

Andres, Maximilian; Bruttel, Lisa; Friedrichsen, Jana

Article — Published Version

How communication makes the difference between a cartel and tacit collusion: A machine learning approach

European Economic Review

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Andres, Maximilian; Bruttel, Lisa; Friedrichsen, Jana (2023) : How communication makes the difference between a cartel and tacit collusion: A machine learning approach, European Economic Review, ISSN 1873-572X, Elsevier, Amsterdam, Vol. 152, pp. --, <https://doi.org/10.1016/j.euroecorev.2022.104331>

This Version is available at:

<https://hdl.handle.net/10419/266561>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>



How communication makes the difference between a cartel and tacit collusion: A machine learning approach[☆]

Maximilian Andres^{a,*}, Lisa Bruttel^a, Jana Friedrichsen^{b,c,d,e}

^a Universität Potsdam, Germany

^b Christian-Albrechts-Universität zu Kiel, Germany

^c Humboldt-Universität zu Berlin, Germany

^d WZB Berlin Social Science Center, Germany

^e DIW Berlin, Germany

ARTICLE INFO

JEL classification:

C92
D43
L41

Keywords:

Cartel
Collusion
Communication
Machine learning
Experiment

ABSTRACT

This paper sheds new light on the role of communication for cartel formation. Using machine learning to evaluate free-form chat communication among firms in a laboratory experiment, we identify typical communication patterns for both explicit cartel formation and indirect attempts to collude tacitly. We document that firms are less likely to communicate explicitly about price fixing and more likely to use indirect messages when sanctioning institutions are present. This effect of sanctions on communication reinforces the direct cartel-detering effect of sanctions as collusion is more difficult to reach and sustain without an explicit agreement. Indirect messages have no, or even a negative, effect on prices.

1. Introduction

Modern competition law generally prohibits firms from joining agreements that target coordinated (pricing) behavior and joint profit maximization.¹ In contrast to this clear prohibition of explicit cartel formation, competition law does not have bite against tacitly collusive behavior, i.e., price coordination without accompanying evidence of agreements between the firms. Thus, firms willing to coordinate their prices have a choice between the formation of an explicit cartel, which comes with the risk of being sanctioned, and tacit collusion, which is not risky with respect to sanctions but may be less effective in terms of coordination.

[☆] We are indebted to Tobias Lettl and Jonathan Beel for providing us with continuous advice from a legal perspective. Further, we thank Joe Harrington and Anja Rösner as well as participants at the annual meeting of the German Association for Experimental Research (GfEW) 2019, the Economic Science Association (ESA) 2020, the workshop “Communication in Markets and Organizations” at Tinbergen Institute Amsterdam, Oligo 2021 at the University of Maastricht, the annual conference of the European Association for Research in Industrial Economics (EARIE) 2021, the annual meeting of the Verein für Socialpolitik (VfS) 2021, and in seminars at the University of Regensburg, the Düsseldorf Institute for Competition Economics (DICE), and the German Institute for Economic Research (DIW Berlin) for helpful comments and suggestions. Luis Koch, Fenja Meinecke, Juri Nithammer, and Max Padubrin provided excellent research support. This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 389447910, which we gratefully acknowledge. Design and hypotheses were preregistered at OSF before the data collection had started (osf.io/s2e3y/).

* Corresponding author.

E-mail addresses: maximilian.andres@uni-potsdam.de (M. Andres), lisa.bruttel@uni-potsdam.de (L. Bruttel), friedrichsen@economics.uni-kiel.de (J. Friedrichsen).

¹ For example, Article 101 of the Treaty on the Functioning of the European Union (2012), prohibits “all agreements between undertakings, decisions by associations of undertakings and concerted practices ... which have as their object or effect the prevention, restriction or distortion of competition...”

<https://doi.org/10.1016/j.eurocorev.2022.104331>

Received 17 June 2021; Received in revised form 29 September 2022; Accepted 23 October 2022

Available online 9 November 2022

0014-2921/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Empirically, the firms' decisions between a cartel and tacit collusion can be inferred from their communication with each other because the difference between the two forms of coordinated behavior is a question of how explicitly firms communicate about coordinating their prices. The variety of communication observed in Harrington et al. (2016), ranging from very indirect signals (such as encoded messages hidden in footnotes, see also Blume and Heidhues, 2008) to highly explicit price communication (see also the examples reported by Harrington, 2006), suggests that firms resolve this tradeoff between criminal liability and effectiveness of communication differently.

Thus far, however, we have only a limited understanding of how sanctioning institutions, communication between firms, and their price setting behavior interact. Existing studies involving communication data either use hand-coded classification into pre-defined categories or rely on anecdotal evidence. In this paper, we shed light on the effect of sanctions on communication as well as the effect of communication content on prices, using a machine learning approach to systematically analyze communication content. In a laboratory experiment, we exogenously manipulate firms' cost to discuss the explicit formation of a cartel by varying whether or not cartel formation is illegal and can be sanctioned. To implement sanctions for illegal agreements in real time during the experiment, our experiment features a participant in the role of the competition authority, who is properly incentivized to judge the communication content and price setting behavior of the firms.²

To quantify differences in communication with and without the threat of sanctions, we organize the content of firms' chat communication using Latent Dirichlet Allocation (LDA, introduced by Blei et al., 2003). This procedure classifies communication into a number of topics covered in the conversation, where both the topics and their number are determined by the algorithm. We then identify the topics that relate to explicit cartel formation and compare how much of the total communication in both treatments can be attributed to these topics. Thus, we can quantify how explicit cartel communication reacts to the presence of sanctioning institutions. Finally, we investigate what share of the total deterring effect of sanctions on price setting behavior is moderated through this variation in the explicitness of communication on prices.

While topics relating to cartel formation are easily identified as those containing the joint profit-maximizing price, there is no obvious indicator for topics covering tacit collusion. Thus, to detect such indirect communication, we rely on a complementary approach: the computation of relative rank differentials. This procedure allows us to compare the most frequent words in the conversation with and without sanctions. We use the words that are relatively more prominent in the treatment with sanctions as an indicator for topics relating to tacit collusion. Then, we test whether markets with a high share of such indirect communication *ceteris paribus* have different prices than other markets.

In our study, we use an experimental approach to study communication between firms. Experiments provide insights into otherwise unobserved aspects of cartels. In particular, we can watch the behavior of undetected cartels and obtain a complete record of the firms' price setting and communication. This constitutes an improvement over the available empirical data on cartels, which provides a biased picture of the universe of cartels and, therefore, also of communication. Specifically, we have some information about legal cartels³ and on illegal cartels that were detected,⁴ but evidence on illegal cartels that remain undetected by the authorities is largely lacking. Economic theory is also not very informative about how firms decide whether the expected profit from forming a cartel is worth the risk of being sanctioned because it typically does not distinguish between an explicit cartel and tacit collusion (Whinston, 2008).⁵

We observe almost perfect adherence to the symmetric joint profit-maximizing price and very explicit communication in the treatment with unrestricted and unsanctioned communication. In contrast, in the presence of sanctioning institutions, fewer markets achieve this coordination and communication is indirect rather than explicit. In particular, our machine learning approach illuminates that firms communicate less often about, or even agree on, specific prices when the competition authority may sanction cartel formation. On the basis of a mediation analysis, we estimate that about one fifth of the total effect of sanctioning institutions on market prices is driven by this inhibiting effect on explicit price communication. Furthermore, we are able to quantify the extent of more indirect communication that tries to initiate a joint price increase without naming it explicitly — for example by discussing contrived excuses for a price increase. Such indirect communication appears to be insufficient to coordinate a joint price increase, but at least it seems to help firms to stabilize prices at their initial level.

With this study, we contribute to an emerging literature using machine learning techniques to evaluate communication in experiments. The LDA topic modeling approach we use is similar to the structural topic model (STM) that Özkes and Hanaki (2020) employ to compare communication among firms. To the best of our knowledge, ours is the first study using LDA to understand how communication affects behavior in experimental markets.⁶ The relative rank differential statistic due to Huerta (2008), which we use to analyze the communication content in different market settings, is also employed in Moellers et al. (2017), Odenkirchen (2018), and Fourberg (2018).

The finding that there is a connection between communication and price levels is consistent with previous studies showing that firms jointly set higher prices in treatments with unrestricted communication than in treatments without communication (see Isaac

² In previous studies, the unanimous decision to communicate fully determined the risk of being fined (e.g. Bigoni et al., 2012), irrespective of the communication content and its effect on prices.

³ In fact, in many countries, cartels were legal during most of the second half of the 20th century. Based on cartel registers that contain information on active and legal cartels and their activities, Hyytinen et al. (2018, 2019) and Fink et al. (2017) investigate how legal cartels in Finland and Austria operated.

⁴ See, e.g., Clark and Houde (2014), Harrington (2006), and Genesove and Mullin (2001).

⁵ For an approach to develop a specific theory of tacit collusion in an auction setting, see Blume and Heidhues (2008).

⁶ In other fields, LDA is used to study, for instance, how transparency affects the deliberation of monetary policy makers (Hansen et al., 2017). The model is also proven useful for the prediction of armed conflicts or economic uncertainty based on newspaper articles (e.g., Rauh, 2019; Mueller and Rauh, 2018). For an overview on the use of text as a data input into economic research see Gentzkow et al. (2019).

et al., 1984; Davis and Holt, 1998; Apesteguia et al., 2007; Cooper and Kühn, 2014; Dijkstra et al., 2021).⁷ In contrast to these previous studies, we keep the availability of communication constant and focus on the effect of sanctioning institutions on the way in which firms communicate. While previous studies modeled tacit collusion as coordinated behavior in the absence of any communication possibility, our design allows for tacit collusion while a communication channel is available, for instance in the form of indirect communication.

The paper proceeds as follows: We describe our experimental design in Section 2 and develop hypotheses in Section 3. We describe our analysis of communication in Section 4 and then present results on how sanctioning institutions affect both, the market outcome and communication among firms, in Section 5. We discuss our results and conclude in Section 6. An appendix complements the paper with the theoretical background (A), additional descriptive statistics (B), additional results (C), illustrative analyses on hand-coded communication data (D), the instructions for firms and authorities (E), details on the text mining results (F), and information on the original German communication content (G).

2. Experimental design and procedures

The experiment features two main treatments, the *SANCTION* treatment, where cartel formation is subject to sanctions, and the *NoSANCTION* treatment without any sanctioning institutions.⁸

General setup. In each session, participants are matched in groups of three participants in *NoSANCTION* and four participants in *SANCTION*. In each group, three participants take the role of firms. In *SANCTION*, the fourth participant takes the role of the competition authority in their group. Role assignments and matching groups remain fixed throughout the repeated interaction described in Figure Fig. 1. Each group represents a market and interacts for at least 25 rounds as described below.

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Chat (60 sec.) - only in rounds 2 to end - chat window opens and closes automatically	Price setting (30 sec.) - self-reporting option available in <i>SANCTION</i>	Feedback 1 (15 sec.) - information about all three prices - self-reporting option available in <i>SANCTION</i> , if the firm has not already reported in stage 2	Investigation (180 sec.) - only in <i>SANCTION</i> , with 10% random investigation probability or after a self-report	Feedback 2 (30 sec.) - own profit (since last investigation excl. and incl. fines in <i>SANCTION</i>) - fine sizes and if a reduction was obtained (for each firm) in <i>SANCTION</i> - recap of all three prices

Fig. 1. Timing of a round in the experiment.

Stage game. In each round, firms simultaneously choose prices in a discrete Bertrand price-setting game with differentiated products.⁹ In this game, a price of three is the Nash equilibrium price and a price of nine is the symmetric joint profit-maximizing price of the stage game. The firms are informed about each others' prices immediately after the price setting stage. Starting from round 2, participants in the role of firms can communicate via free form chat for 60 s before price setting takes place.¹⁰ The chat window opens automatically at the beginning of each round. In *NoSANCTION*, a round is complete with communication, price setting, and feedback.

In *SANCTION*, each round may also contain an investigation by the competition authority. An investigation can take place at random or by a self-report of a firm. The random detection probability is set to 10% in each round and is independent of the firms' behavior.¹¹ Self-reports can be filed to the competition authority during price setting and then again during feedback. Self-reporting is not possible after an investigation has started.

If an investigation takes place, the participant in the role of the authority receives access to the history of chats and prices in their group. He or she judges if, and for how long, a cartel existed and decides about the extent of fines (0%, 50%, or 100%) for

⁷ Relatedly, Fonseca and Normann (2012), Harrington et al. (2016), and Garrod and Olczak (2018) present experimental evidence that explicit cartel formation is most effective in sustaining collusive outcomes when conditions are adverse to tacit collusion, for example because of the market having many firms or the firms being asymmetric in costs or capacities.

⁸ In the treatment *SANCTION*, we further vary whether the first self-reporting firm in a cartel receives amnesty from any potential fine payment (*LENIENCY*) or not (*FINE*). We had already collected data for treatments *FINE* and *LENIENCY* when we started this project. The comparison between these two treatments is the subject of Andres et al. (2021). There we find no effect of a leniency rule, neither on different measures of cartelization nor on communication. Therefore, we pool the data from these two treatments in a joint *SANCTION* treatment for the present paper.

⁹ Our price-setting game and the payoff function for the firms are an adapted three-player version of the setup used by Bigoni et al. (2012). The details are contained in Appendix A.

¹⁰ Communication starts in the second round because we use the price level in the first round as a benchmark for price setting in the absence of communication.

¹¹ This number is consistent with actual cartel detection rates in the European Union between 1985 and 2009 as estimated in Ormosi (2014).

each of the three firms in the respective market. To compute the actual fine, the experimental program takes this percentage value and the cartel duration as an input and applies it to the profits made by the firms' during the rounds that have passed since the last investigation; profits from past rounds are discounted linearly before the fine formula is applied. Participants again receive feedback about the three prices set in their market in the current round, their own profit, and – if applicable – about reporting decisions and realized fine payments.

Repetition. Participants repeat the previously described interaction for a minimum of 25 rounds. From then on, the game ends with a probability of 1/3 after any round; with the complementary probability of 2/3 the game continues for another round. The expected duration of the interaction is 27 rounds. The random termination rule serves the purpose to blur the time horizon to minimize endgame effects.

Instructions and training. Participants were informed about the relationship between their own and the other two firms' prices and their own profit by means of a profit table (cf. Appendix E). The instructions also provide a verbal description of the qualitative impact of own and others' prices on profits. To ensure that participants in the role of firms understand the relatively complex market interaction, they were given access to a computerized training tool before the start of the experiment. In the tool, they could enter their own price and two prices for their competitors and receive feedback on the resulting profits for as many price combinations as they desired.

Participants in the role of an authority received an information sheet explaining in detail when firm behavior is to be considered in violation of competition law as well as how the duration and severity of the infringement are determined.¹² Further, participants in the role of a competition authority interacted with a training tool before the start of the experiment. In the tool, they had to judge three archetypical market constellations in exactly the way they had to during the actual experiment. Participants then received feedback and an explanation for the correct judgments. The experiment only started after everyone had finished their use of the respective training tool.

Payment. Participants in the role of firms were paid their cumulative earnings from the entire interaction, using an exchange rate of 1 Euro = 125 points. Stage payoffs are not discounted. Perfectly competitive behavior according to playing the Nash equilibrium of the stage game across all rounds yields an expected 2700 points and a symmetric joint profit-maximizing cartel subject to the risk of being detected and fined yields 4860 in expectation.

Participants in the role of the competition authority were paid based on the overlap of their judgment with the judgment of an expert in competition law, who we contracted with to independently evaluate the chat messages and price setting behavior of the firms. In each investigation, the competition authority makes four decisions (size of the fine for firms 1, 2, and 3 (0%, 50%, 100% of the relevant profit) as well as the duration of the cartel in rounds). We use a binary scoring rule to evaluate decisions. For each agreement with the expert, a participant in the role of the competition authority receives 900 points meaning that, in each investigation, he or she can make up to 3600 points. In order to compare the duration stated by the experimental competition authority with the expert's round-wise judgment, we computed the sum of rounds since the last investigation in which the expert reported a cartel. Similarly, we computed the average of the expert's judgment of cartel activity by each firm over that interval and counted the decision of the experimental competition authority as correct if the expert's average judgment comes closer to this judgment than to the other two (0%, 50%, or 100%) categories. Authorities were paid the average number of points achieved per investigation, using the same exchange rate of 1 Euro = 125 points. In case no investigation ever took place in his or her group, the respective authority received a payoff of 15 Euros.

Participants in the role of a firm received their payoff from the experiment and a show-up fee of 5 euros immediately after the experiment in cash. Participants in the role of the competition authority received a show-up fee of 10 euros immediately after the experiment in cash and were paid their payoff from the investigations 2–3 weeks after the experiment by bank transfer.

Expert judgment of cartelization. The expert holds a law degree (German: "Volljurist"), was writing a dissertation in the field of competition law at the time of the experiment, and also has practical experience in this area. After each session, the expert received the full chat protocols as well as the history of prices of all firms in all rounds. A few days later, he provided us with a round-wise classification of whether a firm participated in a cartel split up by the same levels that were available to the authority in the experiment (0%, 50%, or 100%). The expert judged firms based on their communication and price-setting as this would be done in an actual investigation. This implies that a cartel might cease to exist without an investigation that dismantled it but a cartel could also continue to exist after having been investigated and fined. The expert judgment gives us a per-period assessment of market conduct that we use to analyze the extent of cartelization.

Procedures. The experiment was programmed in z-Tree (Fischbacher, 2007). We collected our data with a total of 269 participants at the experimental laboratories at the University of Potsdam and at TU Berlin in February to July 2019. The participants were invited for the sessions through the regular invitation procedures of the respective laboratories using ORSEE (Greiner, 2015). Assignment to the different treatments was random in the sense that subjects signing up for a session did not know which treatment would be run. Our sample contains 23 independent markets in NoSANCTION, and 50 in SANCTION (split up into 23 in FINE and 27 in LENIENCY). All treatments were balanced across the two involved laboratories in Potsdam and Berlin. On average each participant earned 36.73 Euro. The experiment was planned to last for a maximum of 2.5 h. If the random continuation mechanism had not stopped the experiment during this time span, we would have manually stopped the experiment at this point in time. Participants were informed about this rule in the instructions. This event was unlikely and did not occur.

¹² We intentionally did not provide this information to participants in the role of firms because we wanted to mimic real conditions in which most firms (except very large ones with their own legal department) are not aware of the precise legal situation.

3. Hypotheses

The innovation of our approach lies in allowing for free form communication and analyzing the content of communication using machine learning. Before turning to communication, however, we introduce two hypotheses regarding the direct economic effect of sanctioning institutions on the main economic variables in our setting – the extent of cartelization and average prices – which are crucial in assessing the effectiveness of sanctions.

Due to the risk of being fined, the incentive compatibility constraint for the symmetric collusive equilibrium is tighter and the critical discount factor higher with sanctioning institutions than without (see Appendix A). In fact, the critical discount factor of an infinitely repeated discounted game with punishment by Nash reversion is below the continuation probability of $2/3$ in the NoSANCTION treatment and above it in the SANCTION treatment, both with and without leniency, implying that in the abstract game without communication, perfect, symmetric collusion is an equilibrium only in the absence of sanctions. Therefore, we expect the extent of cartelization, as measured by the expert judgment of the firms' behavior in each round, to be higher in NoSANCTION than in SANCTION. Further, we expect that average market prices move in parallel to cartelization rates because prices are lower in the absence of a cartel due to competitive price effects than in cartels that fix prices. With less cartelization as a response to the risk of being fined, averaged across all rounds, prices will then be lower in the presence of either type of sanctioning institution than without. Further, collusion at a lower price relaxes the incentive compatibility constraint for collusion, which is more relevant in SANCTION treatments, where the critical discount factor at a price of nine exceeds the continuation probability (see Appendix A).

Hypothesis 1. The extent of cartelization in rounds 2–25 is higher in NoSANCTION than in SANCTION.

Hypothesis 2. Average prices in rounds 2–25 are higher in NoSANCTION than in SANCTION.

Our next hypothesis posits that the communication content exhibits treatment differences in line with those in cartelization rates and prices. We expect that sanctioning institutions make participants more careful in their statements because they will try to avoid punishment for explicit price coordination. Specifically, we expect fewer statements referring explicitly to setting specific supra-competitive prices and, in particular, to the joint profit maximizing price of 9 in the treatment with sanctioning institutions than in the one without.

Hypothesis 3. Communication in NoSANCTION is more explicit about prices and about jointly maximizing profits than communication in SANCTION.

Finally, we also investigate to what extent more explicit communication *causally* drives higher cartelization rates and prices. We expect that explicit communication is more effective in coordinating and raising prices than less explicit statements, irrespective of the treatment condition.

Hypothesis 4. Prices are higher and there is more cartelization with more explicit communication than when communication is less explicit.

Note that the level of explicit communication and average cartelization as judged by the expert are necessarily related because the expert judgment underlying the cartelization measure relies in part on the content of the firms' communication. However, they are not the same: explicit communication followed by non-cartel prices (e.g. when all firms try to exploit each other) is not classified as cartelization by the expert. Non-explicit communication, vice versa, might nevertheless be classified as cartelization by the expert if it was sufficient to initiate a joint price increase of the firms. Thus, the average price is the more appropriate measure than the cartelization rate here. For details on how to judge firm behavior, we refer to the instructions to the authorities that were developed in cooperation with the expert (see Appendix E.2).

4. Evaluating communication

Before testing our hypotheses, let us first explain how we analyze our communication data. Ultimately, we are interested in the role communication plays for cartel formation. This analysis goes far beyond the classification of whether a specific group in the experiment formed a cartel or not because it aims at understanding the patterns of communication. While the judgment whether a cartel exists or not is done by humans both in the real world (by judges at a court) and in our experiment (by the experimental competition authority and the expert), a deeper understanding of communication patterns and a formal test of the related hypotheses require a comprehensive text analysis to map the recorded open chat communication from our experiment into quantified data about the topics discussed in the chat.

Quantifying communication data is a challenging task receiving attention in a variety of disciplines, including economics. The reliance on human raters to hand-code text is the most commonly used approach in the field of experimental economics. In these studies, categories are defined first, either based on an in-depth-analysis of parts of the data (e.g. Cooper and Kagel, 2005), using external experts (e.g. Coffman and Niehaus, 2015), or on the basis of coordination games (e.g. Houser and Xiao, 2011). Then, the entire data set is coded into these categories either by human raters or – less often – using supervised machine learning techniques as in Penczynski (2019).

As they rely on predefined categories, these approaches may be subject to biases introduced in the definition of categories. Therefore, we use an unsupervised machine learning algorithm that does not rely on any pre-classification of text by the researchers

(or others who are contracted by the researchers). This unsupervised machine learning algorithm is fed with unclassified text data and uncovers hidden patterns in the form of meaningful word groupings that form the topics of communication.¹³

4.1. Text corpus

The starting point for our analysis is the entire chat communication from our experimental sessions. We take each group chat, i.e., all messages sent in a specific group throughout rounds 2 to 25, as a separate document. Thus, we have 73 separate documents, which together form the corpus for the analysis. As a first step, we process the text data in the corpus by (1) correcting spelling mistakes, (2) eliminating ‘stopwords’, i.e. words that appear frequently in all texts but have no meaningful content,¹⁴ and (3) reducing the remaining words to their linguistic roots (Hansen et al., 2017).¹⁵ The processed corpus of the communication data consists of 19888 tokens in total and contains 3547 unique tokens. In most cases, such tokens are equivalent to words in the documents, but a token can also be, e.g., a number.¹⁶

At an abstract level, this corpus of communication data can be represented in a 73 by 3547 document–term matrix, where the element (d, v) of the matrix gives the number of times that the v th unique token appeared in d th group chat. This matrix representation has a high sparsity of 96 percent so that it is key to reduce the dimensionality of the data for further analysis.¹⁷

4.2. LDA model

Intuitively speaking, the LDA procedure assumes that the content of each text document is a collection of tokens. The LDA assumes further that each document is a mixture of topics and that topics are characterized by a distribution of tokens. More technically speaking, the LDA uses Dirichlet priors for the distributions of tokens over topics and for the distribution of topics over documents.¹⁸ Then, it uses the observed distribution of tokens over documents and a Gibbs Sampling procedure¹⁹ to generate posterior distributions of tokens over topics and of topics over documents (see Blei et al., 2003; Griffiths and Steyvers, 2004; Hansen et al., 2017).²⁰ These posterior probabilities are what we use in our analysis.

A challenge for any LDA lies in choosing the dimensionality of the latent space, in our case the number of topics K . We rely on the ‘perplexity score’ from cross-validation as a goodness-of-fit measure to determine the appropriate number of topics (Newman et al., 2009).²¹ Fig. 2 illustrates this score for up to 100 topics. The solid line in Fig. 2 depicts the average of a 5-fold cross-validation of the model, where 80% of the data are used to train a model that predicts the remaining 20%, in a round-robin sequence. Lower values of the perplexity score indicate a better fit in out-of-sample prediction. If we choose too few topics, the estimated topics will mix underlying content, which will result in a poor model fit, corresponding to a high perplexity score. As the number of topics increases, the perplexity score decreases because finer grained topics better approximate the true data. But if we choose too many topics, they might become very specific to a particular group and be more difficult to interpret (Chang et al., 2009; Hansen et al., 2017). The statistically optimal number of topics lies at the point where adding one more topic does not reduce the perplexity sufficiently further. In Fig. 2, this corresponds to the point of 25 topics, where the solid line starts bending toward the horizontal. The resulting number of topics that we use for modeling the topics of the chat communication is $K = 25$.

4.3. Estimated topics and explicit communication

Next, we let the LDA estimate the posterior distributions of tokens over topics for $K = 25$.²² Thereby, each topic corresponds to a probability vector over the 3547 unique tokens from the processed corpus telling us how likely it is that a specific token is used

¹³ Brandts et al. (2019) provide an overview of laboratory experiments with communication. Özkes and Hanaki (2020) discuss the different methods for making sense of chat data. Their study is also the only one we are aware of that uses an unsupervised algorithm in an experiment.

¹⁴ English examples are ‘the’ or ‘at’. We added tokens typical for chat messages in German to the list of stopwords provided by Feinerer et al. (2008) such as ‘wat’ meaning ‘what’ in Berlin and Brandenburg.

¹⁵ This procedure is called ‘stemming’. For example, ‘preference’ and ‘prefers’ becomes ‘prefer’. To stem words, we use the standard R package *SnowballC* published by Bouchet-Valat (2019).

¹⁶ Figure 14 in Appendix F shows the token frequency per treatment.

¹⁷ Compared to previous studies, our document–term matrix is not that sparse. We attribute the “low” sparsity to the fact that we have a homogeneous group of participants facing the same controlled experimental situation such that it is likely that their vocabulary is very similar.

¹⁸ The Dirichlet priors assign probabilities to tokens over topics in such a way that in each topic few tokens occur with high probability and many other tokens occur with low probability. For the topic-per-document distribution, the Dirichlet prior similarly assigns probabilities such that in each document few topics occur with high probability and many other topics occur with low probability. Such distributions are very typical for all kinds of text data (Griffiths and Steyvers, 2007).

¹⁹ Gibbs Sampling is a form of Markov chain Monte Carlo to obtain sampled values that approximate a target distribution. The method is used when direct sampling is difficult. Broadly speaking, Gibbs Sampling starts with a random token–topic assignment. Then, it picks each token and estimates the probabilities that this token belongs to each topic *conditioning on all other current token–topic assignments*. The resulting new token–topic assignments are the starting point for the next “round” of the estimation procedure (see Griffiths and Steyvers, 2007; Hornik and Grün, 2011).

²⁰ We adopt the LDA implemented in the R package *topicmodels* by Hornik and Grün (2011) and use the suggested values from Griffiths and Steyvers (2004) for the parametrization of the model.

²¹ The perplexity score is computed as the geometric mean per-word likelihood, a standard measure in the machine learning literature.

²² We also ran the LDA with fewer topics. Doing so did not improve the interpretability of topics, which would have justified a deviation from the statistically optimal number of topics according to Blei (2012).

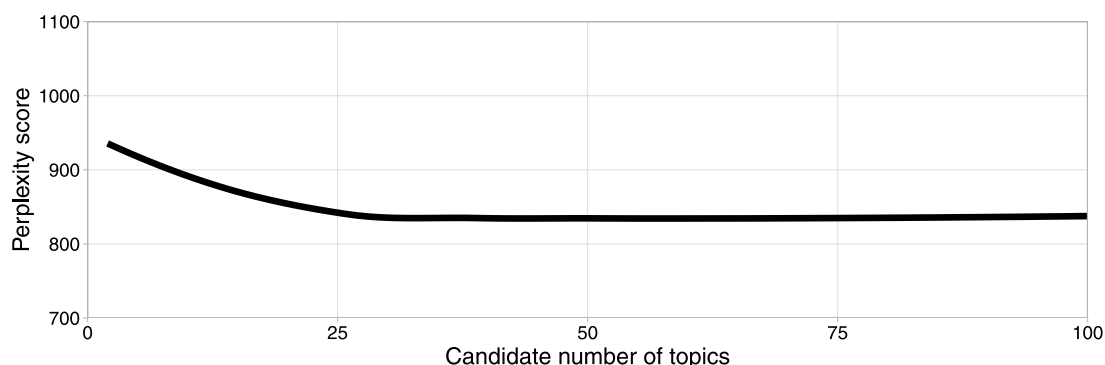


Fig. 2. Perplexity score when fitting the trained model to the hold-out set.

in a given topic.²³ The LDA also provides us with a representation of how much of the communication in a given group chat can be attributed to each of the inferred 25 topics. The representation comes in form of the estimated posterior distribution of the 25 topics over the 73 documents or group chats.²⁴

As Hypotheses 3 and 4 refer to the use of explicit communication about collusive practices, we screen the estimated topics for evidence of such explicitness and concentrate on those topics for all further steps of our analysis. Based on the pre-registration of this study, we define a topic as evidence of explicit cartel formation if the joint profit maximizing price of nine (or ‘9’) appears in the top ten list of tokens of the respective topic. Following this definition, two out of the 25 topics are identified as referring explicitly to cartel formation. Fig. 3 summarizes key information for the two explicit topics.²⁵

In topic 3, depicted in Fig. 3, the joint-profit maximizing price of ‘9’ is ranked fifth and grouped together with several other prices (‘12’, ‘7’, ‘8’, ‘10’) surrounding the symmetric collusive price or relating to an asymmetric collusive equilibrium (cf. Appendix A.2.3) and with the number 2, which probably relates to the suggestion of raising the price by 2.²⁶ These tokens belong to explicit price-fixing agreements and yield supracompetitive profits. Further, this topic contains a strong notion of agreement (‘okay’, ‘yes’). Therefore, we label this topic *Explicit Agreement*. Topic 18, depicted in Fig. 3, consists of a group of tokens related to setting the joint-profit maximizing price (‘9’), to obtain higher earnings (‘get’, ‘remain’, ‘more’, ‘euro’) from the duration of the experiment (‘round’, ‘25’, ‘hour’), and some notion of understanding (‘exact’). Thus, we label this topic *Explicit Reasoning*.

5. Effects of sanctioning institutions

In this section, we first analyze how the presence of sanctioning institutions affects market outcomes, specifically the cartelization rate and average market prices. We then continue to investigate the differences in communication depending on the presence of sanctioning institutions. Finally, we study whether there is a causal link from the extent of explicit and indirect communication to anticompetitive market outcomes.

For the following analysis, we restrict ourselves to the data from rounds 2 to 25. We use this restriction because these rounds are played in all sessions and thus allow for the cleanest treatment comparison. From round 25 onward, the game ends with a probability of 33% after each round, so that the number of rounds played after 25 rounds differs across markets. In the first round, there was no communication stage.

In Andres et al. (2021), we study the effect of a leniency program on cartelization and prices. As we do not find significant differences in cartelization or average market prices between FINE and LENIENCY there, we pool the data from these two treatments under the joint name SANCTION when comparing market outcomes to the NoSANCTION treatment in Section 5.1.

Our sample contains 23 independent markets in NoSANCTION, and 50 in SANCTION. In SANCTION, we observe that an investigation takes place in 11% of rounds on average and in 56% of all investigations conducted, a strictly positive fine is imposed on the investigated firms. By definition, investigations and fines do not occur in NoSANCTION. Table 1 provides a first summary of how sanctions affect market outcomes and communication. The table suggests that sanctions generally have the desired effect: cartelization rates and prices decrease, and communication shifts from explicit cartel talk to indirect collusive attempts. Table 3 in Appendix B provides more detailed summary statistics. If nothing else is stated, all p-values reported in this paper refer to the results of a two-sided Wilcoxon–Mann–Whitney test with continuity correction.

²³ Figure 15 in Appendix F shows the estimated token distributions for all 25 topics.

²⁴ Figure 16 in Appendix F illustrates the estimated distributions of topics separately for each treatment.

²⁵ The algorithm numbered these two topics as topics 3 and 18. In order to facilitate the comparison with Figures 15 and 16 in Appendix F, we maintain this numbering here.

²⁶ In the absence of collusion, markets are typically not fully competitive but many markets have prices of around 6 to 7 in early rounds so that raising the price by 2 would get the market close to the symmetric collusive outcome.

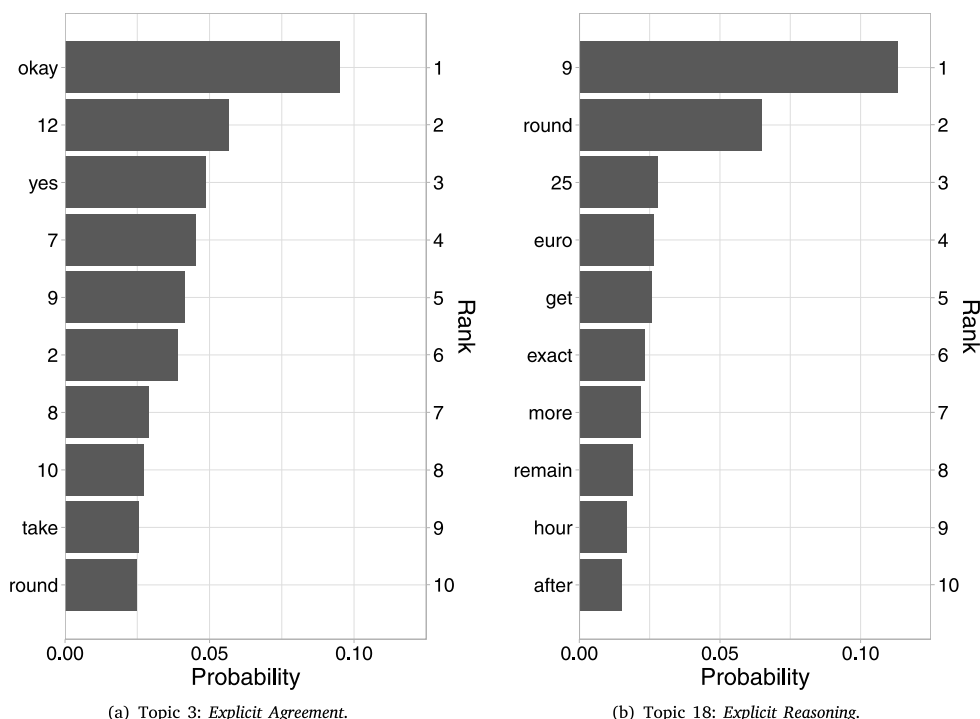


Fig. 3. Top ten token probabilities per explicit topic. The rank of a given token within the topic is given on the y-axis, the estimated probability of a token within the topic is given by the length of the bar on the x-axis.

Table 1

Average cartelization rate, market price, explicit and indirect communication split up by treatment. Standard deviations in parentheses.

	Cartelization	Market price	Communication	
			Explicit	Indirect
NoSanction	0.95 (0.11)	8.84 (0.45)	0.32 (0.14)	0.04 (0.02)
Sanction	0.33 (0.29)	6.64 (1.26)	0.09 (0.06)	0.10 (0.09)
Treatment difference	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

5.1. Market outcomes

Cartelization. To examine how the risk of sanctions affects the extent of cartelization in an average market, we compare the ratio of rounds in which a cartel existed across treatments. Our measure for the extent of cartelization is based on the roundwise judgment of the expert. As the expert classified individual firms' behavior per round into three categories (0%, 50%, or 100%), we can build two different measures of cartelization at the market level. First, we compute the extent of cartelization per market by taking into account the level of punishment when averaging the roundwise judgments to a market-level measure. In other words, we directly average over the three categories of judgment (0%, 50%, or 100%) per round per firm to arrive at the extent of cartelization in a market. This measure accounts for the fact that anticompetitive behavior may be more or less severe and provides a precise measure of cartelization. We, therefore, use it as our primary measure. Less severe infringements, however, constitute also illegal cartels, which speaks for a binary measure where anticompetitive behavior, irrespective of the severity of an infringement (50% or 100%), is treated as a cartel. Therefore, as a secondary measure for the extent of cartelization, we also report treatment comparisons based on the unweighted extent of cartelization, treating cartels of the 50% category and the 100% category equally. Again, the measure that we use for our analysis is at the market level as we average over all binary judgments within a market²⁷

²⁷ If we consider the binary judgment of whether or not a firm participated in a cartel, participants in the role of the competition authority come to the same judgment as the expert in 76.49% of the cases. If we consider the weighted judgment, which takes into account the severity of an infringement and the duration of a cartel, the overlap between participant and expert judgment still amounts to 61.05%. Most importantly for us, the *difference* between the judgment

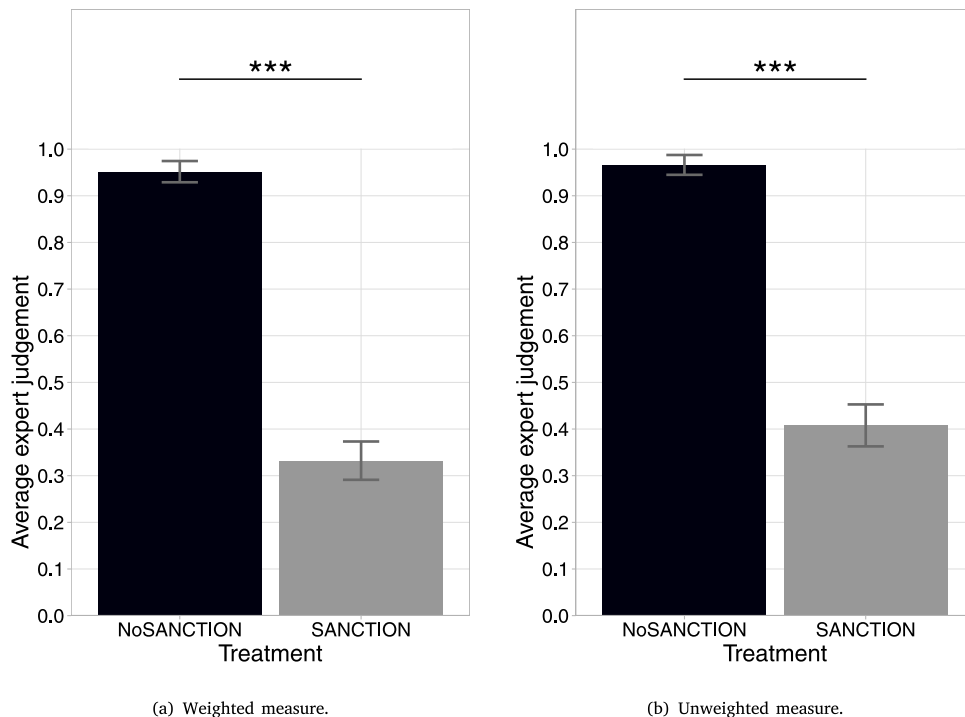


Fig. 4. Cartelization according to the expert judgment split up by treatment; averages are taken over group averages of the weighted expert judgment that reflects the severity of the cartel for Panel (a) and the unweighted expert judgment that reflects whether a cartel is present or not for Panel (b). ‘***’ refers to $p \leq 0.01$. Error bars indicate standard errors.

We find that the average weighted cartelization rate is 0.95 in NoSANCTION ($N = 23$, $SD = 0.11$) and 0.33 in SANCTION ($N = 50$, $SD = 0.29$). The difference is statistically significant ($p < 0.001$). The result is very similar if we instead consider the unweighted expert judgment. In this case, we observe on average a cartelization rate of 0.97 in NoSANCTION ($N = 23$, $SD = 0.1$) versus 0.41 in SANCTION ($N = 50$, $SD = 0.32$). The difference is again statistically significant ($p < 0.001$). Fig. 4 illustrates this finding. Thus, our data clearly supports Hypothesis 1 that sanctioning institutions reduce cartelization.

Prices. We average prices per market over time in rounds 2 to 25 and then test whether the average market prices differ between the two treatments. Prices are substantially higher in NoSANCTION, with an average price of 8.84 points ($N = 23$, $SD = 0.45$), than in SANCTION, where we observe an average price of 6.64 points ($N = 50$, $SD = 1.26$). The difference is statistically significant ($p < 0.001$). Hence, our data supports Hypothesis 2 that prices are higher in NoSANCTION than in SANCTION. Fig. 5 illustrates that this difference also persists at the level of the individual round and does not change over time.

Even if we restrict attention to cartel phases (taking 50% and 100% cartels together), sanctioning institutions have a significantly negative effect on prices. The average cartel price of 8.96 points in NoSANCTION is significantly higher than the average cartel price of 7.82 points in SANCTION (NoSANCTION: $N = 23$, $SD = 0.24$; SANCTION: $N = 40$, $SD = 1.05$; $p < 0.001$). Thus, even conditional on firms engaging in anticompetitive behavior, the infringements are less harmful to consumer surplus in the presence of sanctioning institutions.

Next, we consider the price setting in competitive phases. The average competitive price of 5.46 points in NoSANCTION is not significantly different from the average competitive price of 5.71 points in SANCTION (NoSANCTION: $N = 5$, $SD = 0.88$; SANCTION: $N = 50$, $SD = 1.15$; $p = 0.79$). Thus, we find no evidence that the presence of sanctioning institutions affects the price setting of firms when they are not colluding.

Also if we take a closer look at the full distribution of expert judgments, i.e. distinguishing between 0%, 50%, and 100% entries, we find that average prices correlate with the expert judgment as expected. Table 2 contains the fraction of expert’s judgments at the firm level as well as average prices for each category per market across treatments. While in NoSANCTION the vast majority of firms participate in a cartel most of the time and obtain a judgment 100%, in SANCTION most entries correspond to not participating in a cartel (0%) and only a quarter of firm-period observations are classified as full collusion (100%). In line with the above analysis

of the participant and the one of the expert is not systematically different in the two treatments, neither with the former ($p = 0.75$) nor with the latter measure ($p = 0.31$) in a two-tailed Wilcoxon–Mann–Whitney test.

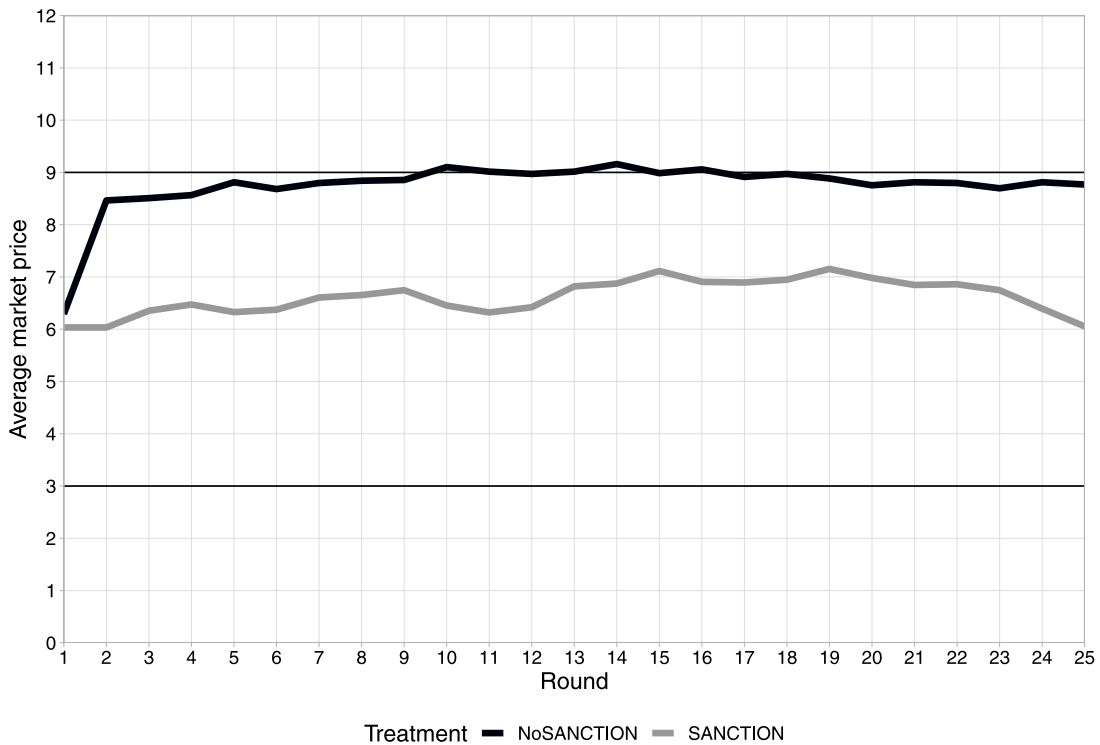


Fig. 5. Average market prices over time split up by treatment. Nine is the joint profit maximizing price and three the Nash equilibrium price of the stage game.

Table 2

Fraction of markets that are categorized as 0%, 50%, or 100% by the expert and average prices within each category per market split up by treatment. Standard deviations in parentheses.

		Expert's judgment	0%	50%	100%
NoSANCTION	Fraction		0.03	0.03	0.94
	Average price		5.46 (0.88)	8.20 (0.84)	8.98 (0.28)
SANCTION	Fraction		0.59	0.15	0.26
	Average price		5.71 (1.15)	7.28 (0.95)	8.63 (0.92)
Treatment difference	Average price		$p = 0.79$	$p = 0.02$	$p = 0.01$

of cartel vs. non-cartel phases, Table 2 shows that average prices in cartel phases are lower in markets with sanctions than in those without, both for moderate (50%) and severe (100%) infringements, but there are no significant treatment differences in prices outside of cartel phases (0%).

5.2. Sanctioning institutions and communication

We now analyze the effect of sanctioning institutions on the extent of communication that is explicit about forming a cartel, using the classification of the chat data and the definition of explicit communication from Section 4. To test whether there are differences in explicit cartel agreements during the communication, we compare the average posterior probabilities of the topics *Explicit Reasoning* and *Explicit Agreement* across treatments. Fig. 6 illustrates these posteriors which tell us with which probability the firms' communication falls into these two topics according to the output from the LDA. It can be seen that the extent of explicit communication is far greater in NoSANCTION than in the treatments with sanctioning institutions.²⁸

²⁸ Note that the relatively small numbers are standard for communication data. Common estimates suggest that about 75% of all communication does not relate to the main theme of the conversation (see Dunbar, 1998).

Indeed, formal tests on our data fully support [Hypothesis 3](#) that communication in NoSANCTION is more explicit about prices than communication in SANCTION.²⁹ For the tests, we compare average posterior probabilities of the explicit topics, i.e. the shares of the chat protocols that can be attributed to one of the explicit topics. The average posterior probability of the topic *Explicit Agreement* is 0.18 in NoSANCTION ($N = 23$, $SD = 0.16$) and, thereby, significantly higher than the probability of only 0.05 in SANCTION ($N = 50$, $SD = 0.05$; $p < 0.001$). Similarly, the average posterior probability of the topic *Explicit Reasoning* is 0.15 in NoSANCTION ($N = 23$, $SD = 0.09$) and 0.04 in SANCTION ($N = 50$, $SD = 0.03$); again, these probabilities differ significantly from each other ($p < 0.001$).

When we also consider the total amount of explicit communication by summing up the average posterior probabilities of *Explicit Reasoning* and *Explicit Agreement*, we find that the average posterior probability of such explicit communication is significantly higher in NoSANCTION, with an average of 0.32 ($N = 23$, $SD = 0.14$), than in SANCTION, with an average of 0.09 ($N = 50$, $SD = 0.06$; $p < 0.001$). Thus, in line with [Hypothesis 3](#), our results show that communication refers more explicitly to cartel formation in the absence of sanctioning institutions than when a competition authority is present that may sanction such agreements.³⁰

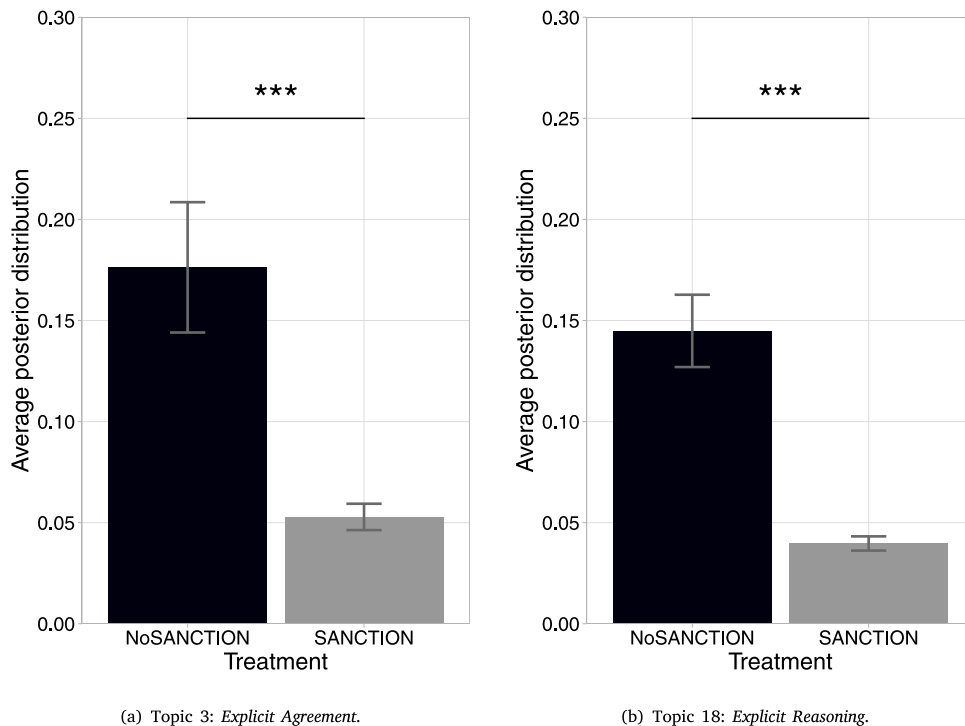


Fig. 6. Average posterior distribution per topic and treatment. ‘***’ refers to $p \leq 0.01$. Error bars indicate standard errors.

5.3. Communication and price setting

Explicit communication. Finally, we turn to investigating whether, and to what extent, the content of communication affects prices. As stated in [Hypothesis 4](#), we expect average prices to be higher with explicit communication than when communication is less explicit. As the presence of sanctioning institutions is likely to affect both the way in which firms communicate with each other and their price setting behavior, we use complementary approaches to shed light on the effect of communication on price setting. First, we restrict attention to the SANCTION treatment and, thereby, hold constant the presence of sanctioning institutions so that we engage in a *ceteris paribus* comparison of prices in markets with more or less explicit communication. Second, we apply causal mediation

²⁹ In Appendix C we provide statistical test results for the separate comparisons between NoSANCTION and FINE OF LENIENCY because we did not analyze communication patterns in [Andres et al. \(2021\)](#) and, thus, cannot be sure that there are no differences in communication patterns between these sub-treatments.

³⁰ A complementary human coding of the communication data finds qualitatively similar results to the machine learning approach. For comparison, two coders independently classified whether a message contains an explicit suggestion to collude, an indirect attempt to do so, or neither of the two. Coding was binary with a message coded one if it contained the relevant content and zero otherwise. The inter-coder agreement is 99.40% for the coding of explicit and 99.57% for indirect communication. The analysis of the results uses means across coders in rounds 2–25. Similar to the topic model, this coding shows that the share of explicit messages is higher in NoSANCTION, at 0.15 ($SD = 0.09$), than in SANCTION, at 0.04 ($SD = 0.04$; $p < 0.001$). These results are also in line with those reported in [Andres et al. \(2021\)](#).

analysis (Imai et al., 2010, 2011, 2013) on the full sample to estimate how much of the treatment effect of sanctioning institutions (treatment variable) on prices (outcome variable) is driven by their effect on communication (mediator variable).³¹

First, we consider only data from the *SANCTION* treatment. We compute the share of communication in a market that can be attributed to the two explicit cartel formation topics according to the LDA and split the sample at the median. We then compare average prices in markets with above-median levels of explicit communication to average prices in markets with below-median levels of explicit communication. In line with Hypothesis 3, the left panel of Fig. 7 shows that average prices in markets with above-median levels of explicit communication (7.01 points, $N = 25$, $SD = 1.14$) are significantly higher than in markets with below-median levels of explicit communication (6.27 points, $N = 25$, $SD = 1.29$; $p = 0.05$). In the right panel of Fig. 7, we compare the cartelization rate according to the weighted average expert judgment for the same median split. Average cartelization as judged by the expert is 0.44 ($N = 25$, $SD = 0.27$) when explicit communication exceeds the median level and it is 0.23 ($N = 25$, $SD = 0.28$) when explicit communication is below the median. This difference is statistically significant ($p = 0.008$), indicating that explicit communication drives cartelization.

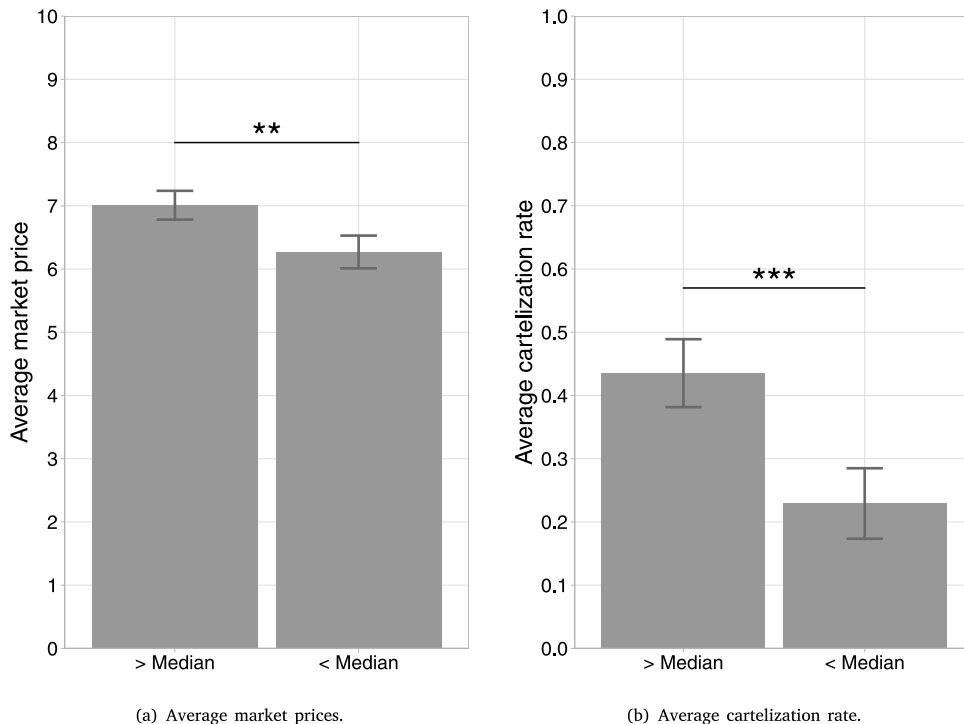


Fig. 7. Price setting and cartelization behavior according to the expert judgment in markets with above and below median levels of explicit communication in *SANCTION*. *** refers to $p \leq 0.05$. **** refers to $p \leq 0.01$. Error bars indicate standard errors.

Second, we run a causal mediation analysis on the full sample to estimate how explicit communication mediates the effect of sanctioning institutions on price setting behavior. We find that the presence of sanctioning institutions decreases the market price directly by 1.82 points on average (95% Confidence interval lower = -2.44, upper = -1.2). The direct effect is statistically significant ($p < 0.001$) and accounts for 82.63% of the total effect on prices of 2.2 points on average (95% Confidence interval lower = -2.6, upper = -1.81). In addition, the presence of sanctioning institutions has an indirect effect through a change in explicit communication. We find that the drop in explicit communication caused by the presence of sanctioning institutions decreases the market price by an additional 0.39 points on average (95% Confidence interval lower = -0.91, upper = 0.05). Fig. 8 illustrates this result.

The mediator effect accounts for 17.53% of the total effect of sanctioning on prices but is not statistically significant at conventional levels with $p = 0.09$.³² Given the specific variance structure of firm behavior in our sample, we interpret this value as a lower bound of the true mediation relationship between explicit communication and prices: while in *NoSANCTION* prices vary only very little around the join-profit-maximizing price of 9 but the dispersion of explicit communication is large, prices in *SANCTION* vary

³¹ For the application of a causal mediation analysis in an experimental setting, the following assumptions must hold: (1) The treatment variable is randomized. (2) The mediator and outcome variables are observed without any intervention of the experimenter. Both assumptions are satisfied in our experimental design.

³² Percentage shares are computed with the original unrounded effects.

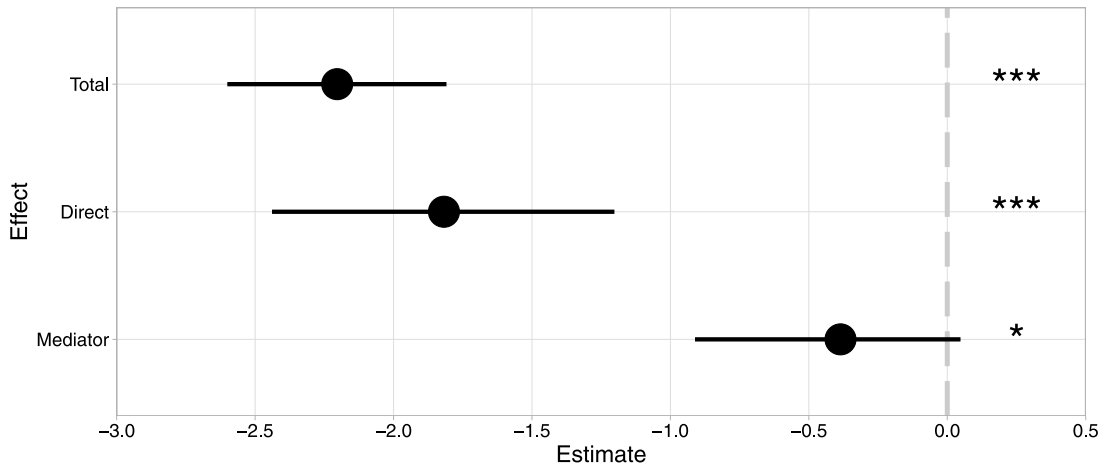


Fig. 8. Average causal mediation analysis: Sanctioning institution (treatment), price setting (outcome) and explicit communication (mediator). ‘**’ refers to $p \leq 0.1$. ‘***’ refers to $p \leq 0.01$. Lines indicate 95% Confidence interval.

substantially but the amount of explicit communication is relatively similar across markets.³³ Thus, the test has little bite because it can neither use the large differences in explicit communication in NoSANCTION to explain the small price differences in this part of the data set, nor can it explain the large prices differences in SANCTION because in this treatment there is not much variation in explicit communication.

Taken together, both the median split within the SANCTION data and the mediation analysis of treatment differences indicate that explicit communication has a positive effect on average market prices. Hence, our data supports Hypothesis 4 that average market prices are higher with explicit communication than when communication is less explicit.³⁴

Indirect communication. In the previous subsections, we focused on the prevalence of communication that explicitly attempts to coordinate on a specific price. To complement these analyses, we now explore alternative communication patterns that are not accounted for by the focus on explicit communication. As we have seen that explicit communication differs substantially between treatments with and without sanctioning institutions, we start with the same between-treatment comparison for the study of indirect communication.

In Fig. 9, we depict the 50 most frequent tokens in treatments NoSANCTION and SANCTION and their relative rank differentials (see Huerta, 2008; Fischer and Normann, 2019; Özkes and Hanaki, 2020), with the most frequently used word having rank 1. We compute the relative rank differential for treatment SANCTION as $\frac{r_{NoS} - r_S}{r_S}$, where r_{NoS} and r_S indicate the rank of a token in treatment NoSANCTION and SANCTION, respectively. The relative rank differential for treatment NoSANCTION is defined analogously. Following Fischer and Normann (2019), we define tokens as more frequent in one treatment than in the other if the relative rank statistic is larger than or equal to one. Tokens that appear to the lower left of the shaded area in Fig. 9 are more frequent in SANCTION than in NoSANCTION. Three tokens stand out: ‘higher’, ‘price’, and ‘authority’. While the latter is clearly an artifact of the experimental design—there was no authority in treatment NoSANCTION—the former two suggest themselves as tokens relating to indirect coordination attempts.

Thus, we define a topic as evidence of indirect attempts to collude if the tokens ‘higher’ or ‘price’ appear in the top ten list of tokens of the respective topic. Following this definition, three topics, numbered 11, 20, and 25, are identified as referring indirectly to collusion. Figure 15 in Appendix F provides information about the most frequent words in these topics. Topic 11 contains tokens such as ‘product’, ‘quality’, ‘customer’, and ‘welfare’ in addition to the defining token ‘price’. As these tokens suggest that the topic relates to discussions about innocent reasons for coordinated price increases, we label this topic *Excuses for high prices*. Topic 20 relates the token ‘higher’ to indirect cartel formation by connecting it to tokens such as ‘go’, ‘slow’, ‘high’, and ‘understand’. Thus, we label this topic *Unspecific appeal to go higher*. Topic 25 connects the token ‘higher’ to some notion of joint actions (‘same’) to increase the benefit (‘welfare’, ‘receives’, ‘needs’). Therefore, we label this topic *Joint benefit*. All three topics explain a significantly larger share of the communication in SANCTION than in NoSANCTION.³⁵

³³ The variance of average pricing behavior equals 0.2 in NoSANCTION and 1.6 SANCTION. The variance of explicitness equals 0.02 in NoSANCTION and 0.003 in SANCTION. The small numbers for the variances of explicitness account for the Dirichlet distribution of communication. Both differences in dispersion are statistically significant between treatments in an Ansari–Bradley test, with $p = 0.003$ for average prices and $p = 0.005$ for explicitness.

³⁴ As we compare averages at the market level, these comparisons cannot inform us on the dynamic interplay between communication and price setting. Section 5.4 complements the analysis with exploratory insights from the hand-coded chat data.

³⁵ *Excuses for high prices*: SANCTION 0.04 ($N = 50$, $SD = 0.07$), NoSANCTION 0.01 ($N = 23$, $SD = 0.01$), $p < 0.001$. *Unspecific appeal to go higher*: SANCTION 0.04 ($N = 50$, $SD = 0.05$), NoSANCTION 0.02 ($N = 23$, $SD = 0.01$), $p < 0.001$. *Joint benefit*: SANCTION 0.03 ($N = 50$, $SD = 0.03$), NoSANCTION 0.02 ($N = 23$, $SD = 0.01$), $p < 0.001$.

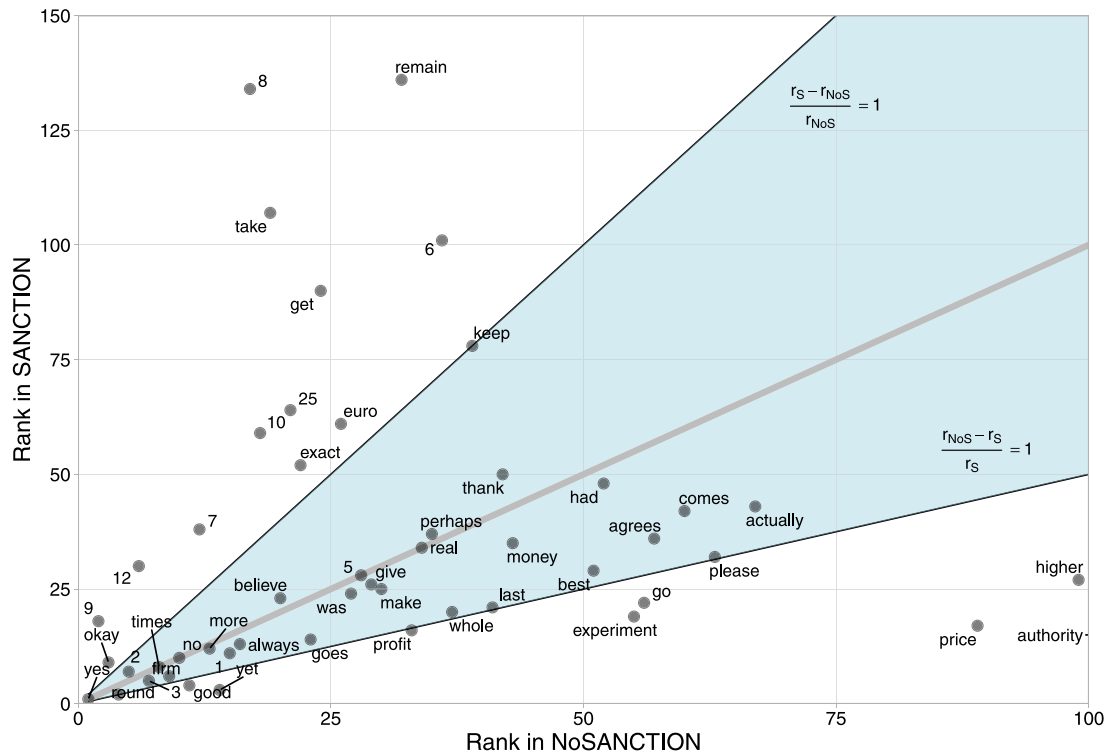


Fig. 9. Frequency rankings of the 50 most used tokens in both treatments. The two black lines indicate the border where the two relative rank differentials $\frac{r_{NoS} - r_S}{r_S}$ and $\frac{r_S - r_{NoS}}{r_{NoS}}$ are equal to 1. Tokens outside the shaded area have a relative rank differential exceeding 1.

To understand how indirect communication affects price setting behavior in the SANCTION treatment, we first compute the share of communication in a market that can be attributed to the three topics. To see how indirect communication adds to explicit communication, Fig. 10 plots the share of communication that can be attributed to indirect and explicit communication in treatment SANCTION.

Next, we compare average prices in markets with more or less indirect communication. To account for explicit communication that appears along with low incidence of indirect messages, we divide the data into an upper and a lower part along a median split according to the share of explicit communication. We then compare average prices in markets with above- and below-median levels of indirect communication, holding explicitness either above or below the median. Both in markets with above- and below-median levels of explicit communication, average prices in markets with above-median levels of indirect communication tend to be lower than average prices in markets with below-median levels of indirect communication.³⁶ Thus, indirect communication does not seem to have a positive effect on prices, rather there is a slightly negative effect.

A final piece of evidence concerning the effect of communication on prices comes from a comparison of prices in the first round without any communication to those in the second round, where communication sets in. While average prices increase sharply between these two rounds in NoSANCTION, they do not change in SANCTION as can be seen in Fig. 5. In NoSANCTION, the average market price of 6.29 points ($N = 23$, $SD = 1.1$) in the first round increases significantly to 8.46 points ($N = 23$, $SD = 1.18$) in the second round ($p < 0.001$). In contrast, average market prices in SANCTION do not change significantly between the first (6.03, $N = 50$, $SD = 1.04$) and the second rounds (6.03, $N = 50$, $SD = 1.55$; $p = 0.97$).³⁷ Thus, the indirect communication in SANCTION appears to be too implicit to allow firms to coordinate on jointly optimal price setting right away, implying that prices do not increase above the

$p = 0.002$. Again, the hand-coding yields similar results to the topic model: the share of indirect messages is higher in SANCTION, at 0.09 ($SD = 0.12$), than in NoSANCTION, at 0.01 ($SD = 0.01$; $p < 0.001$).

³⁶ Explicit above median: the average price with below-median levels of indirect communication is 7.46 points ($N = 11$, $SD = 1.2$), the average price with above-median levels of indirect communication is 6.66 points ($N = 14$, $SD = 1$), $p = 0.05$. Explicit below median: the average price with below-median levels of indirect communication is 6.62 points ($N = 14$, $SD = 1.25$), the average price with above-median levels of indirect communication is 5.81 points ($N = 11$, $SD = 1.26$), $p = 0.09$.

³⁷ Both p -values reported refer to the results of a two-sided Wilcoxon signed-rank test with continuity correction. Average market prices in the first round do not differ significantly between NoSANCTION and SANCTION ($p = 0.3$) but the price setting differs significantly between the treatments as soon as communication sets in ($p < 0.001$).

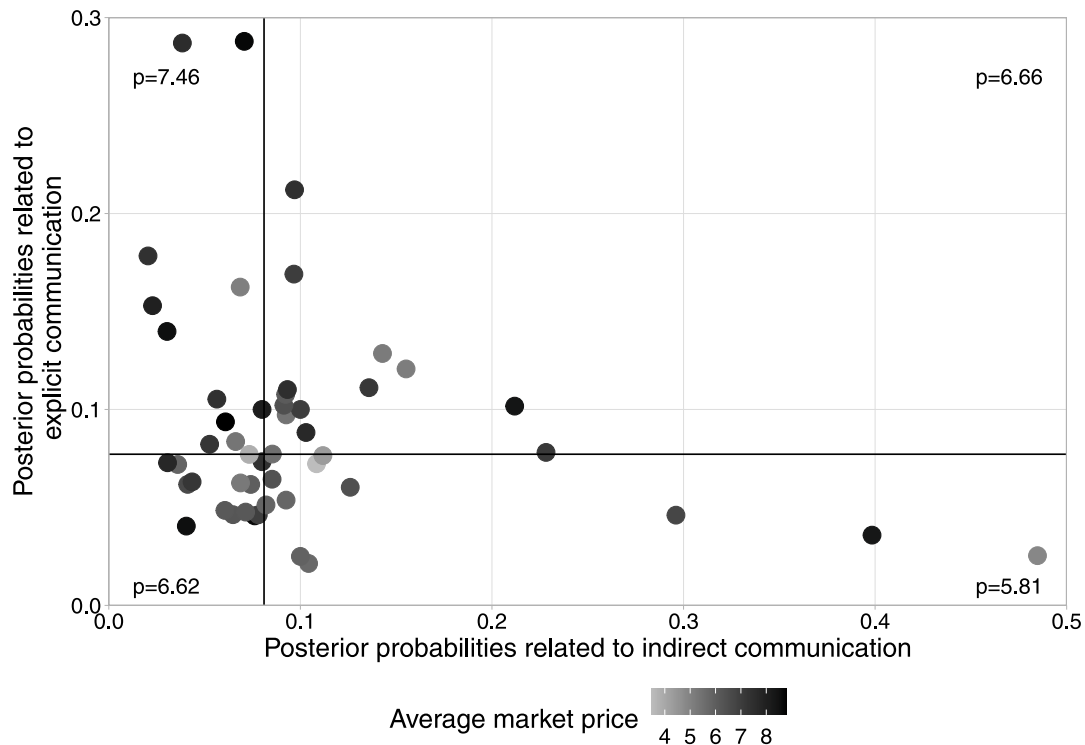


Fig. 10. Average pricing behavior, posterior probabilities related to indirect and explicit communication per market in SANCTION. Black lines along the axis show the median. The average market price for each quadrant is displayed in the respective corner.

first round benchmark. Indirect communication (or the remaining, not cartel-related communication), however, is sufficient to keep average market prices at the level observed in the first round. This result suggests some positive effect of indirect communication because previous studies show that firms tend to converge downward to Nash pricing in the absence of communication. But indirect communication is apparently insufficient to sustain collusion at, or close to, the joint-profit-maximizing price of nine. With respect to the question of “how indirect can communication be and still be reasonably effective?”, raised by Harrington et al. (2016), our data suggests that effective coordination needs explicitness. Already moderate sanctions are sufficient to make communication sufficiently indirect to deter immediate cartel formation and to keep market prices down.

5.4. Explorations into the dynamics of communication, market outcomes, and law enforcement

So far, we have looked at statistics that average behavior over rounds to arrive at market-level averages. While this proved insightful in many aspects, the analysis of sanctioning institutions naturally leads to questions about the dynamics of behavior and communication. The results from the LDA algorithm cannot be used to investigate dynamic aspects as it only provides a classification of communication at the market level and cannot be adjusted to yield reliable round-wise estimates in our sample. In order to provide some insights into the dynamics of communication and cartelization, we had two independent coders classify individual chat messages with respect to two criteria: whether a message explicitly or indirectly attempts to coordinate prices or collude. Based on these classifications, we compute for each market and round the shares of explicit and indirect messages in the total number of messages sent and combine these new variables with information on cartelization and fines. We note that this analysis is exploratory; our experiment was not designed to engage in dynamic analyses and we have not preregistered corresponding hypotheses. Therefore, we mostly rely on descriptive analyses based on the detailed Figures 11, 12, and 13 in Appendix D.

First, we ask whether we can identify specific communication strategies. Noting that we only observe behavioral patterns but not complete contingent action plans, we searched for evidence of two intuitive patterns. On the one hand, firms might try to only communicate to establish a cartel and remain silent as soon as that goal is reached in order not to provide further incriminating evidence of their coordination. On the other hand, firms could try to constantly communicate and, thereby, hide coordination messages in larger amounts of innocuous content. Both patterns are repeatedly observed in our sample but the second pattern is more common. Many markets exhibit close to constant communication in which explicit and indirect attempts to coordinate occur here and there. One group implements this strategy particularly impressively as they spell out that they want to spam the authority with content to hinder the detection of their agreement and indeed send close to 60 messages per round resulting in an excessively

long chat protocol. In contrast, two groups explicitly describe the first strategy of talk then silence but do not manage to implement it (see details in Appendix D). When looking more closely at Figures 11 and 12, where collusion can be fined, it is apparent that most of the time, not talk alone but only communication that relates to the issue of coordination pushes a group to full cartelization. However, whereas such communication appears to be sufficient for successful cartelization in the NoSanction treatment (see Figure 13), it is not so when sanctioning institutions are present (see Figures 11 and 12).³⁸ A potential explanation is that explicit proposals to coordinate are immediately taken up if no sanction must be feared whereas firms are more cautious to agree with such a proposal if collusion is subject to fines.

Second, we use Figures 11 and 12 to investigate how communication patterns affect profits and relate to fines. We observe that the pattern of “talk and hide behind innocuous content” is not only more common but it also relates to higher levels of cartelization which by definition are associated with higher profits. We further observe that groups in which firms continuously talk appear better able to survive an investigation without breaking down than groups in which talk is rare. However, those cartels that remain silent after being established appear to be fined less in case an investigation takes place and more explicit communication increases the risk of a fine, consistent with the idea that little evidence of an infringement is less likely to result in a fine than a large and clear body of evidence.

We hope that these glimpses into the potential dynamics of communication, how they relate to market outcomes and how they interact with law enforcement will be taken up in future studies tailored to these important questions. A further question left for future studies is how unilateral deviations relate to communication behavior and whether sanctions affect this relationship.

6. Discussion and conclusion

Existing experimental studies find that sanctioning coordinated pricing behavior is an effective instrument to hinder cartel formation. Our study is a first approach to understand *how* the sanctioning of cartel formation affects the coordination process of firms and why sanctions are effective. To investigate how sanctioning institutions affect the communication between firms, we use an innovative experimental setup where a free-form communication channel is always open and sanctions are decided upon by properly incentivized participants in the role of the competition authority. Using a machine learning approach, we quantify the content of the firms’ communication such that we can study the degree to which communication contains explicit and indirect attempts to form a cartel.

In line with the literature, we find that sanctions reduce the prevalence of cartel formation and average market prices significantly. This complements the few empirical studies showing evidence for a positive welfare effect of competition policy (Buccrossi et al., 2013) and cartel prohibition (Normann and Tan, 2014). In addition, our experimental approach provides evidence with respect to the firms’ communication, which is even harder to collect in the field: the experimental results reveal that the presence of sanctioning institutions reduces the extent of explicit price coordination in the communication of firms by about two thirds compared to the situation without any sanctions. An additional analysis using a quantitative measure of explicit communication, suggests that the reduction in explicit communication makes up about one fifth of the total treatment effect, thus strengthening the direct deterring effect of sanctions on cartel formation. An exploratory analysis of the remaining chat communication indicates that firms switch to indirect price coordination when sanctioning institutions are in place. While effective in preventing unraveling toward the Nash equilibrium, however, these indirect approaches are insufficient in raising average market prices above the price level observed in the first round of the interaction where no communication was possible.

We expect our findings to be useful in at least two respects. First, we show that explicit communication is effective in achieving a joint increase in the firms’ prices whereas indirect price communication is not. This result proves a link between explicit communication and illegal conduct that may inform courts in their judgment of whether or not a certain conduct violates competition law. Specifically, we show that the detailed analysis of communication data may help to define the boundary between tacit collusion and explicit cartel formation. Second, our study provides potentially useful insights for screening approaches such as e-discovery that are already used in practice. As part of their compliance policy, many companies try to uncover and then eliminate unlawful behavior of their own employees by screening the firm’s internal communication data for suspicious patterns and content — before legal institutions start an investigation. Our study suggests that the presence of screening will already improve compliance by making communication less explicit and thereby less effective.

When we compare the topics as identified by our machine learning algorithm to a hand-coded classification of the same messages into explicit and indirect attempts to collude, we find similar results: irrespective of the classification method, there are more explicit messages in the treatment without sanctioning institutions and more indirect messages in the treatment with sanctions. The topics as identified by the algorithm also relate to those identified in the hand-coded approach of Cooper and Kühn (2014) or Dijkstra et al. (2021): similar to our *Explicit Agreement* and *Explicit Reasoning* topics they report that a large share of messages is coded as proposals to play a specific strategy and agreement with such proposals; and similar to our *Joint Benefit* topic, Cooper and Kühn (2014) report a category of appeal to the mutual benefits of cooperation. Thus, our machine learning approach provides results that are very similar to those from the traditional hand-coding method. We also used the hand-coded data from our experiment to verify that our results are robust to using the traditional approach (see Footnotes 30 and 35). In addition, the machine-learning approach reveals novel insights. In particular, the topic analysis helps to better understand how words relate to other words and it

³⁸ For instance, groups 1, and 4, in FINE and groups 5, 8, 9, 24, 25, 39, 49 in LENIENCY communicate but never manage to form a cartel. Several of these competitive groups exhibit clearly positive shares of explicit or indirect coordination attempts.

can quantify the share of communication related to a specific topic. Furthermore, the procedure could easily deal with far larger data sets, which is an advantage that is becoming increasingly important as sample sizes in experimental studies are increasing in an attempt to obtain more reliable and replicable results.

In the future, our machine learning approach might also help explain why sanctioning institutions sometimes miss their goal. For example, in the experimental studies by Andersson and Wengström (2007) and Bigoni et al. (2012), sanctions (implemented as a monetary cost of communication) reduce cartelization but tend to increase prices. Furthermore, Hinloopen and Onderstal (2014) report results from an experimental first price auction where a leniency rule increases the stability of cartels among the bidders, while Berlin et al. (2018) present empirical evidence on a poorly designed anti-corruption program that failed to reduce bribery. Text analysis along the lines of our approach might be useful to understand such failure of sanctions to affect behavior into the desired direction.

In a next step, it would be interesting to endogenize the mode of coordination and study why some firms decide in favor of explicit communication while others prefer more indirect forms of communication or even remain entirely silent. Such heterogeneity is also observed by Harrington et al. (2016). Possibly influential factors in firm behavior are differential beliefs about the success probability of indirect communication, misperceptions with respect to the authority's judgment of what counts as a cartel, i.e., where exactly communication switches from being innocuous to being evidence of unlawful agreements, and the risk attitude of decision makers.

Appendix A. Supplementary material: 1) replication package 2) appendix with theoretical background, results from additional analyses, and instructions.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.euroecorev.2022.104331>.

References

- Andersson, O., Wengström, E., 2007. Do antitrust laws facilitate collusion? Experimental evidence on costly communication in duopolies. *Scand. J. Econ.* 109 (2), 321–339.
- Andres, M., Bruttel, L., Friedrichsen, J., 2021. The leniency rule revisited: Experiments on cartel formation with open communication. *Int. J. Ind. Organ.* 76 (102728).
- Apesteguia, J., Dufwenberg, M., Selten, R., 2007. Blowing the whistle. *Econom. Theory* 31 (1), 143–166.
- Berlin, M., Qin, B., Spagnolo, G., 2018. Leniency, Asymmetric Punishment and Corruption: Evidence from China. CEIS Working Paper No. 431.
- Bigoni, M., Fridolfsson, S.-O., Le Coq, C., Spagnolo, G., 2012. Fines, leniency, and rewards in antitrust. *Rand J. Econ.* 43 (2), 368–390.
- Blei, D.M., 2012. Probabilistic topic models. *Commun. ACM* 55 (4).
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.
- Blume, A., Heidhues, P., 2008. Modeling tacit collusion in auctions. *J. Inst. Theor. Econ.* 163–184.
- Bouchet-Valat, M., 2019. SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library. R Package Version 0.6.0.
- Brandts, J., Cooper, D.J., Rott, C., 2019. Communication in laboratory experiments. In: *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing.
- Bucciossi, P., Ciari, L., Duso, T., Spagnolo, G., Vitale, C., 2013. Competition policy and productivity growth: An empirical assessment. *Rev. Econ. Stat.* 95 (4), 1324–1336.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., Blei, D.M., 2009. Reading tea leaves: How humans interpret topic models. In: *Neural Information Processing Systems*. pp. 288–296.
- Clark, R., Houde, J.-F., 2014. The effect of explicit communication on pricing: Evidence from the collapse of a gasoline cartel. *J. Ind. Econ.* 62 (2), 191–228.
- Coffman, L., Niehaus, P., 2015. Pathways of persuasion. Mimeo.
- Cooper, D.J., Kagel, J.H., 2005. Are two heads better than one? Team versus individual play in signaling games. *Amer. Econ. Rev.* 95 (3), 477–509.
- Cooper, D.J., Kühn, K.-U., 2014. Communication, renegotiation, and the scope for collusion. *Am. Econ. J. Microecon.* 6 (2), 247–278.
- Davis, D.D., Holt, C.A., 1998. Conspiracies and secret discounts in laboratory markets. *Econ. J.* 108 (448), 736–756.
- Dijkstra, P.T., Haan, M.A., Schoonbeek, L., 2021. Leniency programs and the design of antitrust: experimental evidence with free-form communication. *Rev. Ind. Org.* 59 (1), 13–36.
- Dunbar, R.I.M., 1998. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.
- European Union, 2012. Consolidated version of the treaty on the functioning of the European union. Official J. Eur. Union 101 (1), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012E/TXT&from=DE>. (Accessed 23 January 2020).
- Feinerer, I., Hornik, K., Meyer, D., 2008. Text mining infrastructure in R. *J. Stat. Softw.* 5 (5), 1–54.
- Fink, N., Schmidt-Dengler, P., Stahl, K., Zulehner, C., 2017. Registered cartels in Austria: An overview. *Eur. J. Law Econ.* 44 (3), 385–422.
- Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10 (2), 171–178.
- Fischer, C., Normann, H.-T., 2019. Collusion and bargaining in asymmetric Cournot duopoly—An experiment. *Eur. Econ. Rev.* 111 (2), 360–379.
- Fonseca, M.A., Normann, H.-T., 2012. Explicit vs. tacit collusion – the impact of communication in oligopoly experiments. *Eur. Econ. Rev.* 56 (8), 1759–1772.
- Fourberg, N., 2018. Let's Lock Them In: Collusion Under Consumer Switching Costs. DICE Discussion Paper No. 296, (296).
- Garrod, L., Olczak, M., 2018. Explicit vs tacit collusion: The effects of firm numbers and asymmetries. *Int. J. Ind. Organ.* 56, 1–25.
- Genesove, D., Mullin, W.P., 2001. Rules, communication, and collusion: Narrative evidence from the sugar institute case. *Amer. Econ. Rev.* 91 (3), 379–398.
- Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as data. *J. Econ. Lit.* 57 (3), 535–574.
- Greiner, B., 2015. Subject pool recruitment procedures: Organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1 (1), 114–125.
- Griffiths, T.L., Steyvers, M., 2004. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* 101 (Suppl. 1), 5228–5235.
- Griffiths, T.L., Steyvers, M., 2007. Probabilistic topic models. *Handb. Latent Semant. Anal.* 427 (7), 424–440.
- Hansen, S., McMahon, M., Prat, A., 2017. Transparency and deliberation within the FOMC: A computational linguistics approach. *Q. J. Econ.* 133 (2), 801–870.
- Harrington, J.E., 2006. How do cartels operate? *Found. Trends Microecon.* 2 (1), 1–105.
- Harrington, J.E., Gonzalez, R.H., Kujal, P., 2016. The relative efficacy of price announcements and express communication for collusion: Experimental findings. *J. Econ. Behav. Organ.* 128, 251–264.
- Hinloopen, J., Onderstal, S., 2014. Going once, going twice, reported! Cartel activity and the effectiveness of antitrust policies in experimental auctions. *Eur. Econ. Rev.* 70, 317–336.
- Hornik, K., Grün, B., 2011. Topicmodels: An R package for fitting topic models. *J. Stat. Softw.* 40 (13), 1–30.
- Houser, D., Xiao, E., 2011. Classification of natural language messages using a coordination game. *Exp. Econ.* 14 (1), 1–14.

- Huerta, J., 2008. Relative rank statistics for dialog analysis. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. pp. 965–972.
- Hyttinen, A., Steen, F., Toivanen, O., 2018. Cartels uncovered. *Am. Econ. J. Microecon.* 10 (4), 190–222.
- Hyttinen, A., Steen, F., Toivanen, O., 2019. An anatomy of cartel contracts. *Econ. J.* 129 (621), 2155–2191.
- Imai, K., Luke, K., Tingley, D., 2010. A general approach to causal mediation analysis. *Psychol. Methods* 15 (4), 309–334.
- Imai, K., Luke, K., Tingley, D., Yamamoto, T., 2011. Unpacking the black box: Learning about causal mechanisms from experimental and observational studies. *Am. Polit. Sci. Rev.* 105 (4), 765–789.
- Imai, K., Tingley, D., Yamamoto, T., 2013. Experimental designs for identifying causal mechanisms. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* 176 (1), 5–51.
- Isaac, R.M., Ramey, V., Williams, A.W., 1984. The effects of market organization on conspiracies in restraint of trade. *J. Econ. Behav. Organ.* 5 (2), 191–222.
- Moellers, C., Normann, H.-T., Snyder, C.M., 2017. Communication in vertical markets: Experimental evidence. *Int. J. Ind. Organ.* 50, 214–258.
- Mueller, H., Rauh, C., 2018. Reading between the lines: Prediction of political violence using newspaper text. *Am. Polit. Sci. Rev.* 112 (2), 358–375.
- Newman, D., Asuncion, A., Smyth, P., Welling, M., 2009. Distributed algorithms for topic models. *J. Mach. Learn. Res.* 10, 1801–18028.
- Normann, H.-T., Tan, E.S., 2014. Effects of different cartel policies: evidence from the German power-cable industry. *Ind. Corp. Change* 23 (4), 1037–1057.
- Odenkirchen, J., 2018. Pricing Behavior in Partial Cartels. DICE Discussion Paper No. 299, (299).
- Ormosi, P.L., 2014. A tip of the iceberg? The probability of catching cartels. *J. Appl. Econometrics* 29 (4), 549–566.
- Özkes, A., Hanaki, N., 2020. Talkin' bout cooperation. WU Vienna University of Economics and Business, Department of Strategy and Innovation Working Paper Series No. 08/2020.
- Penczynski, S.P., 2019. Using machine learning for communication classification. *Exp. Econ.* 22 (4), 1002–1029.
- Rauh, C., 2019. Measuring Uncertainty at the Regional Level Using Newspaper Text. Université de Montréal. Département de Sciences Économiques.
- Whinston, M.D., 2008. *Lectures on Antitrust Economics*. Vol. 1. The MIT Press.

Further reading

- Blonski, M., Ockenfels, P., Spagnolo, G., 2011. Equilibrium selection in the repeated Prisoner's dilemma: Axiomatic approach and experimental evidence. *Am. Econ. J.: Microecon.* 3 (3), 164–192.
- Dal Bó, P., Fréchette, G.R., 2011. The evolution of cooperation in infinitely repeated games: Experimental evidence. *Am. Econ. Rev.* 101 (1), 411–429.
- Albors-Llorens, A., 2006. Horizontal agreements and concerted practices in EC competition law: Unlawful and legitimate contacts between competitors. *Antitrust Bull.* 51 (4), 837–876.
- European Union, 2019. Eu rules on concerted practices and agreements between companies. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:4348235&from=EN>, Accessed: 2020-02-28.
- Odudu, O., 2010. The wider concerns of competition law. *Oxford J. Legal Stud.* 30 (3), 599–613.
- Whish, R., Bailey, D., 2015. *Competition Law*, Vol. 8. Oxford University Press.