing expected profit. Proposition 3(i) extends the result of Katsoulacos et al. (2015) to a setting where recidivism is imperfect, defectors can be fined, and cartels formed in previous periods can be detected.

In contrast to the revenue and profit regimes, defecting from the cartel agreement does not increase the fine in an overcharge-based regime, but does increase before-fine profit n-fold. This has two effects. First, antitrust *always* affects the critical discount factor, which increases in both the the penalty rate and the detection probability. Second, defection is incentivized compared to the other two fining regimes, where defection increases both the before-fine profit and the fine by a factor n.

For discount factors $\bar{\delta}_O < \delta < \delta^*_O$, stable cartels that set the joint-profit-maximizing price do not exist, but stable cartels that set a lower price can be part of a subgame-perfect Nash equilibrium. This price–given in the proof of Proposition 3–is always below the joint-profit-maximizing price, and the $\bar{\delta}_O$ is always higher than the critical discount factors of the revenue

[13] Although several jurisdictions mention the overcharge as relevant for determining fines, to our knowledge, an overcharge base has not been implemented in practice. For example, the U.S. fining guidelines mention the overcharge substantially differing from 10 percent as one of the factors determining which fine is selected from the range of possible fines (U.S. Sentencing Commission, 2023, p.312).

and profit regimes. We compare joint-profit-maximizing prices and the corresponding critical discount factors across fining regimes in the remainder of this paper.

2.5 Comparison of fining regimes

Propositions 1 to 3 allow us to obtain our main theoretical result:

Proposition 4. *Comparing across fining regimes shows that, regardless of the recidivism level ρ,*

i) *The price set by a stable cartel is highest if fines are based on revenue and lowest if fines are based on the overcharge: $p_R^C > p_\pi^C > p_O^C$.*

ii) *The critical discount factor is highest if fines are based on the overcharge: $\delta_O^* > \delta_R^* = \delta_\pi^*$.*

Revenue-based fines incentivize cartels to increase prices above the monopoly price as the slight reduction in before-fine profit is more than offset in expected profit by a lower penalty base. Overcharge-based fines reduce prices compared to the monopoly price, as they directly target the distortion created by the cartel: a price above the competitive price. Profit-based fines leave prices unaffected as they are essentially a proportional tax on firm profit.

Overcharge-based fines always increase the critical discount factor above which cartels are stable compared to the other fining regimes. This effect arises because defectors can be fined. As defecting from a collusive agreement increases before-fine profit and revenue n-fold, defecting increases the fine n-fold in revenue and profit-based regimes, but the fine does not increase in the overcharge regime. As a result, the critical discount factor is always highest in the overcharge regime, even though antitrust can still deter cartels under imperfect recidivism in the revenue and profit regimes.

Consider a non-degenerate distribution of discount factors δ over different markets. The average price–averaged over stable cartels and competitive markets–follows the same ranking as in Proposition 4i). This follows as overcharge-based fines result in the fewest number of stable cartels and the lowest cartel price of all fining regimes. While revenue and profit bases induce identical deterrence, prices of undeterred cartels are higher when fines are based on revenue.

Proposition 4 raises concerns about the current fining practice, as revenue bases are commonly encountered while overcharge bases have yet to be implemented. Why, then, is practice not more aligned with the theory? One potential reason is that the theoretical results are based on a host of assumptions that guarantee a unique subgame-perfect Nash equilibrium.[14] In particular, we assume–as does the broader literature–that firms coordinate on the joint-profit maximum, but coordination on an infinite amount of other prices is also possible.[15] Therefore, we conduct a laboratory experiment to test whether subjects select the

[14] We discuss implementability in Section 5.

[15] “The multiplicity of equilibria is an embarrassment of riches,” as famously noted by Tirole (1988).

same equilbria as the theory presumes. In addition, an experiment allows us to randomize the fining regime and accurately track cartel formation, demise, and pricing. In contrast, data on discovered cartels suffer sample selection bias, and identifying the cartel's duration and marginal costs is challenging.

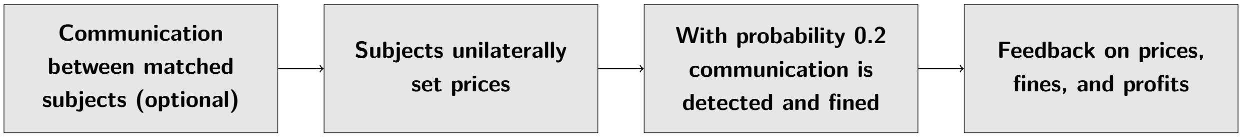

Communication between matched subjects (optional)
Subjects unilaterally set prices
With probability 0.2 communication is detected and fined
Feedback on prices, fines, and profits

Figure 1: Timeline of a single period

3 Experimental design, procedures, and hypotheses

3.1 Experimental design and procedures

Our experiment tests how the different cartel fining regimes studied in Section 2 affect prices. Subjects play an infinitely repeated Bertrand triopoly game.[16] Each period in each treatment follows the timeline displayed in Figure 1. Subjects first engage in optional communication and then set their price unilaterally. Next, the market clears, and cartels are detected and punished with a fixed probability of 0.2. We vary fines across different treatments by basing them either on a firm's revenue, profit, or overcharge. Finally, subjects receive feedback on the prices, fines, and profits. With probability 0.9, the three matched subjects play another period, while with probability 0.1, each subject is re-matched with two new subjects before playing the next period. We now explain all phases of a period in more detail.

After being matched with two subjects, each subject unilaterally votes for or against cartel formation. Only if all three subjects vote in favor, a cartel is formed and a chat window becomes available to the subjects. This free-chat is available for 60 seconds in the cartel's first period and 30 seconds in all subsequent periods until the cartel is detected by the competition authority or subjects are re-matched.[17] If no cartel is formed, subjects start the next period by voting on cartel formation. We implement communication by using a chat as this facilitates coordination on the joint-profit-maximizing outcome and stable cartels to a much larger extent than restricted communication protocols such as suggesting prices (Cooper and Kühn (2014); Harrington et al. (2016)). In addition, unrestricted communication using

[16] Some authors speak of 'indefinitely' rather than 'infinitely' repeated games. We follow Dal Bó and Fréchette (2018) and use 'infinitely repeated' as a reference to the theoretical framework under consideration rather than a description of the implementation in the laboratory.

[17] We do not allow for partial cartels as this would further complicate the already challenging decision problem for subjects. Moreover, Clemens and Rau (2022) show that subjects are more likely to form complete than partial cartels when both are part of a Nash equilibrium.

natural language is a central feature of discovered hard-core cartels (Genesove and Mullin (2001); Harrington (2006)).

Market demand in period t of the Bertrand triopoly is given by $q(p_t) = 100 - p_t$, and marginal costs equal 47. We opt for a triopoly as tacit collusion is frequently observed in oligopoly experiments with no more than two players (Huck et al., 2004). If subjects can earn more by tacitly colluding than by engaging in potentially costly communication, cartels will rarely form, making the study of their behavior challenging.[18] With more than two players, tacit collusion is unlikely to occur as market prices in Bertrand experiments closely resemble the competitive price in the absence of explicit communication (e.g., Dufwenberg and Gneezy (2000); Fonseca and Normann (2012); Freitag et al. (2021)).

We believe a Bertrand game stimulates subjects' understanding of the game. In addition, a Bertrand setting is the norm in existing theory on cartel fines and is used in many oligopoly experiments (e.g., Dufwenberg and Gneezy (2000); Fonseca and Normann (2012); Hinloopen et al. (2024)). Fines can vary substantially over periods in this setting as defectors capture the entire market. While extreme, a Bertrand game magnifies the incentives that also exist in Cournot games or differentiated goods games, thereby facilitating subjects' understanding. Several authors have instead employed differentiated goods price-setting duopolies when investigating antitrust in the laboratory (e.g., Bigoni et al. (2012, 2015)). While attractive in the duopoly case, differentiated goods price-setting games with more than two players are challenging to implement and place strong demands on experimental subjects.[19] To further aid subjects' understanding, an on-screen profit calculator was made available.

After setting prices, the market clears, and members of active cartels–subjects with access to the chat that period–are discovered and fined with probability 0.2.[20] We implement three treatments. In REVENUE, $F_{it} = p_{it}q_{it}$. In PROFIT, $F_{it} = 2.33(p_{it} - 47)q_{it}$. Finally, in OVERCHARGE, $F_{it} = 1.85(p_{it} - 47)\left(\frac{53}{3}\right)$.[21] Penalty rates are selected to equalize fines across treatments. This ensures that our results are not driven by behavioral responses to the size of the fine and align well with practice, where the principle of proportionality puts a cap on permissible changes of the total fine following penalty base adjustments. We refrain from including a treatment without antitrust. We are interested in studying how the cartel fining regime influences cartel pricing and stability rather than comparing the behavior of legal (or unprosecuted) and illegal cartels. Our results are, therefore, informative for countries with

[18] Indeed, even without antitrust Fonseca and Normann (2014) find that more cartels are formed in four-firm experimental oligopolies than in duopolies. The monetary gains from explicit communication appear lowest for Bertrand duopolies (Fonseca and Normann, 2012).

[19] For instance, Bigoni et al. (2012, 2015) restrict the action space and provide the subjects with payoff tables. Implementing payoff tables with more than two subjects and a larger set of actions is difficult.

[20] Estimates of yearly cartel detection lie between 10 and 20 percent (Bryant and Eckard, 1991; Ormosi, 2014). Random draws prior to the first session determined detection, which was identical across sessions.

[21] Subjects see all numbers rounded to two decimal places–32.69 in this case–and are informed about this rounding in the instructions.

antitrust authorities that enforce a cartel prohibition.[22]

After each period, with probability 0.9, subjects play another period against the same rivals. With probability 0.1, subjects are matched to different subjects before playing the next period. Such random termination, introduced by Roth and Murnighan (1978), is the standard way to implement an infinitely repeated game in the lab (Dal Bó and Fréchette (2018)).[23] This implementation allows a subject to play multiple repeated games–'supergames'–in one session. Random draws prior to the first session determined that each session consists of four supergames, with, respectively, eight, twelve, seven, and four periods.[24] Subjects could not be matched to the same subject in different supergames (perfect stranger matching), and their payment was based on all periods of play. A random continuation probability, together with a cumulative payment scheme, induces preferences that are theoretically equivalent to maximizing the discounted sum of utilities with discount factor $\delta = 0.9$.[25]

Table 1: Subject and observation count, by treatment

	REVENUE	PROFIT	OVERCHARGE	Total
Subjects	90	99	90	279
Markets	120	132	120	372
Market-periods	930	1,023	930	2,883
Observations	2,790	3,069	2,790	8,649

Notes: Count of subjects, markets, market-periods, and observations, by treatment.

The computerized experiment was conducted at the Center for Research in Experimental Economics and political Decision making (CREED) of the University of Amsterdam in September 2023 using oTree (Chen et al., 2016). Students were recruited by public announcement. In total, 279 students, mainly from the university's undergraduate population, participated across 21 sessions covering the three treatments. Each session had either 9 or 18 participants.[26] We employed a between-subject design–each subject participated in only

[22] Europe, North America, and many countries in Africa, Asia, Oceania, and South America.

[23] Random termination rules are commonly used in oligopoly experiments (e.g., Bigoni et al. (2012, 2015); Fonseca et al. (2022)). Alternatively, a fixed number of periods followed by a random termination rule has been used (e.g., Hinloopen et al. (2020)).

[24] Detection was similarly determined to occur in period six of the first supergame, periods two and ten of supergame two, periods five and six of supergame three, and never in the final supergame.

[25] This theoretical equivalence requires risk neutrality. However, Sherstyuk et al. (2013) provide evidence that subjects' behavior in infinitely repeated games does not change if the payoff scheme is altered to allow for deviations from risk neutrality.

[26] The share of sessions with only 9 subjects was equal across treatments.

one treatment. At the start of each session, matching groups of nine subjects were randomly formed. These groups did not change during the sessions. In each supergame, subjects were randomly re-matched to subjects they had never faced before in their matching group. Before the first supergame was played, subjects completed a test measuring their risk attitude, the outcome of which was communicated to them after the final supergame had finished (details are in Appendix C). Table 1 lists the number of subjects, supergames, and observations across treatments.

Sessions took 70-90 minutes to complete. Subjects earned points which were exchanged for euros at the end of the experiment at the rate of 300 points per euro. In addition, subjects received a show-up fee of 7 euros. In the rare occurrence of a loss, subjects were still paid the 7 euro show-up fee.[27] Average earnings were 16.1 euros per subject. To ensure that all subjects understood the experiment, they had to answer several test questions correctly before the experiment started. The instructions and test questions of REVENUE are in Appendix B.

Table 2: Theoretical predictions, by treatment

	REVENUE	PROFIT	OVERCHARGE
p^C	$\frac{635}{8} \approx 79.38$	73.5	$\frac{482+\sqrt{\frac{1517}{2}}}{8} \approx 63.69$
δ^*	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2(318-\sqrt{\frac{1517}{2}})}{3(318-\sqrt{\frac{1517}{2}})-2(106-\sqrt{\frac{1517}{2}})} \approx 0.81$

Notes: p^C is the joint-profit-maximizing price set by a stable cartel, and δ^* the critical discount factor above which this price can be part of a subgame-perfect Nash equilibrium, derived in Section 2.

3.2 Hypotheses

Table 2 displays the theoretical predictions on prices and critical discount factors which are based on the model in Section 2, the parameters introduced in Section 3.1, and assuming perfect recidivism ($\rho = 1$).[28] Parameters were selected based on simplifying the presentation towards subjects while ensuring that (expected) fines are equalized across treatments and no focal prices emerge that could guide subject behavior. We test the following hypotheses against the null of no differences between treatments.

H1: Market prices are highest in REVENUE and lowest in OVERCHARGE
H2: Stable cartels are least likely in OVERCHARGE, and equally likely in REVENUE and PROFIT

[27] Out of 279 participants, this happened 11 times.

[28] This assumption is regularly made in related work (e.g., Motta and Polo (2003); Chen and Rey (2013)).

H1 and **H2** are derived from claims (i) and (ii) in Proposition 4, respectively. While the price of uncartelized markets is independent of the fining regime, the price of stable cartels ranks according to **H1**. Therefore, if in all treatments, stable cartels are formed in the same fraction of markets, our theoretical model predicts that market prices follow the ranking in **H1**. Notice that for all treatments, the continuation probability in the experiment–0.9–exceeds the critical discount factor, which implies that stable cartels can, in theory, be the norm regardless of fining base.

Given the continuation probability of 0.9, our theoretical model provides no reason to reject the null hypothesis of no differences in cartel stability across treatments for perfect recidivism ($\rho = 1$). However, we posit as an alternative hypothesis regarding cartelization that stable cartels are less likely to emerge in `OVERCHARGE` than in `PROFIT` and `REVENUE`. First, for sufficiently low recidivism rates cartels might be stable in `PROFIT` and `REVENUE`, but not in overcharge. Second, the experimental literature on infinitely repeated games generally finds that the discount factor exceeding the critical discount is a necessary, but not sufficient, condition for coordination. Indeed, this literature suggests that subjects are more likely to cooperate as the discount factor increases beyond the critical discount factor (Dal Bó and Fréchette, 2018). **H2** follows as this difference is smallest in `OVERCHARGE`.

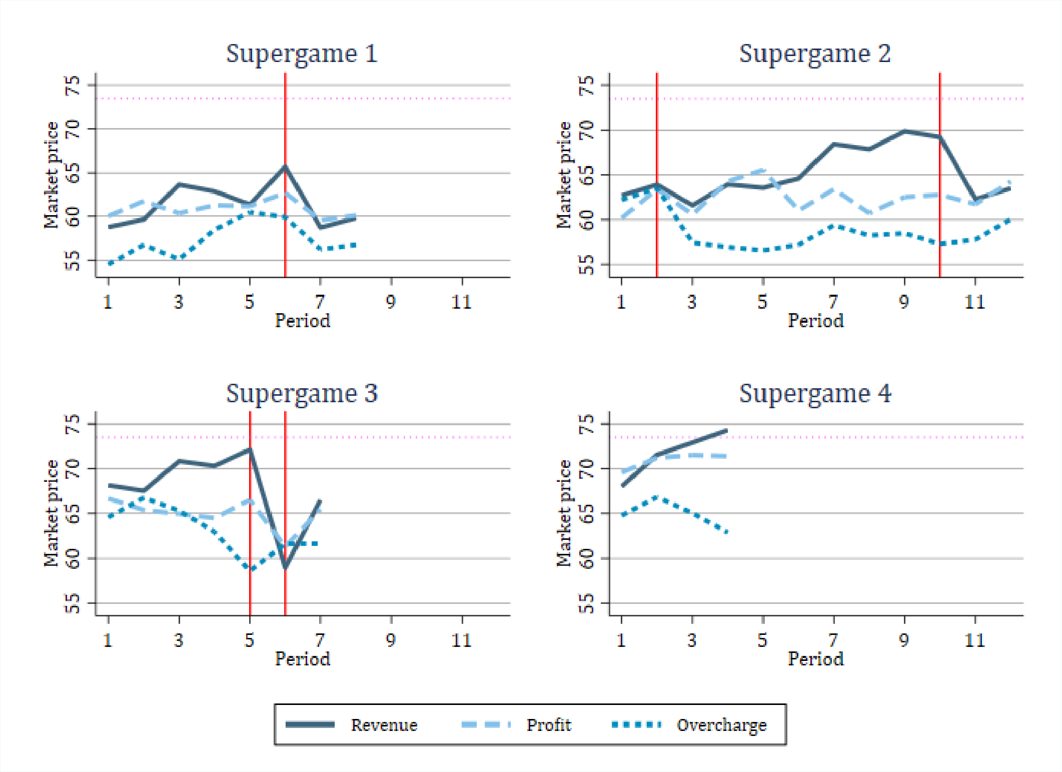
4 Experimental results

In this section, we analyze the data from the experiment. In Section 4.1, we compare `REVENUE`, `PROFIT`, and `OVERCHARGE` in terms of market prices and submitted prices. Section 4.2 presents the relative performance of the three fining regimes in terms of measures of cartelization.[29] We show that differences in prices across treatments follow our theoretical predictions and are driven by differences in prices of cartels rather than differences in the prevalence of cartels or prices in uncartelized markets. In Sections 4.3 and 4.4, therefore, we use the communication data to show that our aggregate results on pricing originate in the pricing of stable cartels rather than differences in cartel stability.[30]

[29] Throughout this section, a cartel is said to exist in a market-period if the chat is active. This aligns with the experimental literature and legal practice, where explicit attempts to coordinate are typically of central importance (Motta, 2004).

[30] We use the Kruskal-Wallis (KW) test to compare all three treatments, Mann-Whitney U (MWU) test for pairwise comparisons across treatments, and the Wilcoxon signed-rank test for within-treatment comparisons. All tests are two-sided, with the average of a variable within a matching group taken as one independent observation in the non-parametric tests. All main results are robust to using less conservative approaches such as regressions with market, market-period or subject-period level data–depending on the outcome–while clustering the standard errors at the matching-group level.

Figure 2: Market price over time, by treatment and supergame



Notes: Average market price over time, by treatment and supergame. Market price = lowest submitted price in a market-period. Red vertical lines indicate a period at the end of which all cartels are detected. Dotted horizontal line indicates the monopoly price of 73.5.

4.1 Prices

Figure 2 plots market prices over time by treatment and supergame, and Table 3 presents the aggregate results on prices across fining regimes. Market prices substantially exceed the one-shot Nash equilibrium price of 47 in all periods of all treatments. Market prices are typically highest in REVENUE (22 out of 31 periods) and lowest in OVERCHARGE (26 out of 31 periods). With experience, subjects learn to set higher prices. The range of average market prices across treatments at the start of each supergame steadily increases, rising from approximately 55-60 in the first supergame to approximately 65-70 by the final supergame. While market prices tend upward over time, they typically decrease to similar levels in all treatments in the period immediately following cartel detection.

Market prices in REVENUE (65.59) and PROFIT (63.74) are higher than market prices in OVERCHARGE (60.14) ($p = 0.005$ and $p = 0.072$, respectively). The concomitant submitted prices, 68.35, 66.25, and 62.33, compare similarly ($p = 0.000$ and $p = 0.030$, respectively). While both price measures are higher in REVENUE than in PROFIT, these differences are not significant at conventional significance levels. Market prices could differ across treatments

Table 3: Prices, across treatments

	Market price	Submitted price	Market price	Market price
	(All markets)	(All markets)	(Cartels)	(Competitive)
REVENUE	65.59 (12.97)	68.35 (12.08)	71.50 (9.08)	51.21 (9.21)
	$\vee$	$\vee$	$\vee^{**}$	$\vee$
PROFIT	63.74 (12.50)	66.25 (11.31)	67.41 (10.73)	50.88 (9.36)
	$\vee^{*}$	$\vee^{**}$	$\vee$	$\vee$
OVERCHARGE	60.14 (11.94)	62.33 (11.99)	64.60 (11.01)	48.82 (4.26)
	$\wedge^{***}$	$\wedge^{***}$	$\wedge^{***}$	$\wedge^{*}$
REVENUE	65.59 (12.97)	68.35 (12.08)	71.50 (9.08)	51.21 (9.21)
KW test	$p = 0.023$	$p = 0.004$	$p = 0.002$	$p = 0.197$

Notes: Table 3 compares prices across treatments; Market price = lowest submitted price in a market-period; Submitted price = price submitted by a subject in a market-period; Cartels = market-periods with a cartel; Competitive = market-periods without a cartel; Standard deviation in brackets; Bottom row reports Kruskal-Wallis p-value; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively (MWU test, two-sided).

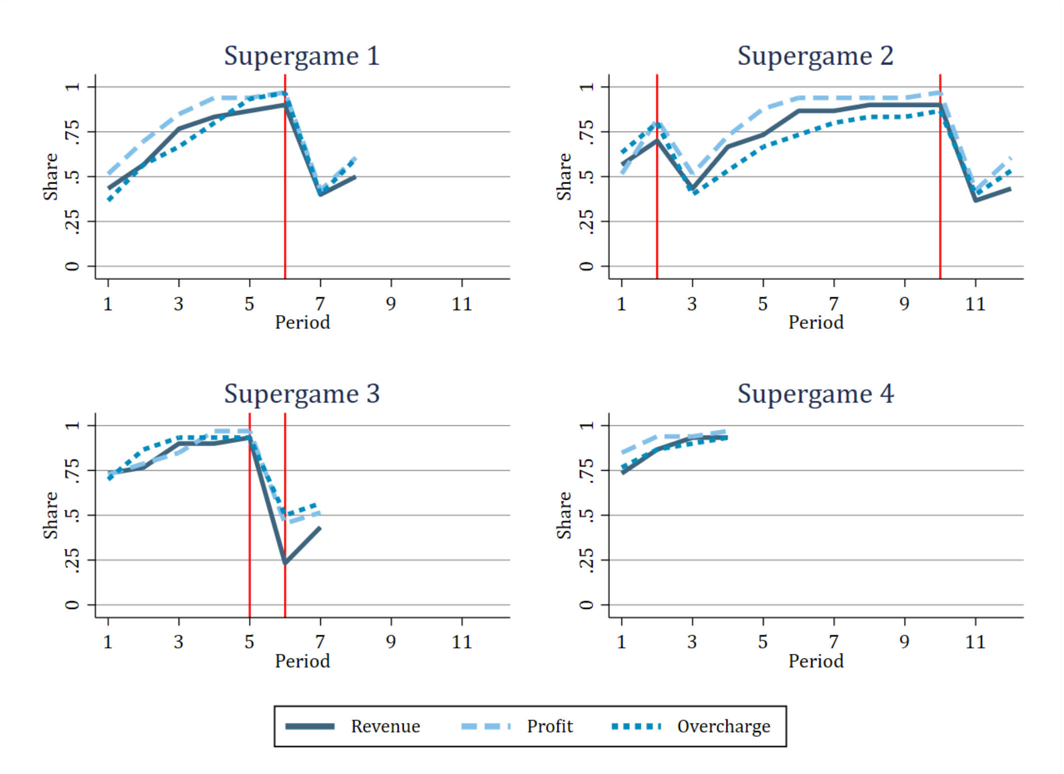
due to differences in cartelization, cartel prices, and prices in uncartelized markets.

Cartel prices in REVENUE (71.50) exceed those in PROFIT (67.41) and OVERCHARGE (64.60) (p = 0.030 and p = 0.000, respectively). In line with the theoretical predictions, market prices exceeding the monopoly price are most common when fines are based on revenue.[31] Market prices in uncartelized markets lie between 51.21 in REVENUE and 48.82 in OVERCHARGE. This is only somewhat above the one-shot Nash equilibrium price of 47, suggesting that subjects do not manage to collude tacitly, and consistent with previous work on repeated Bertrand experiments with more than two players (e.g., Dufwenberg and Gneezy (2000); Fonseca and Normann (2012)). Results are even more in line with the prediction when subjects gain more experience–by the last two supergames–as then cartel prices in all treatments are significantly different while none of the differences in market prices of uncartelized markets are significant.

Summing up, we conclude that the data are in line with alternative hypothesis **H1**: market prices are highest in REVENUE and lowest in OVERCHARGE. While uncartelized markets yield prices close to the one-shot Nash equilibrium price in all three fining regimes,

[31] When fines are based on revenue, 33.98 percent of all market prices exceed the monopoly price of 73.5, significantly more often than 17.11 percent when fines are based on profit and 10.75 when fines are based on the overcharge (p = 0.049 and p = 0.008, respectively; p = 0.564 when comparing profit to overcharge bases).

Figure 3: Cartel incidence over time, by treatment and supergame



Notes: Average cartel incidence over time, by treatment and supergame. Cartel incidence = indicator for a cartel in a market-period. Red vertical lines indicate a period at the end of which all cartels are detected.

cartel prices are highest when fines are based on revenue and lowest when they are based on the overcharge. Moreover, market prices above the monopoly price are common when fines are based on revenue. We next turn to an additional factors that might contribute to the observed difference in market prices across treatments: differences in cartelization.

4.2 Cartel formation, incidence, and recidivism

Figure 3 displays cartel incidence over time by treatment and supergame, and Table 4 presents the aggregate results on cartel formation, incidence, and recidivism. Cartel incidence follows a near-identical trend over time in the three fining regimes. There is a tendency toward complete cartelization in all supergames–i.e., a tendency for all markets to contain a cartel.[32] As subjects gain experience, cartel incidence in the first period of a supergame increases in all treatments, from roughly 50 percent in the first supergame to

[32] Recall that once subjects in a given market have agreed to form a cartel, that cartel remains active until it is detected, regardless of the subjects' behavior. This implies that cartel incidence can only decline over time following a period where all cartels are detected.

about 75 percent in the final supergame. Detection of cartels causes cartel incidence to decline sharply, often below incidence in the first period of the supergame. However, cartel formation picks up again immediately after detection, suggesting that the effects of detection are short-lived.

There are no statistical differences between cartel incidence in `REVENUE` (0.71), `PROFIT` (0.78), and `OVERCHARGE` (0.72) (p-values lie between 0.433 and 0.986). This suggests that the likelihood that a cartel will be formed in an uncartelized market-period is equal across treatments. Indeed, cartel formation rates when fines are based on revenue (0.40), profit (0.50), or the price overcharge (0.43) do not differ significantly (p-values lie between 0.557 and 0.739). Hence, it is unsurprising that the probability with which a subject votes in favor of cartel formation is very similar across treatments–between 71 and 76 percent across the three fining regimes. These results are unchanged when focusing only on cartel formation in market-periods where detection shut down a cartel in the previous period. Such recidivism averages at 47 percent, ranging from 43 percent in `REVENUE`, to 46 percent in `PROFIT`, to 52 percent in `OVERCHARGE` (p-values lie between 0.243 and 0.959).

Forming a cartel comes with the risk of being fined, so failing to balance subjects' risk preferences across treatments might drive results rather than the fining regime. However, Figure C1 in Appendix C shows that the distribution of elicited risk preferences is highly similar across treatments. Indeed, the average of our risk measure across subjects in `REVENUE`, `PROFIT`, and `OVERCHARGE` does not differ significantly (p-values lie between 0.565 and 0.850). As all but one subject participates in a cartel at some point in the experiment, average differences between the risk preferences of cartel members are also absent. Finally, the average of elicited risk preferences over all cartel observations does not differ significantly between the three treatments (p-values between 0.512 and 0.971), suggesting that there are no between-treatment differences in *when* subjects with a particular appetite for risk form a cartel. We conclude that risk-preference-based selection into cartels does not differ across treatments.

That our measures of cartelization do not differ between treatments is surprising given the experimental literature on infinitely repeated games, as the difference between the discount factor implemented in the laboratory and the critical discount factor is found to be an strong predictor of cooperation (Dal Bó and Fréchette, 2018). However, the results surveyed by Dal Bó and Fréchette (2018) are based on infinitely repeated games without the opportunity to communicate, a setting where even experienced subjects rarely achieve sustained cooperation. Our results suggest that when subjects can use unrestricted communication, sustained cooperation is achieved regardless of the distance between the actual and critical discount factor, a finding which has not been pointed out by prior experimental work on collusion as critical discount factors are typically equalized across treatments.

We interpret our results on cartelization as aligning with the null hypothesis of no differences rather than the alternative hypothesis **H2**: stable cartels are equally likely in all

Table 4: Measures of cartelization, across treatments

	Incidence	Formation	Voting	Recidivism
REVENUE	0.71 (0.45)	0.40 (0.49)	0.71 (0.45)	0.43 (0.50)
	∧	∧	∧	∧
PROFIT	0.78 (0.42)	0.50 (0.50)	0.76 (0.43)	0.46 (0.50)
	∨	∨	∨	∨
OVERCHARGE	0.72 (0.45)	0.43 (0.50)	0.71 (0.45)	0.52 (0.50)
	∨	∨	=	∨
REVENUE	0.71 (0.45)	0.40 (0.49)	0.71 (0.45)	0.43 (0.50)
KW test	$p = 0.740$	$p = 0.777$	$p = 0.611$	$p = 0.434$

Notes: Table 4 compares measures of cartelization across treatments; Incidence = indicator for a cartel in a market-period; Formation = indicator for cartel formation in a market-period; Voting = indicator for a vote in favor of a cartel in a market-period; Recidivism = indicator for formation of a cartel in a market the period after it has been detected; Standard deviation in brackets; Bottom row reports Kruskal-Wallis p-value; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively (MWU test, two-sided).

treatments. Together with the fact that market prices in uncartelized markets do not differ across fining regimes either, this implies that all observed variation in market prices across treatments originates in differences between the market prices set by cartels. However, recall that a cartel is said to exist whenever the chat is active. Therefore, to determine the drivers of cartel prices and accurately classify cartel stability, we next turn to the contents of the discussions between cartel members. This allows us to determine whether differences in cartel agreements or cartel stability cause differences in cartel prices across treatments.

4.3 Classifying cartel agreements

The fact that subjects form cartels and send chat messages to each other does not imply that cartel members form agreements on which prices to set. We, therefore, employed a third party with knowledge of competition policy to classify all chat messages to determine whether subjects had agreement to set a specific price. We use two definitions of cartel agreements. An 'explicit price agreement' to set price p in a given period is said to exist if at least one subject proposes price p, and all other subjects explicitly agree before any subject leaves the chat. We also employ a broader definition of cartel agreements–labeled 'price agreements'–that adds implicit agreements where the context makes it clear that all subjects agree on a particular price. An example of an implicit agreement is one subject commenting "great! let's keep it going" after several periods of successful coordination, followed by "ok" from the other two subjects. A detailed explanation of how chat data were classified, as well as several illustrative examples of chat contents, is in Appendix D.[33]

Explicit price agreements are present in 862 of all 2122 market-periods with a cartel (40.62 percent). When an explicit agreement is in place the average market price is 70.48 and subjects manage to successfully coordinate on a market price above the one-shot Nash equilibrium market price of 47 in 716 market-periods (83.06 percent of all explicit price agreements), suggesting that our measure captures cartel agreements. However, cartels without explicit price agreements still manage to coordinate on market prices above 47 in 55.63 percent of all market-periods (701 of 1260), compared to only 10.78 percent in uncartelized market-periods (82 out of 761). Unsurprisingly, therefore, the average market price in cartelized markets without explicit price agreements (65.97) substantially exceeds that of uncartelized markets (50.29). Moreover, Figure E1 in Appendix E shows that the fraction of explicit price agreements tends downward within supergames, while the incidence of successful price coordination does not. These findings suggest that our measure of explicit price agreements substantially underestimates the frequency of cartel agreements, prompting us to construct a broader measure: 'price agreements.'

Price agreements are present in 73.7 percent of all market-periods with a cartel (1564 of 2122 instances). The average market price of cartelized markets with price agreements is 71.28, while market prices of cartelized markets without price agreements (58.05) are now much closer to those of uncartelized markets (50.29). Figure E1 in Appendix E shows that the incidence of price agreements tracks the movement of successful coordination over time, while the level is higher, which suggests that our broader measure of cartel agreements is substantially more accurate than explicit price agreements alone. Hence, we first present results using our broad measure of price agreements, after which we show that our main results are robust to alternative approaches.

Figure E2 in Appendix E reveals no apparent differences in price agreement incidence between treatments. Moreover, the shares of cartelized market-periods with price agreements in `REVENUE` (0.73), `PROFIT` (0.72), and `OVERCHARGE` (0.76) do not differ significantly, as is the case for the average risk aversion in market-periods with price agreements.[34] Therefore, differences in cartel pricing, which determines our aggregate results on market prices, are likely to be explained by the behavior of cartels with price agreements, and such cross-

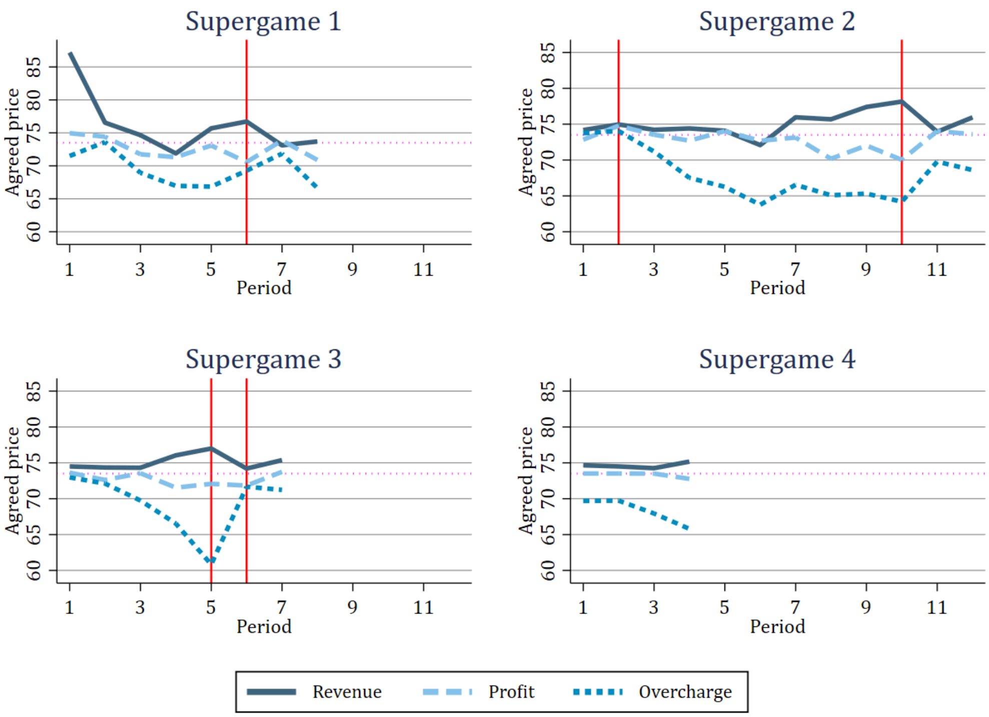
[33] An alternative recent approach is to employ statistical classification techniques to analyze chat data (Andres et al., 2023). We refrain from this exercise as chat contents are often rather sparse, particularly after several periods of stable collusion.

[34] Comparisons of price agreement incidence result in p-values between 0.796 and 0.989, while p-values are between 0.314 and 0.912 for pairwise comparisons across treatments of average risk aversion in market-periods with price agreements.

treatment differences are unlikely to be driven by selection on risk preference.

While price agreements are the norm, note that in cartelized market-periods without price agreements, market prices when fines are based on revenue (64.33) are higher than when fines are based on profit (55.19) or the overcharge (54.88) ($p = 0.024$ and $p = 0.002$, respectively). Recall that, according to theory, when fines are based on revenue, firms that are still detectable but no longer coordinate prices with other firms set price $p_R^{PD} = \frac{c}{1-\alpha r_R} = 58.75$. Hence, one explanation for our aggregate results, albeit minor, given the prominence of price agreements, is that basing fines on revenue leads to higher market prices even when cartels fail to reach a price agreement.[35]

Figure 4: Price agreements over time, by treatment and supergame



Notes: Average price agreement over time, by treatment and supergame. Price agreement = price that a cartel agrees to set in a market-period. Red vertical lines indicate a period at the end of which all cartels are detected. Dotted horizontal line indicates the monopoly price of 73.5.

[35] This is unlikely to be the result of our cartel agreement categorization being too conservative, as in REVENUE cartels without an agreement only coordinate on market prices above 47 in 10 out of 124 market-periods (8.06 percent), which is less than the incidence of such coordination in the absence of cartels (10.78 percent). Learning effects are another unlikely explanation, as the result persists in the final two supergames.

4.4 Price agreements and cartel stability

Figure 4 displays the prices that cartel members agree to set over time, by treatment and supergame. Table 5 presents our aggregate results on price agreements and cartel stability. Price agreements in `REVENUE` (75.06) are significantly higher than those in `PROFIT` (72.79) and `OVERCHARGE` (68.46) ($p = 0.005$ and $p = 0.001$, respectively), as well as the monopoly price of 73.5 ($p = 0.017$, one-sided t-test). When fines are based on profit, price agreements are higher than when fines are based on the overcharge ($p = 0.001$). The ranking of price agreements, therefore, is in accordance with the theoretical predictions.

The degree to which price agreements translate into market prices depends on cartel stability. A cartel is said to be stable if all three subjects set the agreed-upon price. The fraction of price agreements that are adhered to is high in all treatments, ranging from 0.80 in `REVENUE` to 0.89 in `OVERCHARGE` ($p = 0.036$). Over the four supergames, the fraction of cartels that is stable tends towards one (see Figure E3 in Appendix E. Therefore, our aggregate results on market prices are primarily driven by the prices that stable cartels agree to set–particularly as time progresses. As a result, market prices in market-periods with a price agreement in place in `REVENUE` (74.20) are significantly higher than those in `PROFIT` (72.10) and `OVERCHARGE` (67.60) ($p = 0.013$ and $p = 0.000$, respectively), and market prices in `PROFIT` exceed those in `OVERCHARGE` ($p = 0.005$).

Table 5: Price agreements and cartel stability, across treatments

	Agreement incidence	Price agreement	Cartel stability	Market price
	(Cartels)	(Cartels with a price agreement in place)		
`REVENUE`	0.73 (0.45)	75.06 (5.49)	0.80 (0.40)	74.20 (6.07)
	$\vee$	$\vee^{***}$	$\wedge$	$\vee^{**}$
`PROFIT`	0.72 (0.45)	72.79 (4.95)	0.82 (0.38)	72.10 (5.63)
	$\wedge$	$\vee^{***}$	$\wedge$	$\vee^{***}$
`OVERCHARGE`	0.76 (0.42)	68.46 (9.15)	0.89 (0.31)	67.60 (9.48)
	$\vee$	$\wedge^{***}$	$\vee^{**}$	$\wedge^{***}$
`REVENUE`	0.73 (0.45)	75.06 (5.49)	0.80 (0.40)	74.20 (6.07)
KW test	$p = 0.968$	$p = 0.000$	$p = 0.095$	$p = 0.000$

Notes: Table 5 compares measures based on price agreements across treatments; Agreement incidence = Indicator for a cartel with a price agreement in a market-period; Price agreement = Price that the cartel has agreed to set in a market-period; Cartel stability = Indicator for whether all three subjects in a cartel have set the agreed upon price in a market-period; Market price = Lowest submitted price in a market-period; Standard deviation in brackets; Bottom row reports Kruskal-Wallis p-value; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively (MWU test, two-sided).

Price agreements above the monopoly price of 73.5 are common in `REVENUE`. In 283 of all 479 market-periods with a price agreement subjects agree to set such a price (59.1 percent). While price agreements and the fraction of above-monopoly-price agreements are stable over time, the standard deviation of price agreements in `REVENUE` decreases from 6.69 in the first two supergames to 3.40 in the final two supergames as agreements of different cartels converge. Moreover, 11.3 percent of price agreements in `REVENUE` fall between 79 and 80, a percentage that is stable over the supergames. Subjects appear to converge on a price between the monopoly and predicted cartel prices. Therefore, the perverse incentives inherent in revenue-based fines push price agreements above the monopoly price, but to a slightly lesser extent than joint profit maximization would imply.

Price agreements converge to the jointly optimal monopoly price in `PROFIT`. From Figure 4, it is clear that the average price agreement is stable in `PROFIT` and close to 73.5 in all supergames. This average masks significant learning over time. In the first two supergames, price agreements between 73 and 74 characterize 210 of 313 market-periods with an agreement (67.1 percent). This percentage increases to 94.2 in the final two supergames (247 of 262 instances), and the standard deviation of price agreements decreases from 5.94 in the first two supergames to 3.41 in the final two supergames, indicating near-complete convergence to the monopoly price.

While price agreements lie substantially below the monopoly price in OVERCHARGE, most are above the theoretical prediction of 63.69 (382 out of 510). Out of the 128 agreements below 64, 74 come in the first two supergames and only one is between 63 and 64. So, while low prices are not uncommon, we do not observe convergence to the theoretical prediction in overcharge. The standard deviation of price agreements is 9.66 in the first two supergames of OVERCHARGE, and 8.66 in the final two supergames–more than 2.5 times that of the other two treatments–indicating much more limited convergence than in REVENUE or PROFIT. Therefore, while the theoretical predictions for OVERCHARGE hold qualitatively, subjects do not appear to coordinate on the joint-profit maximum.

Figure 4 suggest one possible explanation for why agreements in OVERCHARGE, and to a lesser extent in REVENUE, do not fully converge to the theoretical predictions. Over time, as cartels go undetected in a given supergame, price agreements tend downward in OVERCHARGE, towards the theoretical predictions. Similarly, in REVENUE, price agreements tend up as cartels go undetected. As these patterns are observable in all supergames, learning is an unlikely explanation for these patters. To test this more formally, we run a number of regression using data from the final three supergames (so as to exclude subject learning as an explanation), reported in Tables F1 and F2 in Appendix F

We find that, indeed, price agreements move closer to the theoretical predictions the longer cartels go undetected when fines are based on either revenue or the price overcharge. In addition, cartels with a higher average appetite for risk agree to set higher prices in OVERCHARGE and lower prices in REVENUE–although the final correlation is not statistically significant. Taken together, these results suggest that some share of the subjects in our sample misunderstand probabilities while acting in accordance with the incentives provided by the different fining regimes. More specifically, subjects might initially under-assess the probability of being detected, and as a result agree to set prices close to the monopoly price. As the cartel goes undetected, subject might increasingly become concerned about detection and adjust their agreements closer to the theoretical predictions. This process can help explain why full convergence of price agreements to the theoretical predictions is not achieved, even though such convergence does appear to occur within individual supergames when detection does not occur for several periods. Indeed, the chat contents suggest such forces are in play, one example is provided by Table D3 in Appendix D.

Before discussing the implications of our results, we stress that our findings in this section do not depend on the classification of cartel agreements. Table F3 in Appendix F shows that our results are robust to using explicit price agreements instead of our broader measure of price agreement. In an additional robustness check we do away with the chat data altogether and focus on cartel-periods where cartel members successfully coordinate on a price above marginal costs. While defining agreements and stability for all cartels is not possible in this approach, it most likely tracks the incidence of stable cartels and their prices quite closely as coordination is rare in the absence of a cartel so that tacit collusion appears unlikely (see

Figure E4 in Appendix E). Table F4 in Appendix F shows that the frequency with which such cartels coordinate on prices above marginal costs is equal across treatments, and that the ranking of cartel prices mirrors that of Table 5.

5 Concluding remarks

Using a laboratory experiment, we have investigated the relative performance of three bases for cartel fines: revenue, profit, and price overcharge. While we observe no significant differences across treatments regarding cartel formation, incidence, or recidivism, we find that average prices are lowest under overcharge-based fines and highest under revenue-based fines. Notably, subjects in the experiment frequently agree to set prices above the monopoly price when fines are based on revenue. The behavior of subjects in the lab aligns well with our theoretical assumptions on equilibrium selection. In particular, price agreements exceed the monopoly price when fines are based on revenue, a feature not shared by fines based on profit or overcharge. Consequently, subjects in the REVENUE treatment coordinate on Nash equilibria that reduce consumer welfare more than the other fining regimes. While active antitrust enforcement can temporarily reduce the negative effects of revenue-based fines by dismantling cartels, this effect is short-lived as cartels tend to be re-established regardless of how often detection has previously occurred.

Therefore, the main policy recommendation from our study is to move away from revenue-based antitrust fines. Both profit- and overcharge-based regimes outperform revenue-based fines in our experiment, with the overcharge regime leading to the lowest prices. However, the preferred alternative also depends on the feasibility of implementation. Indeed, the primary argument for using a revenue-based regime is ease of implementation, as information on firm- or product-level revenue is relatively simple to obtain.

Determining the overcharge requires an estimate of the counterfactual price. Katsoulacos et al. (2019) argue that an overcharge-based regime is feasible as counterfactual price calculations are routinely made in cartel damage cases. If fines are accompanied by an estimate of the counterfactual price that can be used in future damage cases, the uncertainty and cost of such cases would substantially drop.[36] However, a clear downside of such calculations is the lengthy duration of damage cases, which could substantially delay the first monetary penalty for firms engaged in price-fixing. Moreover, in complex cases where cartels coordinate on factors other than the price, and in jurisdictions where indirect purchasers have standing to sue for damages, it is not always clear that the price overcharge accurately reflects actual damages.

A profit-based fine is, therefore, a particularly attractive alternative. As discussed in

[36] Rüggeberg and Schinkel (2006) previously argued in favor of consolidating damage estimation for such reasons.

Section 2.3, a measure of firm profit is preferable to one based on the incremental profit generated by the cartel. Such measures should be relatively straightforward to obtain in a reasonable timeframe–compared to estimating the overcharge–and can also be applied when cartel members collude on factors other than prices. Since the overcharge-based and–particularly–profit-based regimes appear feasible alternatives, we conclude that there is little justification for continuing to base fines on revenue.

Policymakers may question whether our experimental results generalize to practice. Economic laboratory experiments frequently replicate in the field and with professional participants rather than students (Camerer, 2015; Fréchette, 2015). Notably, Dal Bó and Fréchette (2018, p.88) state: "*There is no robust evidence that risk aversion, economic training, altruism, gender, intelligence, patience, or psychological traits have a systematic effect on cooperation in infinitely repeated games . . . there is evidence consistent with the idea that the main motivation behind cooperation is strategic.*" While lab experiments are neither superior nor inferior to other methods, they are particularly valuable in our case due to the limitations of observational data on cartels. Thus, we contribute to a long tradition of testing theory in the lab, the merits of which have been widely recognized in the literature (e.g., Falk and Heckman (2009); List (2020)).

Finally, we have abstracted from leniency programs and damage claims to closely align with prior theoretical work on fining bases for antitrust penalties and to keep the experiment comprehensible for subjects. However, incorporating such factors would not eliminate the underlying forces driving our results, as long as fines are imposed with positive probability. In fact, under revenue-based fines, the possibility of future damage claims can further exacerbate the distortions created by revenue-based penalties (Katsoulacos et al., 2020). Integrating leniency programs and damage claims into our experiment is a promising avenue for future work, although it should be noted that even when holding the fining base fixed, incorporating damage claims and leniency results in a challenging-to-implement experiment (e.g., Bodnar et al. (2023)).

References

Andres, M., Bruttel, L., and Friedrichsen, J. (2023). How communication makes the difference between a cartel and tacit collusion: A machine learning approach. *European Economic Review*, 152:104331.

Apesteguia, J., Dufwenberg, M., and Selten, R. (2007). Blowing the whistle. *Economic Theory*, 31:143–166.

Aubert, C., Rey, P., and Kovacic, W. E. (2006). The impact of leniency and whistle-blowing programs on cartels. *International Journal of Industrial Organization*, 24(6):1241–1266.

Awaya, Y. and Krishna, V. (2016). On communication and collusion. *American Economic Review*, 106(2):285–315.

Bageri, V., Katsoulacos, Y., and Spagnolo, G. (2013). The distortive effects of antitrust fines based on revenue. *Economic Journal*, 123(572):F545–F557.

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 75(2):169–217.

Bigoni, M., Fridolfsson, S.-O., Le Coq, C., and Spagnolo, G. (2012). Fines, leniency, and rewards in antitrust. *RAND Journal of Economics*, 43(2):368–390.

Bigoni, M., Fridolfsson, S.-O., Le Coq, C., and Spagnolo, G. (2015). Trust, leniency, and deterrence. *Journal of Law, Economics, and Organization*, 31(4):663–689.

Block, M. K., Nold, F. C., and Sidak, J. G. (1981). The deterrent effect of antitrust enforcement. *Journal of Political Economy*, 89(3):429–445.

Bodnar, O., Fremerey, M., Normann, H.-T., and Schad, J. (2023). The effects of private damage claims on cartel activity: Experimental evidence. *Journal of Law, Economics, and Organization*, 39(1):27–76.

Bryant, P. G. and Eckard, E. W. (1991). Price fixing: The probability of getting caught. *Review of Economics and Statistics*, 73(3):531–536.

Buccirossi, P. and Spagnolo, G. (2007). Optimal fines in the era whistleblowers: Should price fixers still go to prison? In Ghosal, V. and Stennek, J., editors, *The Political Economy of Antitrust*. Elsevier.

Byrne, D. P. and De Roos, N. (2019). Learning to coordinate: A study in retail gasoline. *American Economic Review*, 109(2):591–619.

Camerer, C. F. (2015). The promise and success of lab–field generalizability in experimental economics: A critical reply to Levitt and List. In Fréchette, G. and Schotter, A., editors, *Handbook of Experimental Economic Methodology*. Oxford University Press.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chen, Z. and Rey, P. (2013). On the design of leniency programs. *Journal of Law and Economics*, 56(4):917–957.

Chowdhury, S. M. and Crede, C. J. (2020). Post-cartel tacit collusion: Determinants, consequences, and prevention. *International Journal of Industrial Organization*, 70:102590.

Clemens, G. and Rau, H. A. (2022). Either with us or against us: Experimental evidence on partial cartels. *Theory and Decision*, 93(2):237–257.

Cooper, D. J. and Kühn, K.-U. (2014). Communication, renegotiation, and the scope for collusion. *American Economic Journal: Microeconomics*, 6(2):247–78.

Dal Bó, P. and Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56(1):60–114.

Dufwenberg, M. and Gneezy, U. (2000). Price competition and market concentration: An experimental study. *International Journal of Industrial Organization*, 18(1):7–22.

Duso, T., Röller, L.-H., and Seldeslachts, J. (2014). Collusion through joint R&D: An empirical assessment. *Review of Economics and Statistics*, 96(2):349–370.

European Commission (2006). Guidelines on the method of setting fines imposed pursuant to Article 23(2)(a) of Regulation No 1/2003. *Official Journal of the European Union*, September 1.

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538.

Fonseca, M. A., Gonçalves, R., Pinho, J., and Tabacco, G. A. (2022). How do antitrust regimes impact on cartel formation and managers' labor market? An experiment. *Journal of Economic Behavior & Organization*, pages 643–662.

Fonseca, M. A. and Normann, H.-T. (2012). Explicit vs. tacit collusion—The impact of communication in oligopoly experiments. *European Economic Review*, 56(8):1759–1772.

Fonseca, M. A. and Normann, H.-T. (2014). Endogenous cartel formation: Experimental evidence. *Economics Letters*, 125(2):223–225.

Freitag, A., Roux, C., and Thöni, C. (2021). Communication and market sharing: An experiment on the exchange of soft and hard information. *International Economic Review*, 62(1):175–198.

Fréchette, G. (2015). Laboratory experiments: Professionals versus students. In Fréchette, G. and Schotter, A., editors, *Handbook of Experimental Economic Methodology.* Oxford University Press.

Genesove, D. and Mullin, W. P. (2001). Rules, communication, and collusion: Narrative evidence from the sugar institute case. *American Economic Review*, 91(3):379–398.

Gomez-Martinez, F., Onderstal, S., and Sonnemans, J. (2016). Firm-specific information and explicit collusion in experimental oligopolies. *European Economic Review*, 82:132–141.

Harrington Jr., J. E. (2004). Cartel pricing dynamics in the presence of an antitrust authority. *RAND Journal of Economics*, 35(4):651–673.

Harrington Jr., J. E. (2005). Optimal cartel pricing in the presence of an antitrust authority. *International Economic Review*, 46(1):145–169.

Harrington Jr, J. E. (2006). How do cartels operate? *Foundations and Trends in Microeconomics*, 2(1):1–105.

Harrington Jr., J. E. (2014). Penalties and the deterrence of unlawful collusion. *Economics Letters*, 124(1):33–36.

Harrington Jr, J. E., Gonzalez, R. H., and Kujal, P. (2016). The relative efficacy of price announcements and express communication for collusion: Experimental findings. *Journal of Economic Behavior & Organization*, 128:251–264.

Hinloopen, J., Martin, S., Onderstal, S., and Treuren, L. (2024). Spillovers from legal cooperation to non-competitive prices. *Tinbergen Institute Discussion Paper 2024-078/VII.*

Hinloopen, J., Onderstal, S., and Soetevent, A. (2023). Corporate leniency programs for antitrust: Past, present, and future. *Review of Industrial Organization*, 63(2):111–122.

Hinloopen, J., Onderstal, S., and Treuren, L. (2020). Cartel stability in experimental first-price sealed-bid and English auctions. *International Journal of Industrial Organization*, 71:102642.

Hinloopen, J. and Soetevent, A. R. (2008). Laboratory evidence on the effectiveness of corporate leniency programs. *RAND Journal of Economics*, 39(2):607–616.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Houba, H., Motchenkova, E., and Wen, Q. (2018). Legal principles in antitrust enforcement. *Scandinavian Journal of Economics*, 120(3):859–893.

Huck, S., Normann, H.-T., and Oechssler, J. (1999). Learning in Cournot oligopoly–An experiment. *Economic Journal*, 109(454):80–95.

Huck, S., Normann, H.-T., and Oechssler, J. (2000). Does information about competitors' actions increase or decrease competition in experimental oligopoly markets? *International Journal of Industrial Organization*, 18(1):39–57.

Huck, S., Normann, H.-T., and Oechssler, J. (2004). Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53(4):435–446.

International Competition Network (2017). Setting of fines for cartels in ICN jurisdiction. *Report to the 16th International Competition Network Annual Conference.*

Kandori, M. and Matsushima, H. (1998). Private observation, communication and collusion. *Econometrica*, 66(3):627–652.

Katsoulacos, Y., Motchenkova, E., and Ulph, D. (2015). Penalizing cartels: The case for basing penalties on price overcharge. *International Journal of Industrial Organization*, 42:70–80.

Katsoulacos, Y., Motchenkova, E., and Ulph, D. (2019). Penalizing cartels—a spectrum of regimes. *Journal of Antitrust Enforcement*, 7(3):339–351.

Katsoulacos, Y., Motchenkova, E., and Ulph, D. (2020). Combining cartel penalties and private damage actions: The impact on cartel prices. *International Journal of Industrial Organization*, 73:102604.

Katsoulacos, Y. and Ulph, D. (2013). Antitrust penalties and the implications of empirical evidence on cartel overcharges. *Economic Journal*, 123(572):F558–F581.

Kwoka, J. E. and White, L. J., editors (2018). *The Antitrust Revolution: Economics, Competition, and Policy (7th edition).* Oxford University Press.

List, J. A. (2020). Non est disputandum de generalizability? A glimpse into the external validity trial. NBER Working Paper 27535, National Bureau of Economic Research, Cambridge, MA.

Marshall, R. C. and Marx, L. M. (2012). *The Economics of Collusion: Cartels and Bidding Rings.* MIT Press.

Marvão, C. and Spagnolo, G. (2018). Cartels and leniency: Taking stock of what we learnt. In Corchón, L. and Marini, M., editors, *Handbook of Game Theory and Industrial Organization, Volume II.* Edward Elgar Publishing.

McCutcheon, B. (1997). Do meetings in smoke-filled rooms facilitate collusion? *Journal of Political Economy*, 105(2):330–350.

Motta, M. (2004). *Competition Policy: Theory and Practice.* Cambridge University Press.

Motta, M. and Polo, M. (2003). Leniency programs and cartel prosecution. *International Journal of Industrial Organization*, 21(3):347–379.

Normann, H.-T., Rösch, J., and Schultz, L. M. (2015). Do buyer groups facilitate collusion? *Journal of Economic Behavior & Organization*, 109:72–84.

Offerman, T., Potters, J., and Sonnemans, J. (2002). Imitation and belief learning in an oligopoly experiment. *Review of Economic Studies*, 69(4):973–997.

Ormosi, P. L. (2014). A tip of the iceberg? The probability of catching cartels. *Journal of Applied Econometrics*, 29(4):549–566.

Roth, A. E. and Murnighan, J. K. (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17(2):189–198.

Rüggeberg, J. and Schinkel, M. P. (2006). Consolidating antitrust damages in Europe: A proposal for standing in line with efficient private enforcement. *World Competition*, 29(3):295–420.

Sherstyuk, K., Tarui, N., and Saijo, T. (2013). Payment schemes in infinite-horizon experimental games. *Experimental Economics*, 16:125–153.

Sovinsky, M. (2022). Do research joint ventures serve a collusive function? *Journal of the European Economic Association*, 20(1):430–475.

Tirole, J. (1988). *The Theory of Industrial Organization*. MIT press.

United States Sentencing Commission (2023). *Guidelines Manual §2R1.1*. Washington, D.C.

Vega-Redondo, F. (1997). The evolution of Walrasian behavior. *Econometrica*, 65(2):375–384.

Appendices

A Proofs of propositions

Proof of Proposition 1

i) Assume that the stability condition holds. The cartel's price is then given by

$$p_R^C = \arg\max_p (1-\alpha r_R) q(p) \left(p - \frac{c}{1-\alpha r_R}\right) = p^M \left(\frac{c}{1-\alpha r_R}\right) > p^M(c).$$

The inequality in the above chain follows because p^M is a strictly increasing function. Notice that the cartel acts like a monopolist in the absence of antitrust facing marginal cost $\frac{c}{1-\alpha r_R}$ rather than c. We denote the corresponding firm-level profit and fine by π_R^C and F_R^C respectively.

ii)-iii) As firms facing a revenue-based fine act as if they have marginal cost $\frac{c}{1-\alpha r_R}$ and face no threat of fines, the optimal defection is to slightly undercut p_R^C and capture the entire market. This increases both the defector's before-fine profit π_R^D and its fine F_R^D n-fold compared to the cartel case: $\pi_R^D = n\pi_R^C$ and $F_R^D = nF_R^C$. Post-defection, if the cartel has not been convicted yet, the possibility of being fined implies that expected profit is negative if all firms set price equal to marginal cost. Instead, the unique symmetric pure-strategy one-shot Nash equilibrium market price equals $p_R^{PD} = \frac{c}{1-\alpha r_R}$. As before, the revenue-based fine incentivizes firms to act like firms facing marginal cost $\frac{c}{1-\alpha r_R}$ in the absence of antitrust enforcement. Expected profit is 0 if all firms set p_R^{PD}, $\pi_R^{PD} - \alpha F_R^{PD} = 0$, and as $\pi^N = 0$, the expected net present value of deviation equals $V_R^D = n\left(\pi_R^C - \alpha F_R^C\right)$.

The critical discount factor, δ_R^*, is the δ such that $\sigma_R(\alpha, \delta, \rho) = 1$, where $\sigma_R(\alpha, \delta, \rho) \equiv \frac{V_R^C}{V_R^D} = \frac{f(\alpha,\delta,\rho)}{n(1-\delta)}$, V_R^C is given in equation (1) and $f(\alpha, \delta, \rho)$ in equation (2). Let $\sigma_R^* \equiv \sigma_R(\alpha, \delta_R^*, \rho)$. By the implicit function theorem, we have

$$\frac{\partial \delta_R^*}{\partial \alpha} = -\frac{\partial \sigma_R^* / \partial \alpha}{\partial \sigma_R^* / \partial \delta}.$$

Since $\frac{\partial f(\alpha,\delta,\rho)}{\partial \delta} = \frac{-\alpha(1-\rho)}{(1-\delta(1-\alpha)(1-\rho))^2}$, $\frac{\partial f(\alpha,\delta,\rho)}{\partial \alpha} < 0$ if $\rho \in [0, 1)$, and $\frac{\partial f(\alpha,\delta,\rho)}{\partial \alpha} = 0$ if $\rho = 1$, it follows that

$$\frac{\partial \sigma_R}{\partial \delta} = \frac{\partial f(\alpha,\delta,\rho)/\partial \delta}{n(1-\delta)} + \frac{f(\alpha,\delta,\rho)}{n(1-\delta)^2} > 0 \quad \text{and} \quad \frac{\partial \sigma_R}{\partial \alpha} = \frac{\partial f(\alpha,\delta,\rho)/\partial \alpha}{n(1-\delta)} \begin{cases} = 0 & \text{if } \rho = 1, \\ < 0 & \text{if } \rho \in [0,1). \end{cases}$$

■

Proof of Proposition 2

i) Assume that the stability condition holds. The cartel's price is then given by

$$p_\pi^C = \arg\max_p (1-\alpha r_\pi)(p-c)\, q(p) = p^M(c).$$

Let π_π^C and F_π^C denote the corresponding firm-level profit and expected fine respectively.

ii)-iii) As a profit-based fine does not alter a firm's incentives compared to the no-antitrust case, the optimal defection is to slightly undercut the monopoly price and capture the entire market. This increases the defector's before-fine profit and fine n-fold compared to the cartel case: $\pi_\pi^D = n\pi_\pi^C$ and $F_\pi^D = nF_\pi^C$. Post-defection, regardless of whether the cartel has been convicted, the firms revert to the Nash equilibrium of the static Bertrand game, so $p_\pi^{PD} = c$ and $\pi_\pi^{PD} = 0$. Together with $\pi^N = 0$ and the preceding paragraph, this implies that $V_\pi^D = n\left(\pi_\pi^C - \alpha F_\pi^C\right)$.

Denote the critical discount factor by δ_π^*. As $V_\pi^C = V_R^C$ and $V_\pi^D = V_R^D$ it follows immediately from the proof of Proposition 1 that

$$\frac{\partial \delta_\pi^*}{\partial \alpha}\begin{cases} = 0 & \text{if } \rho = 1, \\ > 0 & \text{if } \rho \in [0,1). \end{cases}$$

■

Proof of Proposition 3

i) Assume that the stability condition holds. For certain parameters, overcharge-based fines allow for stable cartels but constrain the cartel price. Consider first the behavior of a stable cartel whose pricing is not constrained by the stability condition in (3). This 'unconstrained' cartel price is then given by

$$p_O^{UC} = \arg\max_p \,(p-c)\,q(p) - \alpha r_O\left(p - p^N\right)q(p^N) < p^M(c).$$

The inequality follows from the first-order condition underlying the maximization problem: $(p-c)\frac{\partial q(p)}{\partial p} + q(p) - \alpha r_O q(p^N) = 0$. The first two terms define the monopoly price. The overcharge-based fine introduces the third term, which incentivizes the cartel to set a price below $p^M(c)$ to reduce the expected fine and increase expected profit. The difference between $p^M(c)$ and p_O^U increases with the detection probability, the penalty rate, and the Nash-equilibrium quantity of the static Bertrand game, none of which are influenced by the cartel's pricing decision.

To see that the unconstrained cartel price is strictly above marginal cost, rewrite the cartel's maximization problem as

$$\max_q (p(q)-c)(q - \alpha r^O q(p^N)).$$

The associated first-order condition is $p(q) + \frac{\partial p(q)}{\partial q}(q - \alpha r^O q(p^N)) = c$. Note that the left-hand side of the first-order condition is equal to $p(q)$ if $q = \alpha r^O q(p^N)$ and lies below inverse demand $p(q)$ for higher values of q, implying that $p_O^U > c$.

Note that an individual firm's fine is scaled by their share of the competitive output $q^N = \frac{q(p^N)}{n}$ instead of $q(p^N)$. A defector is, therefore, faced with the first-order condition

$(p-c)\frac{\partial q(p)}{\partial p} + q(p) - \alpha r_O \frac{q(p^N)}{n} = 0$. Comparison to the first-order condition in the first paragraph of this Proof shows that a defector would ideally increase the price compared to the cartel. However, this would result in no demand, so the best a defector can do is to slightly undercut the cartel's price and capture the entire market. This increases the defector's before-fine profit n-fold while leaving the fine unchanged. Compared to the cartel case: $\pi_O^D = n\pi_O^C$ and $F_O^D = F_O^C$. Post-defection, regardless of whether the cartel has been convicted, firms revert to the Nash-equilibrium of the static Bertrand game, so $p_O^{PD} = c$ and $\pi_O^{PD} = 0$. Together with $\pi^N = 0$ and the preceding paragraph, this implies that $V_O^D = n\pi_O^C - \alpha F_O^C$.

Let $\sigma_O(\alpha, \delta, \rho, p) \equiv \frac{V_O^C}{V_O^D} = \frac{\left(q(p)/n - \alpha r_O q^N\right) f(\alpha,\delta,\rho)}{(q(p) - \alpha r_O q^N)(1-\delta)}$. Comparing σ_O to σ_R and σ_π, note that $\frac{\partial \sigma_R}{\partial p} = \frac{\partial \sigma_\pi}{\partial p} = 0$, while

$$\frac{\partial \sigma_O}{\partial p} = \frac{\partial q(p)}{\partial p} \frac{(1 - 1/n)\alpha r_O q^N f(\alpha, \delta, \rho)}{\left(q(p) - \alpha r_O q^N\right)^2 (1-\delta)} < 0,$$

as demand strictly decreases with price. In addition, since $\frac{\partial f(\alpha,\delta,\rho)}{\partial \delta} = \frac{-\alpha(1-\rho)}{(1-\delta(1-\alpha)(1-\rho))^2}$

$$\frac{\partial \sigma_O}{\partial \delta} = \underbrace{P}_{>0} \left(\underbrace{\frac{\partial f(\alpha, \delta, \rho)/\partial \delta}{(1-\delta)} + \frac{f(\alpha, \delta, \rho)}{(1-\delta)^2}}_{>0} \right) > 0,$$

where $P = \frac{q(p)/n - \alpha r_O q^N}{q(p) - \alpha r_O q^N}$. The critical discount factor, δ_O^*, is the δ such that $\sigma_O\left(\alpha, \delta, \rho, p_O^{UC}\right) = 1$. Define $\bar{\delta}_O$ as the delta such that $\sigma_O\left(\alpha, \delta, \rho, c\right) = 1$. As $p_O^{UC} > c$, $\delta_O^* > \bar{\delta}_O$. Consider parameter values such that $\sigma_O\left(\alpha, \delta, \rho, p_O^{UC}\right) = 1$, and then decrease δ slightly. The above implies that p_O^{UC} can no longer be part of a subgame-perfect Nash equilibrium, but that $\sigma_O\left(\alpha, \delta, \rho, p\right) = 1$ can be restored by the cartel selecting a 'constrained' cartel price p_O^{CC} that is strictly below p_O^{UC}. The price of stable cartels in the overcharge regime is, therefore, always strictly below the monopoly price and given by

$$p_O^C = \begin{cases} p_O^{UC} & \text{if } \delta_O^* \leq \delta < 1, \\ p_O^{CC} & \text{if } \bar{\delta}_O < \delta < \delta_O^*. \end{cases}$$

ii) By the implicit function theorem, we have

$$\frac{\partial \delta_O^*}{\partial \alpha} = -\frac{\partial \sigma_O^*/\partial \alpha}{\partial \sigma_O^*/\partial \delta}.$$

As before, let $P = \frac{q(p)/n - \alpha r_O q^N}{q(p) - \alpha r_O q^N}$. As $\frac{\partial \sigma_O}{\partial \delta} > 0$ by the proof of Proposition 3i), and as $\frac{\partial f(\alpha,\delta,\rho)}{\partial \alpha} \leq 0$, it follows that

$$\frac{\partial \sigma_O}{\partial \alpha} = \underbrace{\frac{\partial P}{\partial \alpha}}_{<0} \underbrace{\frac{f(\alpha, \delta, \rho)}{1-\delta}}_{>0} + \underbrace{\frac{P}{1-\delta}}_{>0} \underbrace{\frac{\partial f(\alpha, \delta, \rho)}{\partial \alpha}}_{\leq 0} < 0,$$

so that $\frac{\partial \delta_O^*}{\partial \alpha} > 0$, for all $\rho \in [0, 1]$.

■

Proof of Proposition 4

i) Follows directly from Propositions 1 to 3.

ii) By the proofs of Propositions 1 and 2, $V_R^C = V_\pi^C$ and $V_R^D = V_\pi^D$, so that $\delta_R^* = \delta_\pi^*$. By the proofs of Propositions 2 and 3, showing that $\sigma_O(\alpha, \delta, \rho, p) < \sigma_R(\alpha, \delta, \rho)$ (defined in the same proofs) is sufficient to show that $\delta_O^* > \delta_R^*$ for all jointly-optimal cartel prices. Because $\sigma_O(\alpha, \delta, \rho, p) = \frac{\left(q(p)/n - \alpha r_O q^N\right) f(\alpha,\delta,\rho)}{(q(p) - \alpha r_O q^N)(1-\delta)}$ and $\sigma_R(\alpha, \delta, \rho) = \frac{f(\alpha,\delta,\rho)}{n(1-\delta)}$, we have

$$\sigma_O(\alpha, \delta, \rho, p) < \sigma_R(\alpha, \delta, \rho) \iff \frac{q(p)/n - \alpha r_O q^N}{q(p) - \alpha r_O q^N} < \frac{1}{n}. \tag{A1}$$

Note that $\sigma_O(\alpha, \delta, \rho, p) = \sigma_R(\alpha, \delta, \rho)$ if $\alpha r_O q^N = 0$ (i.e., if there is no antitrust enforcement). It follows that $\sigma_O(\alpha, \delta, \rho, p) < \sigma_R(\alpha, \delta, \rho)$ as $\alpha r_O q^N > 0$ and the left-hand side of the final inequality of equation (A1) is strictly decreasing in $\alpha r_O q^N$.

■

B Instructions

Subjects could read through the computerized instructions at their own pace. All test questions needed to be answered correctly for the subject to progress to the experiment. For brevity, we include only the instructions for REVENUE and for the risk preference test–discussed in Appendix C. The instructions for the other treatments are available from the authors upon request.

Introduction

We ask that you do not talk to other people during the experiment. Please refrain from verbally reacting to events that occur during the experiment. The use of mobile phones is not allowed. If you have any questions, or need assistance of any kind, please notify the experimenter by raising your hand.

Please comply with these rules, otherwise you will be asked to leave and you will not be paid.

Your earnings will depend on your decisions, and the decisions of other participants: your rivals. You will be paid privately and in cash at the end of the experiment.

Description of the experiment

In this experiment you will play a game four times. Each game consists of several periods. At the end of each period, there is a 90% chance that another period will be played and a 10% chance that the game ends.

In all periods of a game you will be matched to the same two participants: your rivals. In different games you will have different rivals. You always face the same rivals in different periods of the same game. You never face the same rivals in different games.

In each period of a game you and your rivals pick prices. Before picking prices, you can vote to form a cartel. If you and your two rivals vote in favour of a cartel, a cartel is formed, and you can chat about prices before setting your price. If no cartel is formed, next period you can vote again. Cartels are illegal and there is a 20% chance each period that all active cartels are detected. If your cartel is detected, you will pay a fine. The next period you can vote to form a new cartel. If your cartel is not detected, it is automatically active the next period. The market is described in detail on the next page. All numbers are perioded to two decimal points.

The market

The price you set must be between 0.01 and 99.99 (all inputs are perioded to two decimals). The quantity you sell from setting price p is:

$q = 0$	if you do not set the lowest price
$q = \frac{100-p}{n}$	if n firms set the lowest price

Your before-fine profit from setting price p is:

Before-fine profit $= 0$	if you do not set the lowest price.
Before-fine profit $= (p - 47)\frac{100-p}{n}$	if n firms set the lowest price

Note that setting a price below 47 will result in a loss.

When choosing your price, an on-screen calculator is available, as well as information on the history of the game.

Example 1: Firm 1 and 2 set price 50 and firm 3 sets price 61. Firm 3 did not set the lowest price and makes 0 profit this period. Firms 1 and 2 both set the lowest price so both get a before-fine profit equal to $(50 - 47)\frac{100-50}{2} = 75$.

Example 2: All three firms set price 70. All three firms set the lowest price and so get a before-fine profit equal to $(70-47)\frac{100-70}{3} = 230$.

The next page will give you more information about forming cartels.

Cartels

Before choosing prices, you and your rivals vote to form a cartel. Recall that only if all three vote in favor, a cartel is formed. A cartel gives you access to a chat. In the first period of a cartel, you can chat for 1 minute before setting prices, in other periods you can chat for 30 seconds.

After chatting about prices, you will still need to set a price independently.

Chatting about anything that can be used to identify you in or outside of the lab will result in you not being paid for this experiment.

Forming a cartel is illegal and there is a 20% chance in each period that all active cartels are detected. If your cartel is not detected, the cartel will automatically be active in the next period. If your cartel is detected, you will pay a fine and the cartel is no longer active. If your cartel is detected, you can vote to form a new cartel in the next period.

If your cartel is detected and you set price p in that period, your fine will be:

$\textit{Fine} = 0$ if you did not set the lowest price.

$\textit{Fine} = p\frac{100-p}{n}$ if n firms set the lowest price.

Note that that you cannot be fined if you do not set the lowest price, and that a higher price will not always result in a higher fine. The fine depends on your revenue.

Example 3: All three firms vote to form a cartel. A cartel is formed, and the firms discuss prices. All three firms set a price of 91 and get:

$\textit{Before-fine profit} = (91-47)\frac{100-91}{3} = 132$.

The cartel is not detected this period, so total profit is 132 this period for all three firms. The cartel and chat are automatically active next period.

Example 4: All three firms vote to form a cartel. A cartel is formed, and the firms discuss prices. Firms 1 and 2 set a price of 55. Firm 3 sets a price of 50. Firm 1 and 2 have before-fine profit equal to 0 as they did not set the lowest price. Firm 3 has before-fine profit equal to:

$\textit{Before-fine profit firm 3} = (50 - 47)\frac{100-50}{1} = 150.$

The cartel is detected this period. Firms 1 and 2 pay no fine as they did not set the lowest price. Their profit for this period is 0.

Firm 3 pays the following fine:

$\textit{Fine} = 50\frac{100-50}{1} = 2500.$

Firm 3's profit for this period is 150 - 2500 = -2350.

The next period starts with a new vote to form a cartel.

Payment

During the experiment you will earn points. 3 points equal 1 eurocent. You will be paid based on all points earned in all four games, plus the 7 euro show-up fee.

In the unlikely event that you will make a loss in the experiment, you will still receive the 7 euro show-up fee. You will be paid privately and in cash at the end of the experiment.

You will now have to answer some questions to show that you understand the instructions. The first game begins when everyone has answered all questions correctly.

Question 1

How many games will you play, and against how many other people will you play?
-1 game, against 2 people
-1 game, against 8 people
-4 games, against the same 2 people each game, in total 2 people
-4 games, against 2 different people each game, in total 8 people

---

Question 2

Do all 4 games have the same number of periods?
-Yes

-No
-We can't be sure, after each period there is another period with 90% chance

---

Question 3

Firm 1 and 3 set price 80, firm 2 sets price 90. What is the before-fine profit of firm 3?

Question 4

Firm 1 and firm 2 vote in favor of a cartel, firm 3 votes against a cartel. Is there a cartel? Can firm 1 and 2 be fined?

-Yes and yes: Firm 1 and firm 2 form a cartel together and can therefore be fined
-No and no: No cartel is established and firms can only be fined when they are in a cartel
-No and yes: No cartel is established but since they voted for a cartel they can be fined

---

Question 5

Each period, there is a 20% chance that active cartels are detected and firms are fined. Does this mean that cartels will be discovered once every 5 periods?

-Yes, a 20% chance means once every 5 periods
-No, there is a 20% chance each period, but there could be many periods without detection

---

Question 6

Firms 1, 2 and 3 agreed in the chat to set a certain price, but all three firms set a lower price. Can the cartel still be detected and fined?

-No, the cartel members did not stick to the agreement
-Yes, once a cartel has been formed it can be detected, regardless of the firms' actions

---

Question 7

You are in a cartel. Which of these prices will lead to the highest fine?

-30
-50
-60
-90

Question 8

Firm 3 is part of a cartel and sets a price of 50. Firms 1 and 2 set a price of 40. Firm 3's before-fine profit is, therefore, 0. The cartel is detected. Does firm 3 need to pay a fine?

-No, because the lowest price is 40

-No, because Firm 3 sells nothing

-Yes, because Firm 3's fine depends on Firm 3's price

Instructions risk preference test

Below you see a table with four columns and multiple rows. For each row, you must make a choice between participating in a risky lottery, where there is a 20% chance of a low outcome and an 80% chance of a high outcome, or not participating, in which case you earn 0 points.

During the experiment you will earn points. 3 points equal 1 euro cent. You will be paid based on the outcome of this lottery choice, your performance in the rest of the experiment, plus a 7 euro show-up fee. You will be paid privately and in cash at the end of the experiment.

You must make a choice for every row, but one row has been randomly selected for payment.

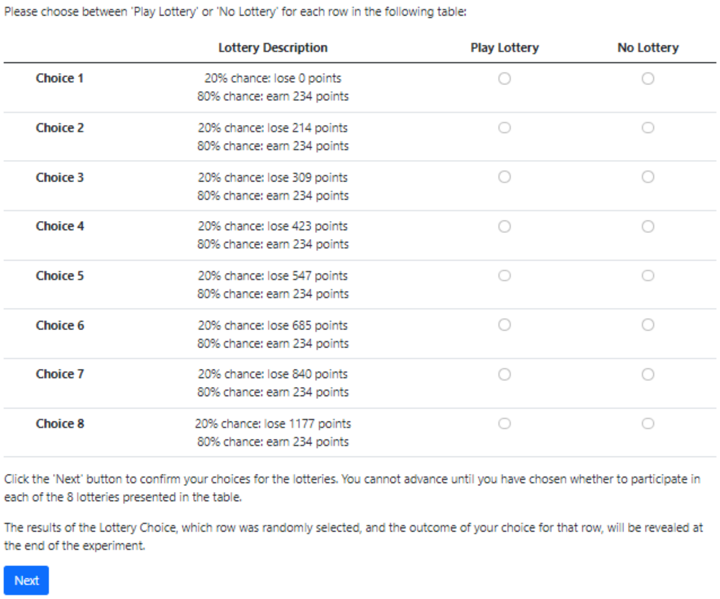
When you go to the next page, all your choices are confirmed. The selected row is revealed at the end of the experiment.
If you chose 'Play Lottery' for the selected row, the lottery is played and you either receive the high or the low outcome.
If you chose 'No Lottery' for the selected row, the lottery is not played, and your payoff will not be affected.

C Risk preference test

Joining a cartel and coordinating prices is risky, as collusion is detectable and punishable in our experiment. We, therefore, measured subjects' risk preferences. Before reading the instructions for and taking part in the repeated Bertrand games, each subject participated in a risk elicitation task based on Holt and Laury (2002), with outcomes chosen to mirror the payoffs in the game that participants would subsequently play. The outcome of this test was communicated to the subjects at the very end of the session, after the conclusion of the Bertrand games.

Each subject needed to indicate for eight lotteries whether she wanted to participate or not. Figure C1 displays the lotteries as seen by the participants, and Appendix B includes the instructions. For each lottery, the chance of ‘winning’ was fixed at 80% (equal to the chance of cartels *not* being detected) and the rewards for winning were 234 points–the single-period before-fine profit of a subject in a cartel that coordinates on the monopoly price. However, the cost of ‘losing’ increased with each lottery. Subjects were paid based on one lottery, drawn randomly before the first session. If a subject had opted to play the randomly chosen lottery, a random draw determined whether any points were added or subtracted to the total earned in the four supergames.

Figure C1: Risk preference test



Please choose between 'Play Lottery' or 'No Lottery' for each row in the following table:

	Lottery Description	Play Lottery	No Lottery
Choice 1	20% chance: lose 0 points 80% chance: earn 234 points	○	○
Choice 2	20% chance: lose 214 points 80% chance: earn 234 points	○	○
Choice 3	20% chance: lose 309 points 80% chance: earn 234 points	○	○
Choice 4	20% chance: lose 423 points 80% chance: earn 234 points	○	○
Choice 5	20% chance: lose 547 points 80% chance: earn 234 points	○	○
Choice 6	20% chance: lose 685 points 80% chance: earn 234 points	○	○
Choice 7	20% chance: lose 840 points 80% chance: earn 234 points	○	○
Choice 8	20% chance: lose 1177 points 80% chance: earn 234 points	○	○

Click the 'Next' button to confirm your choices for the lotteries. You cannot advance until you have chosen whether to participate in each of the 8 lotteries presented in the table.

The results of the Lottery Choice, which row was randomly selected, and the outcome of your choice for that row, will be revealed at the end of the experiment.

We construct as a measure of risk preferences the first row that a subject opts to not play the lottery. This measure ranges from 1 (subject does not play the lottery in row 1) to 9 (subject plays all eight lotteries), with higher values indicating a higher appetite for risk. Table C1 describes this measure by treatment. Our measure of risk preferences is distinctly balanced across treatments. Including this measure as a control variable in regressions that compare outcomes across treatments barely affects point estimates, and leads to at best a modest increase in efficiency. In the main text we, therefore, refrain from such analyses and mainly use information on risk preferences to argue that selection into cartels is similar in the three treatments.

Table C1: Risk preferences, by treatment

	REVENUE	PROFIT	OVERCHARGE
25th percentile	3	3	3
50th percentile	4	4	5
75th percentile	6	6	6
Mean	4.53	4.44	4.63
Standard deviation	1.84	1.99	1.84
Observations	90	99	90

Notes: Descriptive statistics on our measure of risk preferences, number equals last lottery that a subject opted to play.

D Classification of cartel agreements

This Appendix provides a description of how we use the communication data to determine whether cartels coordinate on a particular price. We utilize two definitions of cartel agreements in our analysis. Explicit price agreements are classified purely based on the content of the chat. As discussed in Section 4.3, this definition seems too conservative as it misclassifies both the level and the trend of cartel agreements. Therefore, we construct a broader measure of price agreements that are based on the content of the chat and on the past behavior of the cartel. This is necessary because stable cartels typically reduce communication significantly after successfully coordinating prices, up to the extreme cases where stable cartels at some point require no communication whatsoever but continue coordinating on the previous period's price. We next provide a description of how we construct both measures, followed by chat excerpts that provide examples of implicit agreements or are referenced in the main text.

Explicit price agreements An explicit price agreement to set price p in a given period is said to exist if at least one subject proposes price p, and all other subjects explicitly agree or reaffirm before any subject leaves the chat.

Price agreements A price agreement to set price p in a given period is said to exist if an explicit price agreement is in place, or if an implicit agreement to set price p is in place. An implicit agreement to set price p in a given period is said to exist if i) at least one subject proposes price p, and all other subjects explicitly agree or reaffirm but at least one subject has left the chat before all non-proposers agree or reaffirm, ii) at least one subject suggests to do the same as the previous period without explicitly suggesting a price, and all other subjects explicitly agree or reaffirm, or if iii) at least one subject proposes price p or suggests to do the same as the previous period without explicitly suggesting a price, none of the none-proposers explicitly disagree but at least one non-proposer does not agree or reaffirm, iv) no price is proposed and no suggestions to follow past behavior are made, but coordination on an agreed on price p was achieved in the previous period and none of the subjects voice disagreement with past behavior.

Categories i) to iii) in the definition of price agreements only rely on chat data. Category i) exists because sometimes subjects leave before all subjects have explicitly agreed, so these subjects can not be certain whether an agreement was reached. We construct Category ii) because subjects commonly suggest which price to set based on the previous period (e.g., "same" or "again?"). Category iii) captures cases where some subjects stop responding over time, an extreme case of which is captured by Category iv). We construct this final category because, in stable cartels, subjects occasionally stop communication altogether. However, lack of communication also occurs when subjects cease attempts to coordinate after unsuccessful previous attempts. Therefore, we resort to defining such cases based on past behavior. Table D1 and Table D2 give examples where lack of communication was classified as a price agreement and not as a price agreement, respectively.

Table D1: Lack of communication classified as price agreement

Sg - Period	**Firm 1**	**Firm 2**	**Firm 3**	**Category**
2-1		75?		Explicit price agreement
	85?			
			The max profit is at 73.5	
		73.5 it is then		
			Shall we do that?	
	okay			
	Firm 1 exits chat			
			Cool	
		Firm 2 exits chat		
			Firm 3 exits chat	
2-1	price: *73.5*	price: *73.5*	price: *73.5*	market price: *73.5*
2-2		73.5		Explicit price agreement
	again?			
			nice	
			Yes	
	okay			
			Firm 3 exits chat	
		Firm 2 exits chat		
	Firm 1 exits chat			
2-2	price: *73.5*	price: *73.5*	price: *73.5*	market price: *73.5*
2-3		73.5		Explicit price agreement
	73.5			
			Yep	
		Firm 2 exits chat		
	Firm 1 exits chat			
			Firm 3 exits chat	
2-3	price: *73.5*	price: *73.5*	price: *73.5*	market price: *73.5*
2-4			*Firm 3 exits chat*	Price agreement
		Firm 2 exits chat		
	Firm 1 exits chatt			
2-4	price: *73.5*	price: *73.5*	price: *73.5*	market price: *73.5*

Table D2: Lack of communication not classified as price agreement

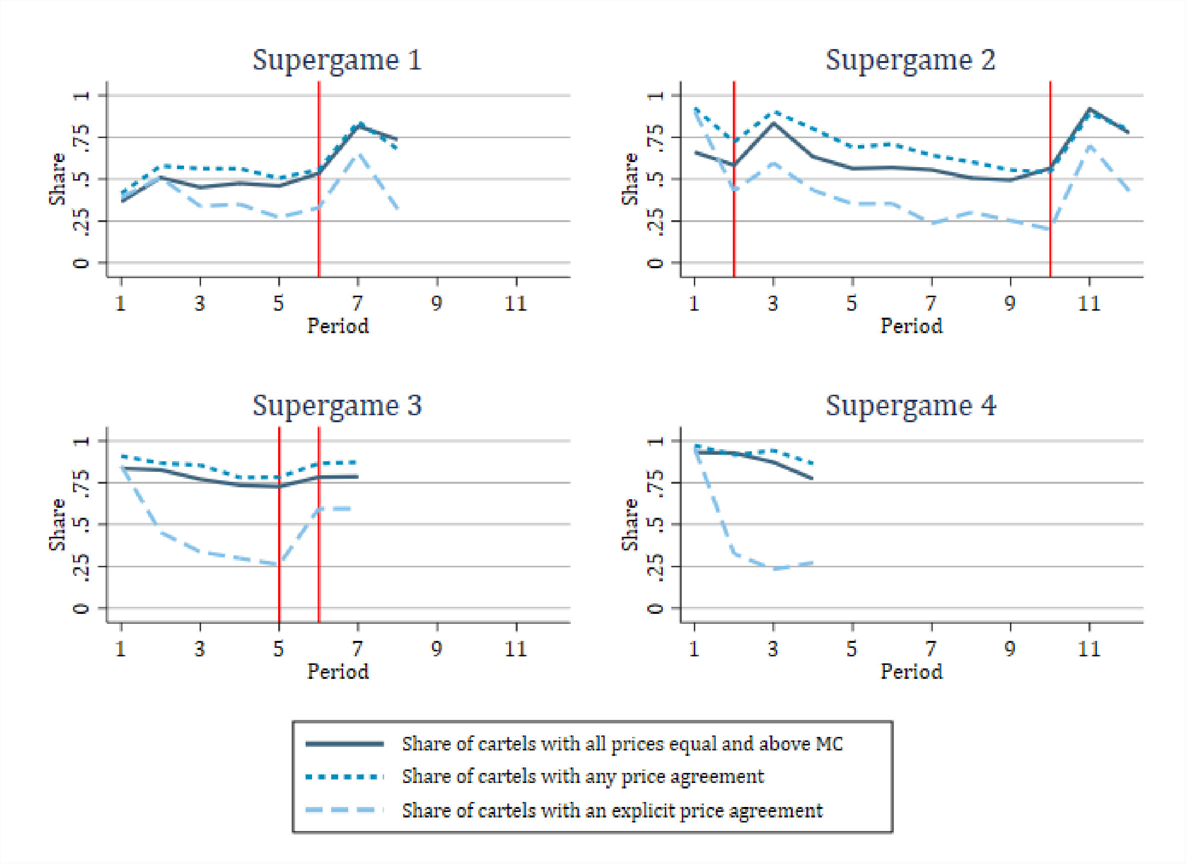
Sg - Period	**Firm 1**	**Firm 2**	**Firm 3**	**Category**
2-7	will we 90 uis bad 75 is bigger win *Firm 1 exits chat*	*Firm 2 exits chat*	90? do 90 *Firm 3 exits chat*	No agreement
2-7	price: *74.9*	price: *88*	price: *90*	market price: *74.9*
2-8	*Firm 1 exits chat*	*Firm 2 exits chat*	*Firm 3 exits chat*	No agreement
2-8	price: *73*	price: *74.9*	price: *69.99*	market price: *69.69*
2-9	*Firm 1 exits chat*	*Firm 2 exits chat*	*Firm 3 exits chat*	No agreement
2-9	price: *75*	price: *66*	price: *59.69*	market price: *59.69*

Table D3: Decreasing price agreement as cartel goes undetected in `OVERCHARGE`

Sg - Period	**Firm 1**	**Firm 2**	**Firm 3**
3-3		last time	
			72 it is
			Firm 3 exits chat
	Firm 1 exits chat		
		72 once more then we all switch to 47	
		Firm 2 exits chat	
3-3	price: *72*	price: *72*	price: *72*
3-4		NOW 47	
			now what
		47	
			yeah lets go
		47 because we are in the 4th round	
	Firm 1 exits chat	*Firm 2 exits chat*	*Firm 3 exits chat*
3-4	price: *47*	price: *47*	price: *47*
3-5		47 again	
	Lets keep 72		
		it's last round	
		we hae profit	
		and 47 allows us to keep it	
			so 72?
		play it safe	
		47 go	
			oka 47
	Firm 1 exits chat	*Firm 2 exits chat*	*Firm 3 exits chat*
3-5	price: *47*	price: *47*	price: *47*

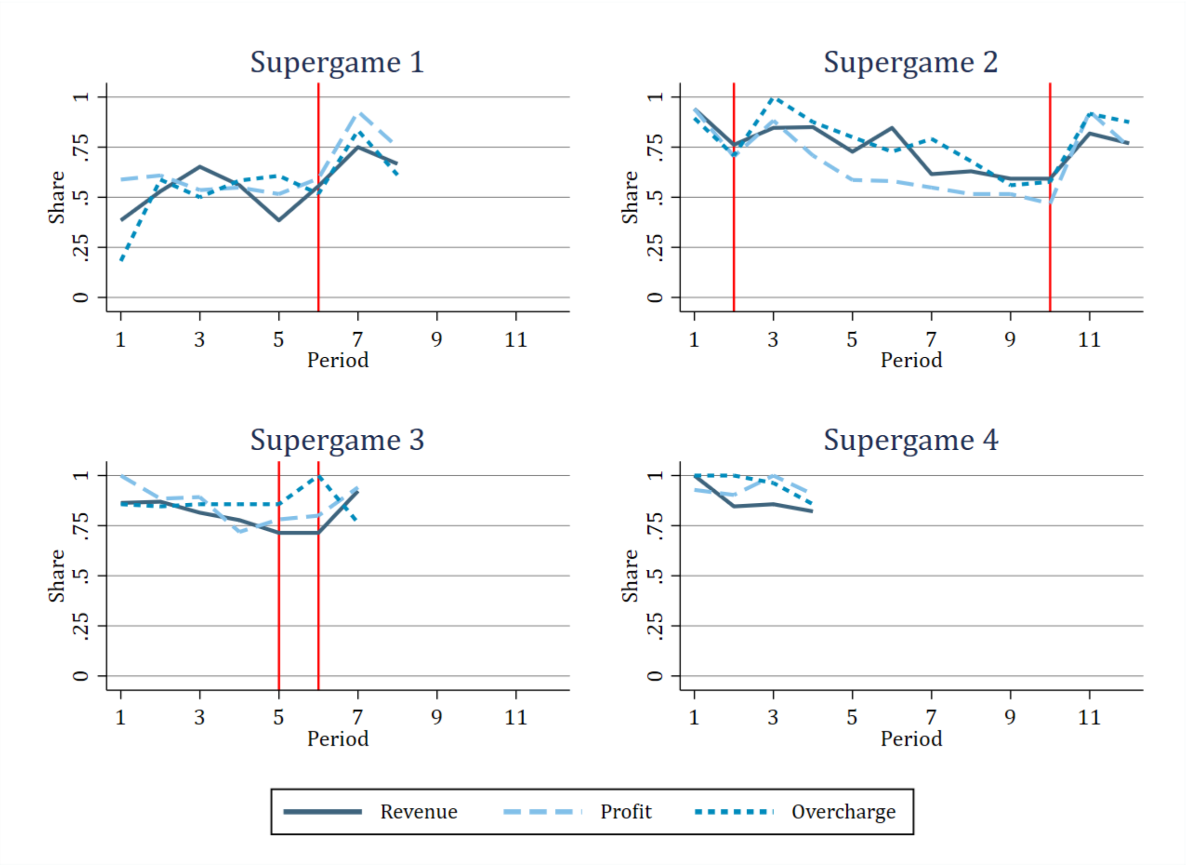
E Additional figures

Figure E1: Coordination and agreement incidence over time, by supergame



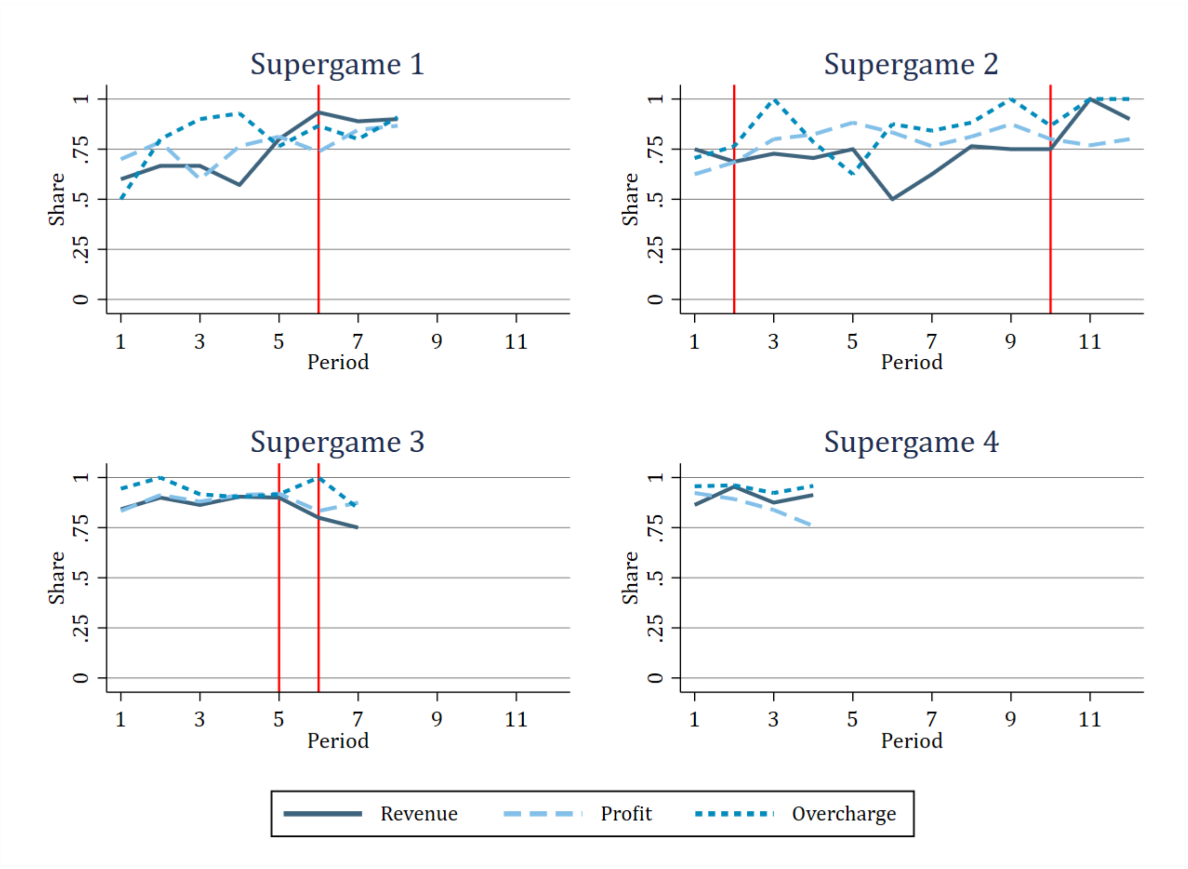
Notes: Share of cartels where all firms set equal prices above marginal cost, share of cartels with an explicit price agreement, and share of cartels with a price agreement in place, over time and by supergame. Red vertical lines indicate a period at the end of which all cartels are detected.

Figure E2: Price agreement incidence over time, by treatment and supergame



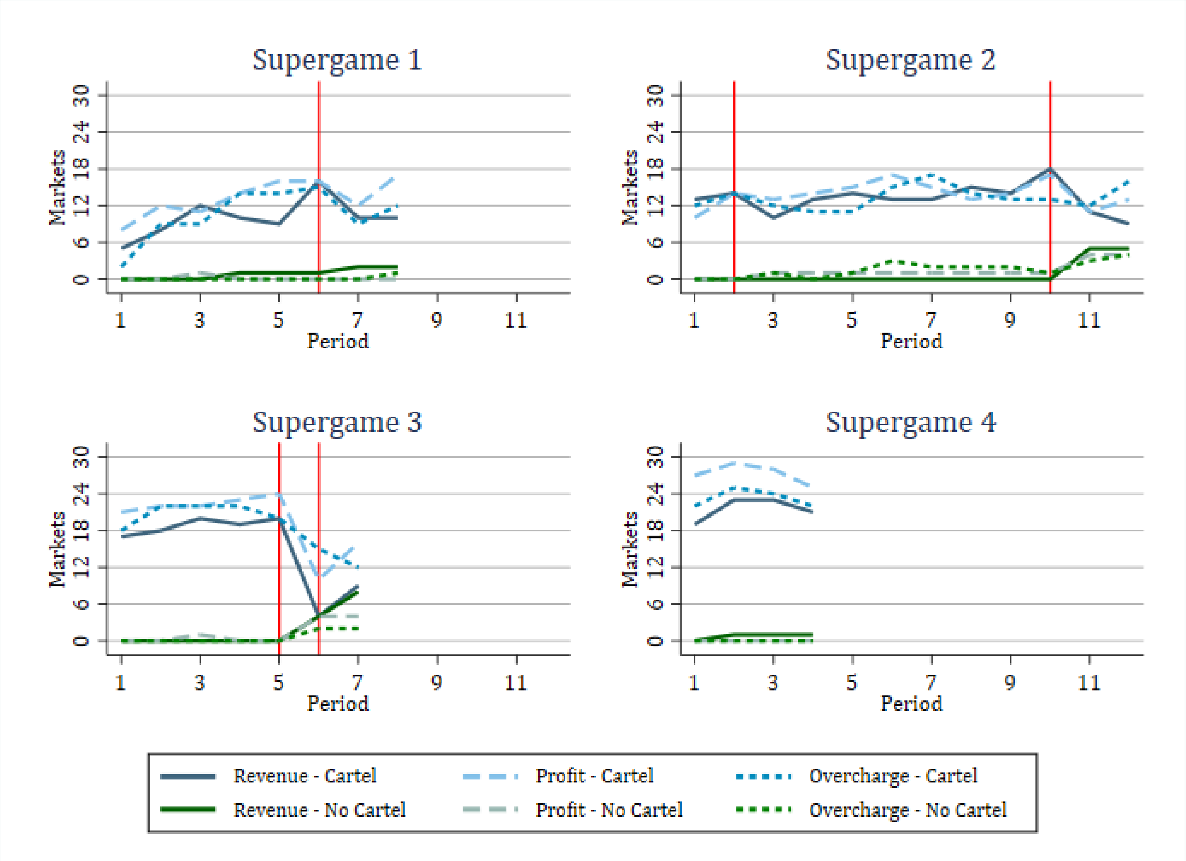
Notes: Average price agreement incidence over time, by treatment and supergame. Price agreement incidence = indicator for a price agreement (explicit or implicit) in a cartelized market-period. Red vertical lines indicate a period at the end of which all cartels are detected.

Figure E3: Cartel stability over time, by treatment and supergame



Notes: Average cartel stability over time, by treatment and supergame. Cartel stability = indicator for a cartel where all subjects set the agreed upon price in a market-period. Red vertical lines indicate a period at the end of which all cartels are detected.

Figure E4: Markets with $p_1 = p_2 = p_3 > 47$ over time, by treatment, cartelization and supergame



Notes: Number of markets where all subjects set the same price (>47) over time, by treatment, supergame, and whether a cartel is active or not. Total number of markets per treatment and supergame are 30 for REVENUE and Overcharge, and 33 for PROFIT . Red vertical lines indicate a period at the end of which all cartels are detected.

F Additional tables

Table F1: Regressions of price agreements on number of consecutive periods with agreement without detection in REVENUE and OVERCHARGE

	REVENUE		OVERCHARGE	
Detectionless periods	0.89	1.53*	-3.47***	-3.41**
	(0.52)	(0.74)	(0.84)	(1.08)
(Detectionless periods)^2	-0.06	-0.13	0.31*	0.30**
	(0.07)	(0.07)	(0.14)	(0.11)
Betrayal last period		-2.08**		-2.32
		(0.92)		(4.11)
Messages sent		0.18*		-0.03
		(0.09)		(0.20)
Reformed cartel		0.16		0.38
		(0.67)		(1.68)
Constant	73.34***	71.48***	74.25***	74.30***
	(0.84)	(1.64)	(1.17)	(3.14)
Observations	392	392	421	421
R^2	0.03	0.06	0.11	0.11
Market Fixed Effects	Yes	Yes	Yes	Yes

Notes: Dependend variable = Price agreement; Detectionless periods = Number of consecutive periods with an agreement leading up to the current period without cartel detection; Betrayal last period = Indicator equal to one if at least one cartel member defected from the price agreement in the previous period; Messages sent = Mean number of messages sent in the chat this cartel-period; Reformed cartel = Indicator equal to one if the current cartel was formed after a previous cartel had been detected; Based on the final three supergames; Standard error clustered at the matching group in parentheses; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table F2: Regressions of price agreements on cartel risk appetite in REVENUE and OVERCHARGE

	REVENUE		OVERCHARGE	
Cartel risk appetite	-0.38	-0.35	2.13*	2.03*
	(0.38)	(0.53)	(1.04)	(0.94)
Cartel irrationality		-0.31		0.85
		(0.94)		(1.71)
Majority male		0.21		-0.99
		(0.90)		(2.01)
Detectionless periods		1.02		-3.70**
		(0.73)		(1.20)
(Detectionless periods)^2		-0.10		0.34**
		(0.09)		(0.13)
Betrayal last period		-1.83*		-0.82
		(0.85)		(3.46)
Messages sent		0.08		-0.01
		(0.10)		(0.14)
Reformed cartel		0.20		-1.28
		(0.55)		(1.63)
Constant	76.76***	74.59***	58.53***	65.15***
	(1.79)	(2.44)	(5.70)	(4.98)
Observations	392	392	421	421
R^2	0.01	0.03	0.06	0.13
Supergame Fixed Effects	No	Yes	No	Yes

Notes: Dependent variable = Price agreement; Cartel risk appetite = Mean (across cartel members) row of the Holt and Laury risk preference test where higher values indicate riskier choices; Cartel irrationality = Indicator that equals one if any cartel subject switched multiple times in the Holt and Laury risk preference test; Majority male = Indicator if at least two of the three cartel members are male; Detectionless periods = Number of consecutive periods with an agreement leading up to the current period without cartel detection; Betrayal last period = Indicator equal to one if at least one cartel member defected from the price agreement in the previous period; Messages sent = Number of messages sent in the chat; Reformed cartel = Indicator equal to one if the current cartel was formed after a previous cartel had been detected; Based on the final three supergames; Standard error clustered at the matching group in parentheses; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table F3: Price agreements and cartel stability using explicit agreements, across treatments

	Agreement incidence	Price agreement	Cartel Stability	Market price
	(Cartels)	(Cartels with an explicit price agreement in place)		
REVENUE	0.38 (0.49)	74.80 (5.84)	0.76 (0.43)	73.53 (6.74)
	$\wedge$	$\vee^{**}$	$\wedge$	$\vee^{*}$
PROFIT	0.36 (0.48)	72.35 (5.63)	0.79 (0.41)	71.52 (6.33)
	$\wedge^{**}$	$\vee^{***}$	$\wedge$	$\vee^{**}$
OVERCHARGE	0.48 (0.50)	68.09 (9.34)	0.88 (0.32)	67.12 (9.66)
	$\vee^{**}$	$\wedge^{***}$	$\vee^{***}$	$\wedge^{***}$
REVENUE	0.38 (0.49)	74.80 (5.84)	0.76 (0.43)	73.53 (6.74)
KW test	$p = 0.042$	$p = 0.000$	$p = 0.051$	$p = 0.002$

Notes: Agreement incidence = Indicator for an explicit price agreement in a market-period; Price agreement = Price that the cartel has explicitly agreed to set in a market-period; Cartel stability = Indicator for whether all three subjects in a cartel have set the agreed-upon price in a market-period; Market price = Lowest submitted price in a market-period; Standard deviation in brackets; Bottom row reports Kruskal-Wallis p-value; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively (MWU test, two-sided).

Table F4: Measures of markets where $p_1 = p_2 = p_3 > 47$, across treatments

	Share	Share	Share	Market price
	(All markets)	(No cartels)	(Cartels)	(Cartels)
REVENUE	0.50 (0.50)	0.12 (0.32)	0.65 (0.48)	75.05 (5.19)
	$\wedge$	$\vee$	$\vee$	$\vee^{***}$
PROFIT	0.53 (0.50)	0.11 (0.32)	0.65 (0.48)	72.52 (5.11)
	$\wedge$	$\vee$	$\wedge$	$\vee^{**}$
OVERCHARGE	0.53 (0.50)	0.09 (0.29)	0.70 (0.46)	68.51 (8.94)
	$\vee$	$\vee$	$\vee$	$\wedge^{***}$
REVENUE	0.50 (0.50)	0.12 (0.32)	0.65 (0.48)	75.05 (5.19)
KW test	$p = 0.823$	$p = 0.933$	$p = 0.869$	$p = 0.000$

Notes: Table F4 compares the share of all, no cartel-, and cartel-periods where $p_1 = p_2 = p_3 > 47$, and market prices of such cartels, across treatments; Standard deviation in brackets; *KW test* = Kruskal-Wallis p-value; ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively (MWU test, two-sided).