CEPA DP No. 53

OCTOBER 2022

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

Maximilian Andres
Lisa Bruttel
Jana Friedrichsen

Center for Economic Policy Analysis

CEPA Discussion Papers

Center for Economic Policy Analysis
https://www.uni-potsdam.de/cepa

University of Potsdam
August-Bebel-Straße 89, 14482 Potsdam
Tel.: +49 331 977-3225
Fax: +49 331 977-3210
E-Mail: dp-cepa@uni-potsdam.de

ISSN (online) 2628-653X

CEPA Discussion Papers can be downloaded from RePEc
https://ideas.repec.org/s/pot/cepadp.html

Published online at the Institutional Repository of the University of Potsdam
https://doi.org/10.25932/publishup-56223

CEPA Discussion Paper No. 53 October 2022

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

Maximilian Andres
University of Potsdam

Lisa Bruttel
University of Potsdam

Jana Friedrichsen
Humboldt-Universität zu Berlin, WZB Berlin Social Science Center, DIW Berlin

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-deterring 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.

Keywords: cartel, collusion, communication, machine learning, experiment
JEL Codes: C92, D43, L41

Corresponding author:
Maximilian Andres
University of Potsdam
Department of Economics
August-Bebel-Str. 89
14482 Potsdam
Germany
Email: maximilian.andres@uni-potsdam.de

*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/).

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.

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

[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..."

[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.

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 cartels[3] 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

[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).

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 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).

[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).

[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.

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

Figure 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

[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.

free form chat for 60 seconds 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 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.

[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).

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

[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.

experiment in cash and were paid their payoff from the investigations 2-3 weeks after the experiment by bank transfer.

Expert judgment o 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 hours. 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.

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

[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).

over documents and a Gibbs Sampling procedure[19] 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] Figure 2 illustrates this score for up to 100 topics. The solid line in Figure 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 Figure 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 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.

[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).

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

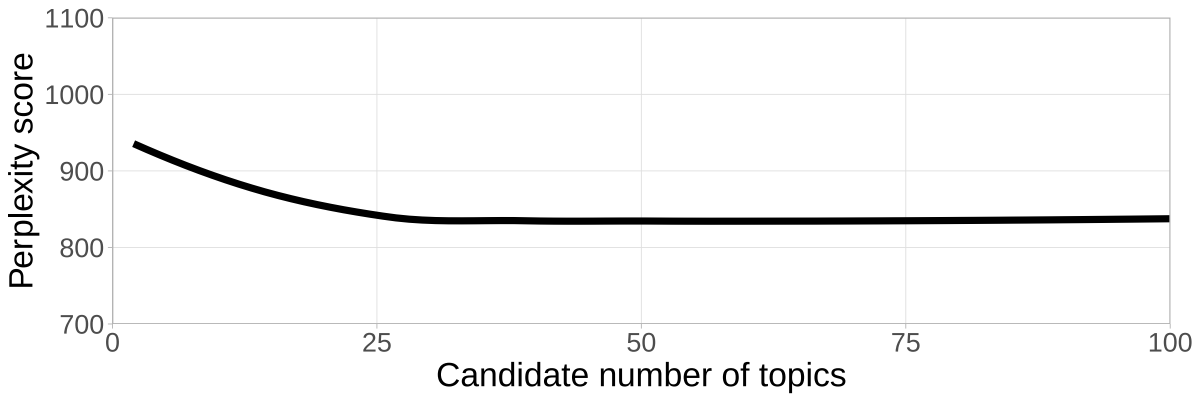


Figure 2: Perplexity score when fitting the trained model to the hold-out set.

The representation comes in form of the estimated posterior distribution of the 25 topics over the 73 documents or group chats.[24]

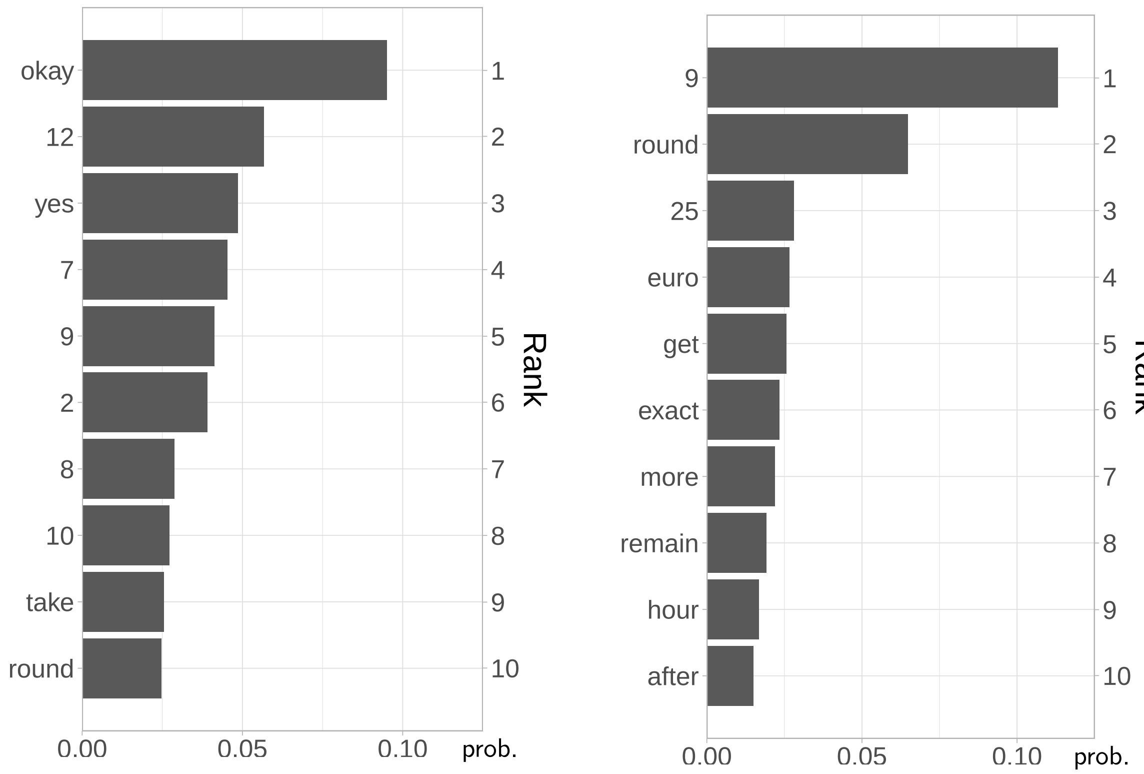
As Hypotheses 3 and 4 refer to the use of explicit communication about collusive practices, we screen the estimated topics for evidence of such explicitness and concentrate on those topics for all further steps of our analysis. Based on the pre-registration of this study, we define a topic as evidence 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. Figure 3 summarizes key information for the two explicit topics.[25]

In topic 3, depicted in Figure 3a, 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 Figure 3b, 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*.

[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.



(a) Topic 3: *Explicit Agreement.*

(b) Topic 18: *Explicit Reasoning.*

Figure 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.

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 rule 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.

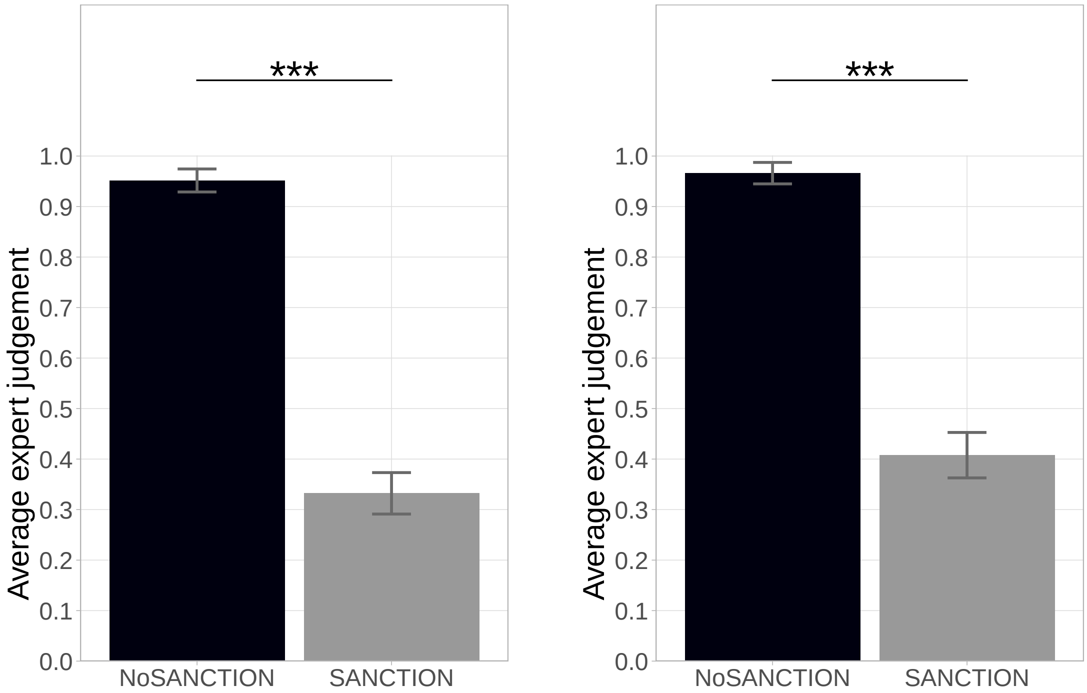
Our sample contains 23 independent markets in NoSanction, and 50 in Sanction. In Sanction, we observe that an investigation takes place in 11% of rounds on average and in 56% of all investigations conducted, a strictly positive fine is imposed on the investigated firms. By definition, investigations 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.

	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	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

Table 1: Average cartelization rate, market price, explicit and indirect communication split up by treatment. Standard deviations in parentheses.

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 market[27]



(a) Weighted measure.

(b) Unweighted measure.

Figure 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 ($p < 0.001$). Figure 4 illustrates this finding. Thus, our data clearly supports Hypothesis 1 that sanctioning institutions reduce cartelization.

[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 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.

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. Figure 5 illustrates that this difference also persists at the level of the individual round and does not change over time.

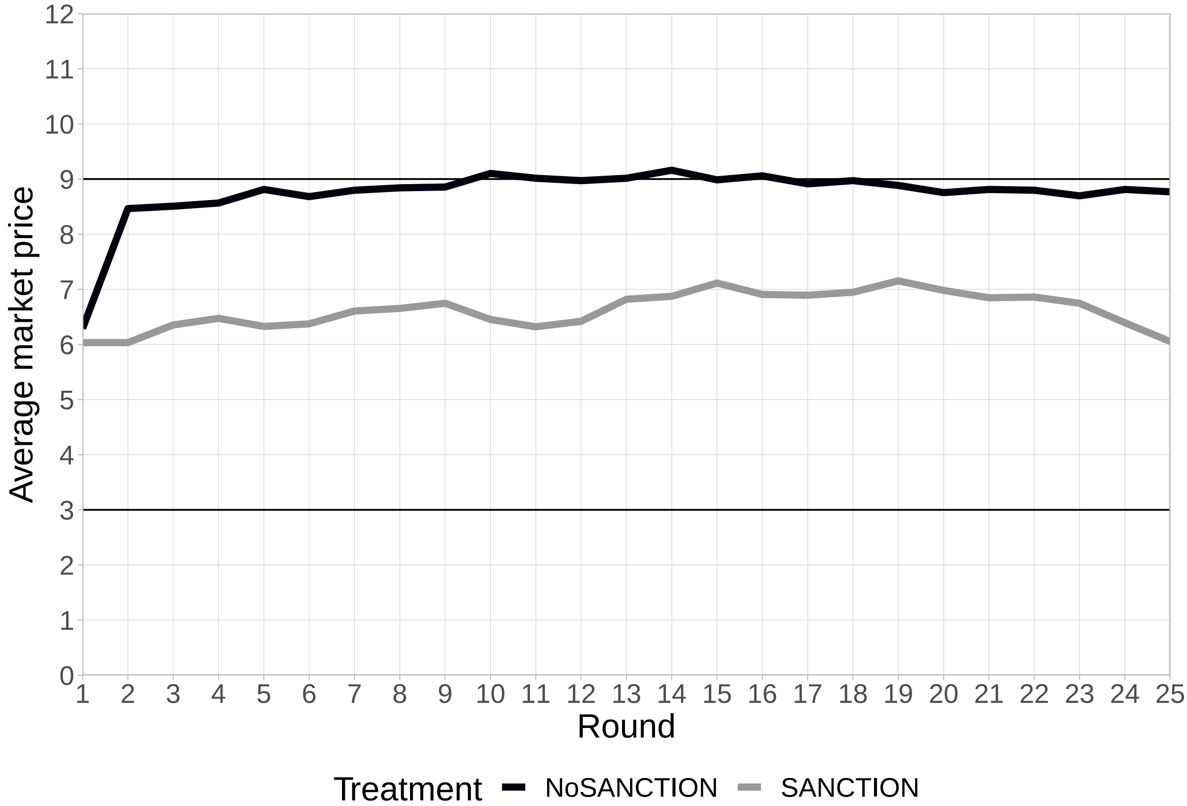


Figure 5: Average market prices over time split up by treatment. Nine is the joint profit maximizing price and three the Nash equilibrium price of the stage game.

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 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 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%).

	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	$p = 0.79$	$p = 0.02$	$p = 0.01$

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.

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. Figure 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 ($p < 0.001$).

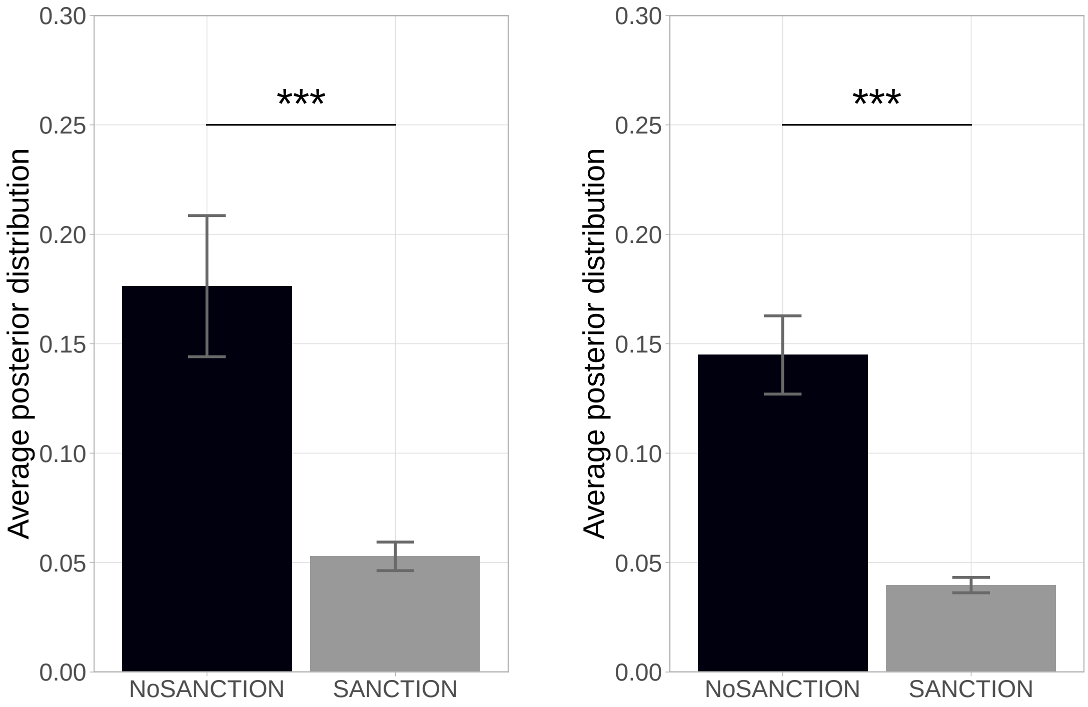
When we also consider the total amount of explicit communication by summing up the average posterior probabilities of *Explicit Reasoning* and *Explicit Agreement*, we find that the average posterior probability of such explicit communication is significantly higher in NoSanction, with an average of 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]

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 analysis (Imai et al., 2010, 2011, 2013) on the full sample to estimate how much

[29] In Appendix C we provide statistical test results for the separate comparisons between NoSanction and Fine or 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 subtreatments.

[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).



(a) Topic 3: *Explicit Agreement*. (b) Topic 18: *Explicit Reasoning*.

Figure 6: Average posterior distribution per topic and treatment. '***' refers to $p \leq 0.01$. Error bars indicate standard errors.

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 Figure 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 Figure 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.

[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.

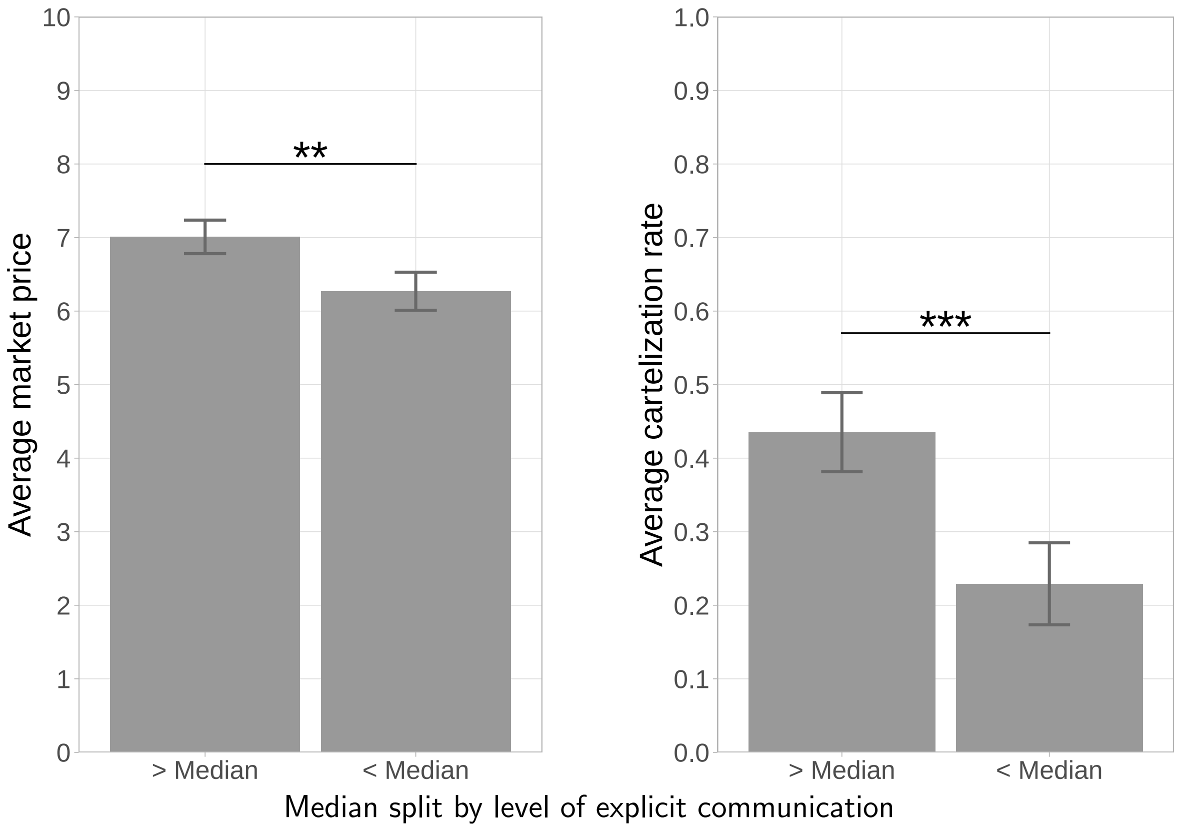


Figure 7: Price setting and cartelization behavior according to the expert judgment in markets with above and below median levels of explicit communication in Sanction. '**' refers to $p \leq 0.05$. '***' refers to $p \leq 0.01$. Error bars indicate standard errors.

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). Figure 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

[32]Percentage shares are computed with the original unrounded effects.

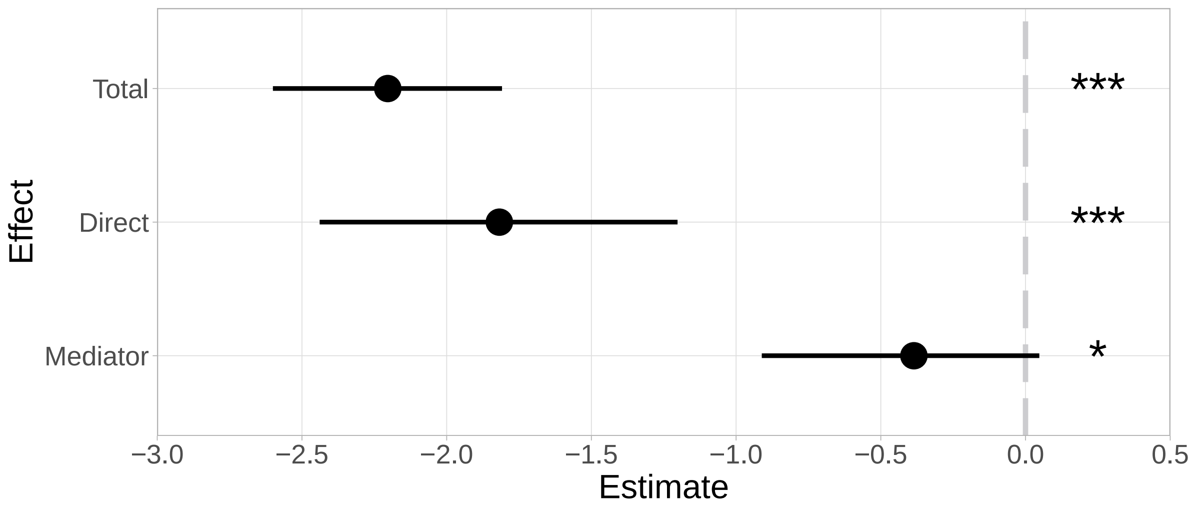


Figure 8: Average causal mediation analysis: Sanctioning institution (treatment), price setting (outcome) and explicit communication (mediator). '*' refers to $p \leq 0.1$. '***' refers to $p \leq 0.01$. Lines indicate 95% Confidence interval.

the dispersion of explicit communication is large, prices in Sanction vary 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 in-

[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 ifrom the hand-coded chat data.

stitutions, we start with the same between-treatment comparison for the study of indirect communication.

In Figure 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 Figure 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.

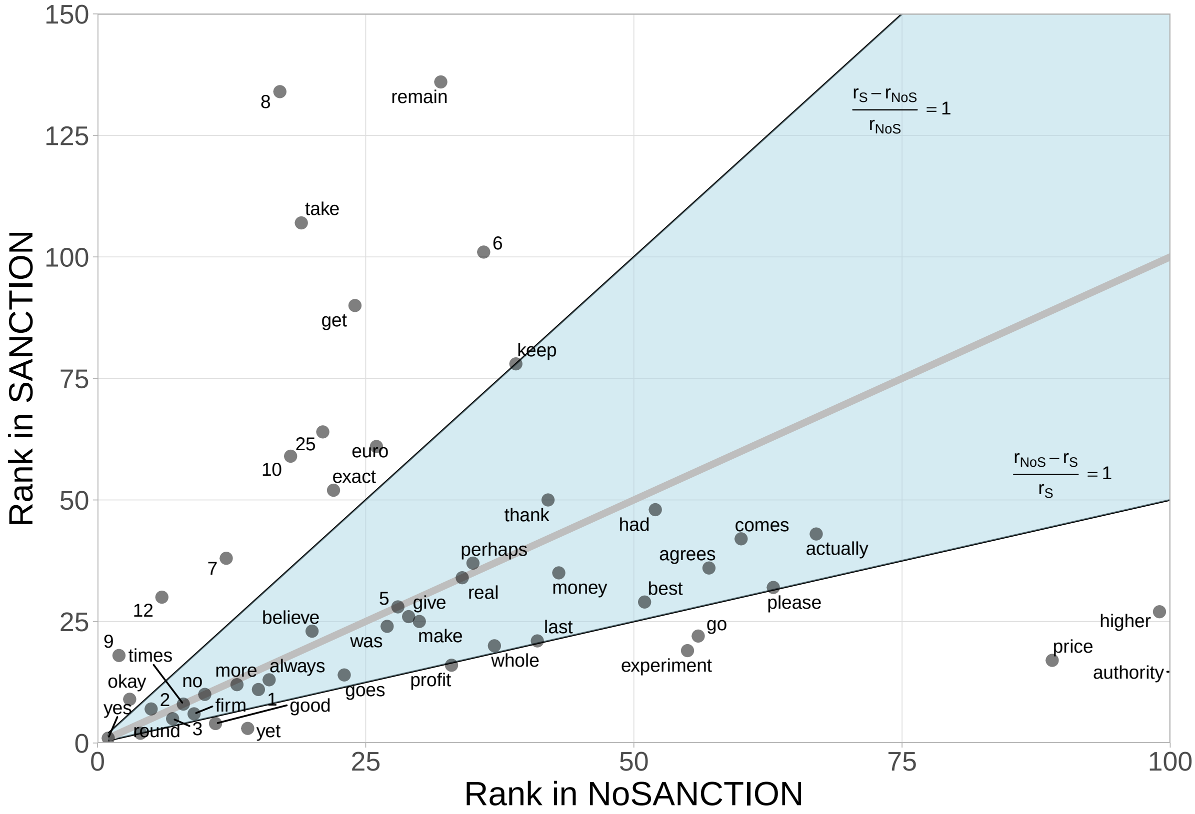


Figure 9: Frequency rankings of the 50 most used tokens in both treatments. The two black lines indicate the border where the two relative rank differentials $\frac{r_{NoS}-r_S}{r_S}$ and $\frac{r_S-r_{NoS}}{r_{NoS}}$ are equal to 1. Tokens outside the shaded area have a relative rank differential exceeding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

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]

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, Figure 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

[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 ($N = 23$, $SD = 0.01$), $p = 0.002$. 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 ($N = 11$, $SD = 1.26$), $p = 0.09$.

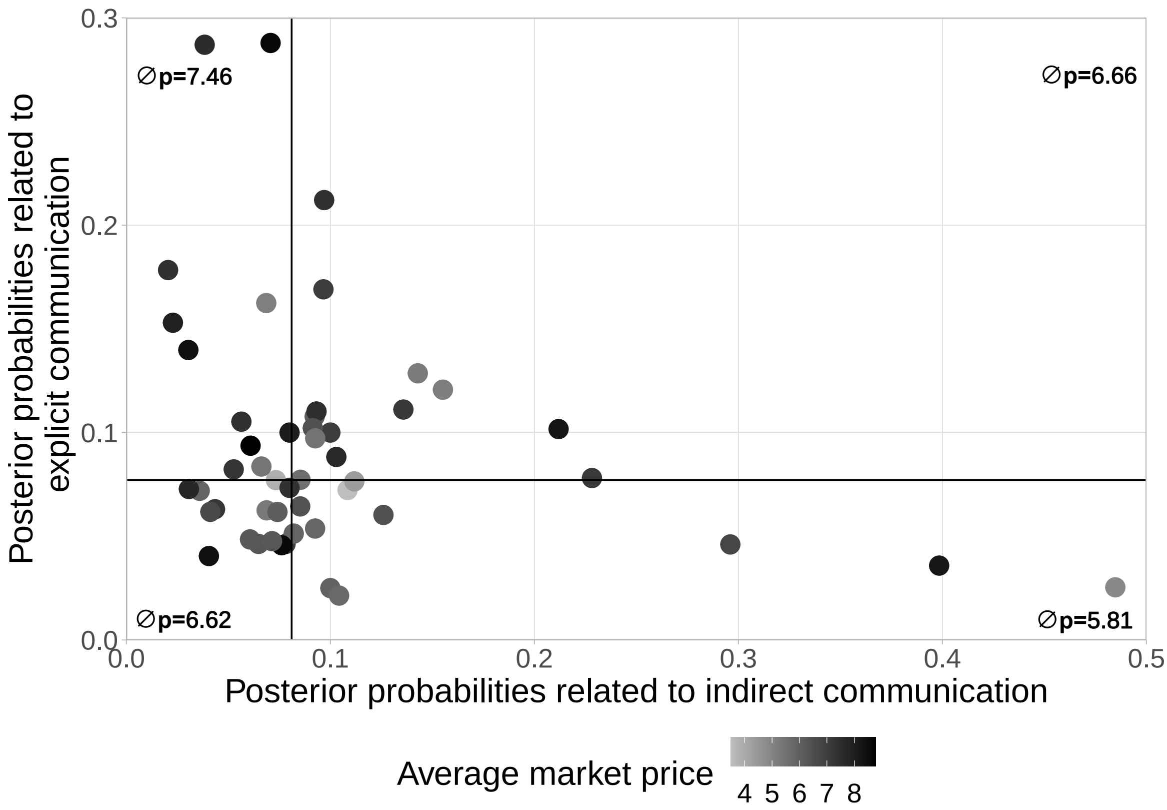


Figure 10: Average pricing behavior, posterior probabilities related to indirect and explicit communication per market in SANCTION. Black lines along the axis show the median. The average market price for each quadrant is displayed in the respective corner.

two rounds in NOSANCTION, they do not change in SANCTION as can be seen in Figure 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, $N = 50$, $SD = 1.55$; $p = 0.9$).[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 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

[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 ($p < 0.001$).

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 (Buccirossi et al., 2013)

[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.

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 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.

References

Albors-Llorens, A. (2006). Horizontal agreements and concerted practices in EC competition law: Unlawful and legitimate contacts between competitors. *The Antitrust Bulletin 51*(4), 837–876.

Andersson, O. and E. Wengström (2007). Do antitrust laws facilitate collusion? Experimental evidence on costly communication in duopolies. *Scandinavian Journal of Economics 109*(2), 321–339.

Andres, M., L. Bruttel, and J. Friedrichsen (2021). The leniency rule revisited: Experiments on cartel formation with open communication. *International Journal of Industrial Organization 76*(102728).

Apesteguia, J., M. Dufwenberg, and R. Selten (2007). Blowing the whistle. *Economic Theory 31*(1), 143–166.

Berlin, M., B. Qin, and G. Spagnolo (2018). Leniency, asymmetric punishment and corruption: Evidence from china. CEIS Working Paper No. 431.

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

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM 55*(4).

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Blonski, M., P. Ockenfels, and G. Spagnolo (2011). Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Economic Journal: Microeconomics 3*(3), 164–192.

Blume, A. and P. Heidhues (2008). Modeling tacit collusion in auctions. *Journal of Institutional and Theoretical Economics*, 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., D. J. Cooper, and C. Rott (2019). Communication in laboratory experiments. In *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing.

Buccirossi, P., L. Ciari, T. Duso, G. Spagnolo, and C. Vitale (2013). Competition policy and productivity growth: An empirical assessment. *Review of Economics and Statistics 95*(4), 1324–1336.

Chang, J., S. Gerrish, C. Wang, J. Boyd-Graber, and D. M. Blei (2009). Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*, pp. 288–296.

Clark, R. and J.-F. Houde (2014). The effect of explicit communication on pricing: Evidence from the collapse of a gasoline cartel. *The Journal of Industrial Economics 62*(2), 191–228.

Coffman, L. and P. Niehaus (2015). Pathways of persuasion. Mimeo.

Cooper, D. J. and J. H. Kagel (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review 95*(3), 477–509.

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

Dal Bó, P. and G. R. Fréchette (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review 101*(1), 411–29.

Davis, D. D. and C. A. Holt (1998). Conspiracies and secret discounts in laboratory markets. *The Economic Journal 108*(448), 736–756.

Dijkstra, P. T., M. A. Haan, and L. Schoonbeek (2021). Leniency programs and the design of antitrust: experimental evidence with free-form communication. *Review of Industrial Organization 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 Journal of the European Union 101*(1). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012E/TXT&from=DE Accessed: 2020-01-23.

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.

Feinerer, I., K. Hornik, and D. Meyer (2008). Text mining infrastructure in R. *Journal of Statistical Software 5*(5), 1–54.

Fink, N., P. Schmidt-Dengler, K. Stahl, and C. Zulehner (2017). Registered cartels in austria: An overview. *European Journal of Law and Economics 44*(3), 385–422.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischer, C. and H.-T. Normann (2019). Collusion and bargaining in asymmetric Cournot duopoly–An experiment. *European Economic Review 111*(2), 360–379.

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

Fourberg, N. (2018). Let's lock them in: Collusion under consumer switching costs. DICE Discussion Paper No. 296.

Garrod, L. and M. Olczak (2018). Explicit vs tacit collusion: The effects of firm numbers and asymmetries. *International Journal of Industrial Organization 56*, 1–25.

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

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–574.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America 101*(Suppl. 1), 5228–5235.

Griffiths, T. L. and M. Steyvers (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis 427*(7), 424–440.

Hansen, S., M. McMahon, and A. Prat (2017). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economic 133*(2), 801–870.

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

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

Hinloopen, J. and S. Onderstal (2014). Going once, going twice, reported! Cartel activity and the effectiveness of antitrust policies in experimental auctions. *European Economic Review 70*, 317–336.

Hornik, K. and B. Grün (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software 40*(13), 1–30.

Houser, D. and E. Xiao (2011). Classification of natural language messages using a coordination game. *Experimental Economics 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.

Hyytinen, A., F. Steen, and O. Toivanen (2018). Cartels uncovered. *American Economic Journal: Microeconomics 10*(4), 190–222.

Hyytinen, A., F. Steen, and O. Toivanen (2019). An anatomy of cartel contracts. *The Economic Journal 129*(621), 2155–2191.

Imai, K., K. Luke, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological Methods 15*(4), 309–334.

Imai, K., K. Luke, D. Tingley, and T. Yamamoto (2011). Unpacking the black box: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review 105*(4), 765–789.

Imai, K., D. Tingley, and T. Yamamoto (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 176*(1), 5–51.

Isaac, R. M., V. Ramey, and A. W. Williams (1984). The effects of market organization on conspiracies in restraint of trade. *Journal of Economic Behavior & Organization 5*(2), 191–222.

Moellers, C., H.-T. Normann, and C. M. Snyder (2017). Communication in vertical markets: Experimental evidence. *International Journal of Industrial Organization 50*, 214–258.

Mueller, H. and C. Rauh (2018). Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review 112*(2), 358–375.

Newman, D., A. Asuncion, P. Smyth, and M. Welling (2009). Distributed algorithms for topic models. *Journal of Machine Learning Research 10*, 1801–18028.

Normann, H.-T. and E. S. Tan (2014). Effects of different cartel policies: evidence from the german power-cable industry. *Industrial and Corporate Change 23*(4), 1037–1057.

Odenkirchen, J. (2018). Pricing behavior in partial cartels. DICE Discussion Paper No. 299.

Odudu, O. (2010). The wider concerns of competition law. *Oxford Journal of Legal Studies 30*(3), 599–613.

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

Özkes, A. and N. Hanaki (2020, March). 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. *Experimental Economics 22*(4), 1002–1029.

Rauh, C. (2019). Measuring uncertainty at the regional level using newspaper text.

Whinston, M. D. (2008). Lectures on antitrust economics. *MIT Press Books 1*.

Whish, R. and D. Bailey (2015). *Competition law*, Volume 8. Oxford University Press.

Appendix

A Theoretical background

In this section, we derive the critical discount factors for a repeated game that provides the background for our experimental design.

A.1 Modeling framework

In the experiment, participants interact in groups, consisting of a market of three firms and one competition authority when sanctioning institutions are present. The interaction between the firms is characterized by Bertrand competition with differentiated products. The same firms play the following stage game repeatedly.

Stage game: We let the quantity sold by each firm i given its own price p_i and the prices of its two competitors j and k, p_j and p_k, be given by:

$$Q_i[p_i, p_j, p_k] = 40 - \frac{100}{9}p_i + \frac{80}{9}(p_j + p_k), \tag{1}$$

where firms may choose only integer prices so that $p_i, p_j, p_k \in \mathbb{N}_0$.

Per period profit for each firm is computed as $(p_i - c)Q_i$ where c is the unit cost of production that we normalize to zero for simplicity. Then firm i's profit as a function of its own and the competitors' prices is given by:

$$\Pi_i[p_i, p_j, p_k] = 40p_i - \frac{100}{9}p_i^2 + \frac{80}{9}p_i(p_j + p_k) \tag{2}$$

Deriving the individual best-response functions and solving for the symmetric Nash equilibrium yields $p = 3$ as the equilibrium price of the stage game with a corresponding per firm profit of $\Pi = 100$. If we instead consider the maximization of joint profits, we find a symmetric joint profit maximizing price of $p = 9$, which yields a per firm profit of 180. Given that the other two firms choose a price of $p = 9$, the optimal unilateral undercutting price is $p = 5$. Deviating to $p = 5$ yields a deviation profit of 322.22 (rounded to 322 in the profit table of the experiment). The other two firms that continue to charge the collusive price $p = 9$ make a profit of only 20 in the respective period.

For the implementation in the laboratory experiment, we restrict the price setting range to the integers from 0 to 12. All prices above 12 are at least weakly dominated by those prices in the restricted range. Thus, this only helps to simplify the experiment.

Investigations and fines: A cartel can be detected and fined during its existence and after its end. In each round an investigation of the competition authority is launched with an exogenous probability of 10% or because a firm self-reported its cartel. If an investigation is launched an existing or past cartel is detected and fined with certainty.

A cartel member is fined based on its cumulative profits during the participation in a collusive agreement as judged by the competition authority. Past profits can, however, only to some extent be reduced by a fine. For the computation of the cumulative profits on which the fine is applied, profits from period t are taken into account with 100%, profits from period t-1 with 80%, profits from period t-2 with 60%, profits from period t-3 with 40%, and profits from period t-4 with 20%. Profits from period t-5 or earlier are only relevant for the computation of a potential fine (chosen by the authorities and the expert as 0%, 50% or 100% that will be applied to the cumulative profits), but the fine is not applied to these profits. This ensures that fine sizes in our setup correspond approximately to the magnitude of real cartel cases.

The experimental program does not know in which rounds a cartel existed because the authorities are only asked to evaluate for how many rounds since the last investigation a cartel existed but do not specify the rounds. Therefore, the program uses the following approximation: Based on the cartel duration as specified by the authority and the number of rounds that passed since the last investigation, the program computes an adjustment factor in the form of the percentage of rounds since the last investigation during which a cartel existed. This factor is then multiplied with the discounted cumulative profits from the five rounds preceding the investigation as detailed above. In the case where firms either always collude or always compete, the program yields exactly the fines specified above.

Feedback, fines, punishment of deviations: We assume that a deviation from a cartel is detected by the other firms immediately due to the complete feedback about each firm's price setting. Expected fines are increasing during the first five rounds of each cartel phase. For the computations that relate to perfectly collusive behavior, we focus on the maximum fine that a firm would incur from the optimal collusive agreement, i.e., when all members always set the joint profit-maximizing price resulting in per-period-per-firm profits of $\Pi^c = 180$. Then, using the linear depreciation of fine-relevant profits as introduced above, the fine in an infinitely repeated game when this cartels is detected equals $F = 540$ for each participating firm. We assume that deviations from the collusive agreement as well as reports will be punished by playing the Nash equilibrium of the stage game forever after.

Repetition: Suppose that time is discrete and that the stage game is repeated infinitely often with the participants discounting future payoffs with a discount factor δ.[39] For the analysis of the repeated game, we restrict attention to the following set of stage game payoffs: the payoff from the Nash equilibrium in the stage game, $\Pi^n = 100$, the payoff from the joint-profit-maximizing price in the stage game (the collusive or cartel payoff), $\Pi^c = 180$, the deviation payoff that is made from an optimal unilateral deviation from the collusive agreement, $\Pi^d = 322$, and the payoff that is made by the remaining cartel members when one member deviates, $\Pi^b = 20$. It holds that $\Pi^b < \Pi^n < \Pi^c < \Pi^d$.

A.2 Participation and incentive compatibility constraints

Firms will only choose the collusive equilibrium if this will yield a greater payoff than playing the Nash equilibrium. Furthermore, in a collusive equilibrium, it does not pay for any firm to deviate unilaterally in any round. In this subsection, we investigate these conditions for both treatments.

A.2.1 Collusion without sanctions

Participation constraints without sanctions: First, consider the setting without sanctioning institutions (corresponding to the treatment NoSanction). The participation constraint without sanctions for collusion reads as

$$\frac{\Pi^c - \Pi^n}{1-\delta} > 0. \tag{3}$$

With the parameters in the experiment, this is clearly fulfilled because $\Pi^n < \Pi^c$.

Incentive compatibility without sanctions: Next, consider the incentive compatibility constraint of collusion without sanctioning institutions. The value of the strategy "sticking to the collusive agreement", i.e., of setting each period the joint-profit-maximizing price is:

$$V^c = \frac{\Pi^c}{1-\delta}. \tag{4}$$

Consider now the possibility of deviating from the collusive agreement. Any such deviation is immediately observed by the cartel members (there is feedback on all prices

[39] We restrict attention to a standard stationary repeated game because we see our experimental design as one way to bring the repeated game to the laboratory even though it diverges from theory in certain aspects.

set in a period, making it easy to observe the deviation). We assume that a deviation is punished by reverting to the Nash equilibrium of the stage game forever after. The value from deviating once and being punished is

$$V^d = \Pi^d + \delta \frac{\Pi^n}{1-\delta}. \tag{5}$$

Thus, the incentive compatibility constraint for collusion without sanctioning institutions is

$$\frac{\Pi^c}{1-\delta} > \Pi^d + \delta \frac{\Pi^n}{1-\delta}. \tag{6}$$

From this constraint, we compute the critical discount factor $\delta_{NoS} = 0.6396$ which determines the range of discount factors for which, given all the other parameters in our experiment, collusion can be sustained as an equilibrium. As the continuation probability of 2/3 in our experiment exceeds the critical discount factor, collusion is a subgame perfect Nash equilibrium of the continuation game starting in round 25 and, therefore, also of the entire repeated game.

A.2.2 Collusion with sanctions

Participation constraints with sanctions: Second, consider the participation constraint for collusion with sanctions (corresponding to the treatment SANCTION). This reads in both the leniency and the no-leniency setting as

$$\frac{\Pi^c - \Pi^n}{1-\delta} > \frac{\alpha F}{1-\delta} \tag{7}$$

With the parameters in the experiment, this is clearly fulfilled because $80 > 54$. Next, consider the incentive compatibility constraints of collusion.

Incentive compatibility without a leniency rule: Without a leniency rule, the value of the strategy "sticking to the collusive agreement", i.e., setting each period the joint-profit-maximizing price and doing so even if the cartel has been detected through the exogenous detection mechanism, is:

$$V^c = \frac{\Pi^c + \alpha(\delta V^c - F)}{1-(1-\alpha)\delta} = \frac{\Pi^c + \alpha\delta V^c - \alpha F}{1-(1-\alpha)\delta} \tag{8}$$

Solving for V^c this yields

$$V^c = \frac{\Pi^c - \alpha F}{1 - \delta} \tag{9}$$

We assume that as part of the strategy "sticking to the collusive agreement" cartel members continue to collude if their cartel has been detected due to an investigation that was triggered by the exogenous detection probability. This implies that their cartel continues to exist after such an investigation; and it also continues to face the exogenous risk of being detected and fined in every single period.

Consider now the possibility of deviating from the collusive agreement. Any such deviation is immediately observed by the cartel members (there is feedback on all prices set in a period, making it easy to observe the deviation). We assume that a deviation is punished by reverting to the Nash equilibrium of the stage game forever after. The value from deviating once and being punished is

$$V^d = \Pi^d + \delta \frac{\Pi^n}{1 - \delta} - \frac{\alpha F}{1 - (1 - \alpha)\delta} \tag{10}$$

The third term results from the possibility of a cartel being detected and fined with exogenous probability also after it has broken down. As the cartel is assumed to never reform, the cartel can only be detected once after the deviation.

The incentive compatibility constraint in a setting without leniency (our treatment named Fine) is therefore

$$\frac{\Pi^c - \alpha F}{1 - \delta} > \Pi^d + \delta \frac{\Pi^n}{1 - \delta} - \frac{\alpha F}{1 - (1 - \alpha)\delta} \tag{11}$$

From this constraint, we compute the critical discount factor which determines the range of discount factors for which, given all the other parameters in our experiment, collusion can be sustained as an equilibrium.

Solving the above constraint for δ, we obtain a quadratic equation which has only one solution that lies in the interval $[0, 1]$ and therefore has a unique admissible solution $\delta_N = 0.6827$.

Incentive compatibility with a leniency rule: Consider now a setting with a leniency rule, i.e., the first firm that self-reports a collusive agreement is exempt from paying a fine. This implies that any deviation from the collusive agreement is coupled with a self-report in order to pre-empt the other firms that would report the cartel once they learn about the deviation. Thus, the value from defecting from the collusive agreement becomes:

$$V^d = \Pi^d + \delta \frac{\Pi^n}{1-\delta} \tag{12}$$

Reporting the cartel leads to an immediate fine to the other cartel members but not the self-reporting deviator. Moreover, the self-report implies that the cartel, which is assumed not to be reformed because of the Nash reversion punishment, does not face any detection risk in the future.

Thus, the incentive compatibility constraint in a setting with a leniency rule (named LENIENCY) is

$$\frac{\Pi^c - \alpha F}{1-\delta} > \Pi^d + \delta \frac{\Pi^n}{1-\delta} \tag{13}$$

From this constraint, we also compute the critical discount factor given all other parameters. Setting the above incentive constraint to bind and solving for δ, we obtain the unique solution $\delta_L = 0.8829$.

Incentive compatibility with sanctions: The above shows that the critical discount factor of an infinitely repeated discounted game with punishment by Nash reversion exceeds 2/3 in the cases with and without leniency. Thus, collusion on the symmetric joint-profit maximizing price of the stage game is not a subgame perfect Nash equilibrium (SPNE) of the continuation game starting in round 25 in the presence of sanctions and, therefore, also not of the entire repeated game, neither in FINE nor in LENIENCY.

A.2.3 Discussion

We have shown above that joint-profit maximizing collusion is not sustainable as SPNE when we have sanctions in place because it fails to satisfy the non-deviation constraint in the continuation game. Only a continuation probability larger than 88.3 percent would make the collusive agreement a subgame perfect Nash equilibrium for the continuation game, in which case the expected duration of the experiment would exceed three hours. Therefore, we opted for a lower continuation probability, which is below the critical level for both treatments with sanctions. There are two main reasons why we nevertheless expect a substantial amount of collusion in the SANCTION treatments. First, also a random continuation probability of 2/3 blurs the end of the experiment and serves the purpose to minimize endgame effects. Second, there is evidence from infinitely repeated prisoner's dilemma games that the cooperation rate does not discretely change when the critical discount factor exceeds or falls short of the continuation probability (Blonski et al., 2011) but other aspects of the game also play a role. Indeed, for the first repeated game that subjects play, which is the relevant comparison for our experiment where subjects only play one repeated game, Dal Bó and Fréchette (2011) report clearly positive cooperation rates in a setting where this is not SPNE and no significant differences to a comparison setting where this is SPNE during the first 10 rounds of play.

We further note that the experimental setting also allows for asymmetric collusive strategies. Specifically, the three firms may alternate in choosing the prices 7 – 7 – 12 yielding an average per-period profit of 217.78 for each firm. Assuming again that any deviation will be punished by reversion to the Nash equilibrium of the stage game, the incentive compatibility constraint for this strategy yields a critical discount factor clearly below 2/3 in all treatments. Specifically, in the leniency setting, for a firm supposed to set a price of 7, the optimal unilateral deviation is $p = 5$ with a one-time deviation profit of 344.44 which – using these values in the incentive compatibility constraint (13) – yields a critical discount factor of 0.613, and for a firm supposed to set a price of 12, the optimal unilateral deviation is also $p = 5$ with a deviation profit of 233.33 which yields a critical discount factor of 0.292. The analogous critical discount factors are even lower in the setting without a leniency rule and are easily derived from the incentive compatibility constraint (11). The repeated game may have additional asymmetric equilibria that we have not identified.

In principle, collusion may occur at prices different from the jointly optimal price of 9. This will lead to lower expected profits but relaxes the incentive compatibility constraint. For the parameters of our experiment and a maximum fine in a steady state equilibrium with stable collusion of $F = 3\Pi^c$, where Π^c is the per-firm profit per period from continued collusion on the respective price, we find that a symmetric collusive agreement on an anti-competitive price below 9 – e.g. on a price of 8, 7, or 6 – yields a critical discount factor that lies below our continuation probability of 2/3 for the Fine treatment while such an agreement is still not sustainable in Leniency. The exact critical discount factors can be derived directly from the incentive compatibility constraints specified above.

B Descriptive Data

	NoSanction	Sanction	Sanction before fine	Sanction after fine
Unweighted cartelization	0.97	0.41	1.00	0.42
	(0.10)	(0.32)	(0.00)	(0.45)
Average Market Price	8.84	6.64	7.92	6.73
	(0.45)	(1.26)	(1.33)	(1.63)
Explicit Communication	0.32	0.09	0.17	0.07
	(0.14)	(0.06)	(0.32)	(0.16)
Indirect Communication	0.04	0.10	0.32	0.14
	(0.02)	(0.09)	(0.35)	(0.24)
	NoSanction	Sanction	Sanction explicit comm.	Sanction indirect comm.
Investigations	-	0.11	0.11	0.11
	-	(0.07)	(0.07)	(0.08)
Fines	-	0.56	0.83	0.43
	-	(0.41)	(0.21)	(0.44)

Table 3: Descriptive means split up by treatments. Standard deviations in parentheses.

Table 3 provides descriptive means split up by treatments. To calculate the means that refer to before a fine and after a fine, respectively, we average the variable in cartelized rounds before a fine and compare them to the variable for the following three rounds or the twenty-fifth round, whichever comes first. The explicit communication and indirect communication variable before a fine and after a fine, respectively, refer to our hand-coded data. Investigations averages over the number of investigations in a market over rounds 2 to 25. Fines averages over whether a fine applies in an investigation taking place in rounds 2 to 25, overall and in the subsamples exhibiting explicit or indirect communication.

C Comparison of treatments Fine and Leniency

In this appendix, we repeat the analysis presented in Section 5.2 on the relation between sanctioning institutions and communication, comparing the data from treatment NoSanction to Fine and Leniency separately.

The average posterior probability of the topic *Explicit Agreement* is 0.05 in Fine ($N = 23$, $SD = 0.02$) and 0.06 in Leniency ($N = 27$, $SD = 0.06$). As in the main analysis comparing Sanction to NoSanction, these separate values are significantly different from the value in NoSanction (in either test, $p < 0.001$). The average posterior probability of the topic *Explicit Reasoning* is 0.04 in Fine ($N = 23$, $SD = 0.03$) and 0.04 in Leniency ($N = 27$, $SD = 0.02$). Both values differ significantly from the value in NoSanction (in either test, $p < 0.001$).

When we consider the total amount of explicit communication by summing up the average posterior probabilities of *Explicit Agreement* and *Explicit Reasoning*, the average posterior probability of explicit communication is 0.09 in Fine ($N = 23$, $SD = 0.04$) and 0.10 in Leniency ($N = 23$, $SD = 0.07$), which is, in both cases, significantly different from the average in NoSanction ($p < 0.001$).

We further note that groups appear to use explicit communication slightly more often in Leniency than in Fine but the difference fails to reach significance at reasonable levels ($p = 0.79$). The probability of the topic *Explicit Reasoning* is statistically indistinguishable between Leniency and Fine ($p = 1$).

D Additional analyses on hand-coded communication data

D.1 Example for communication strategies

Attempt to "hide incriminating evidence in innocuous talk" (translated to English) Communication patterns in many groups appear consistent with this idea but one group spells it out explicitly:

Round 15, firm 2: "here on page 4 [in the instructions] it says that they investigate our entire communication [. . .] we can simply spam the chat as we like"

Thereafter, they indeed spam the chat, sometimes with meaningful sentences but most of the time with super short messages that contain only gibberish, e.g., "asfjha", "asfasf", "sdg", or nouns, meaningless without context. In Figure 12, the bars marking the number of messages of group 52 are censored as the spamming leads to close to 60 individuals messages per round in certain rounds.

Two attempts to "talk once to collude and remain silent thereafter" (translated to English) We see several groups for which the communication pattern suggests they are following this idea. The following two spell it out explicitly but then do not follow through.

Round	Firm	Message
2	2	All 9
3	1	Say nothing and do something else afterwards
3	3	I thought so too
4	1	Lol that was perfect
4	1	But one can do better
4	2	again all a little higher the same
4	2	I also join in
4	1	Say nothing..
5	1	you could take the better but ok
5	1	there is a super price, let's try?
Communication in further rounds omitted here.		

Table 4: Excerpt from chat protocol of group 3 in treatment Leniency. See also Figure 12.

Round	Firm	Message
3	1	And?
3	2	The profit is too small for me and you?
3	1	higher!
5	2	Rightly firm 1 that was a rotten number
6	2	Who initiated the last audit?
6	1	nobody
6	1	Coincidence
7	2	No further agreements ;)
7	1	the educating effect of an audit ;)
9	2	was coincidence again or?
Communication in further rounds omitted here.		

Table 5: Excerpt from chat protocol of group 1 in treatment Leniency. See also Figure 12.

D.2 Figures with communication and outcomes over time at the group-level

Figures 11, 12, and 13 depict the development of communication, cartelization, and fines at the market level over time. We depict the total number of messages sent in a group in a given round in thin dark gray bars; these are measured on the right y-axis, ranging from 0 to 25 per round.[40] In addition, the figures show the share of indirect messages (black dashed) and the share of explicit messages (solid black) in the number of total messages per group per round; the scale for these lines is on the left y-axis. We further include the extent of cartelization as derived from the weighted expert measure in wide light gray bars. Lastly, we include markers for the average fines that was decided on by the actual authorities in the experimental sessions for each group in those rounds, where an investigation took place; triangles indicate that the fine resulted from an investigation triggered by a self-report and circles indicate that the fine resulted from a random investigation.

[40]For group 52 in treatment Leniency, this variable is censored at 25 but the true values reach up to close to 60 messages per round. See also discussion on spam strategy in previous subsection.

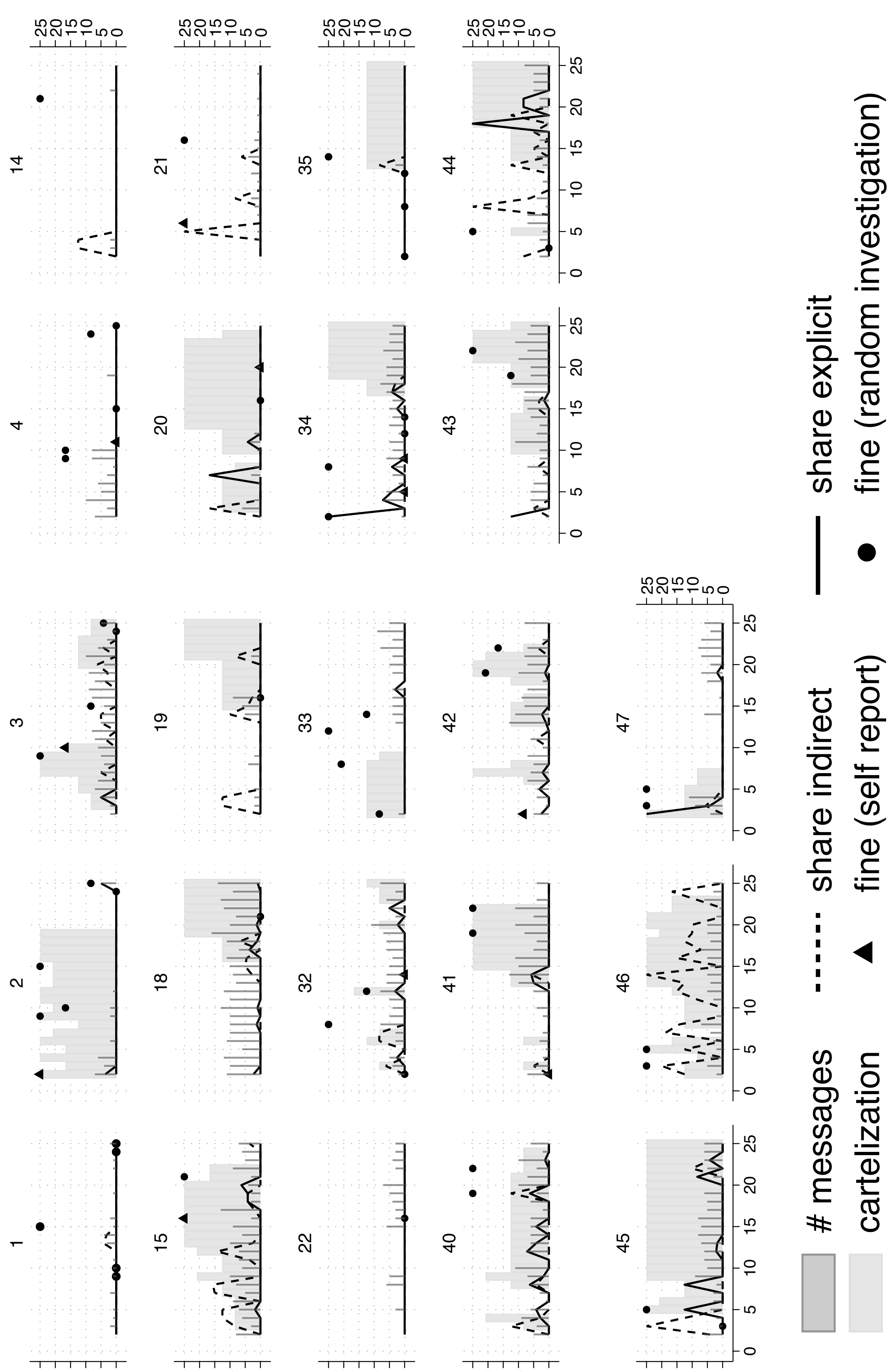


Figure 11: Communication and cartelization over time in FINE treatment.

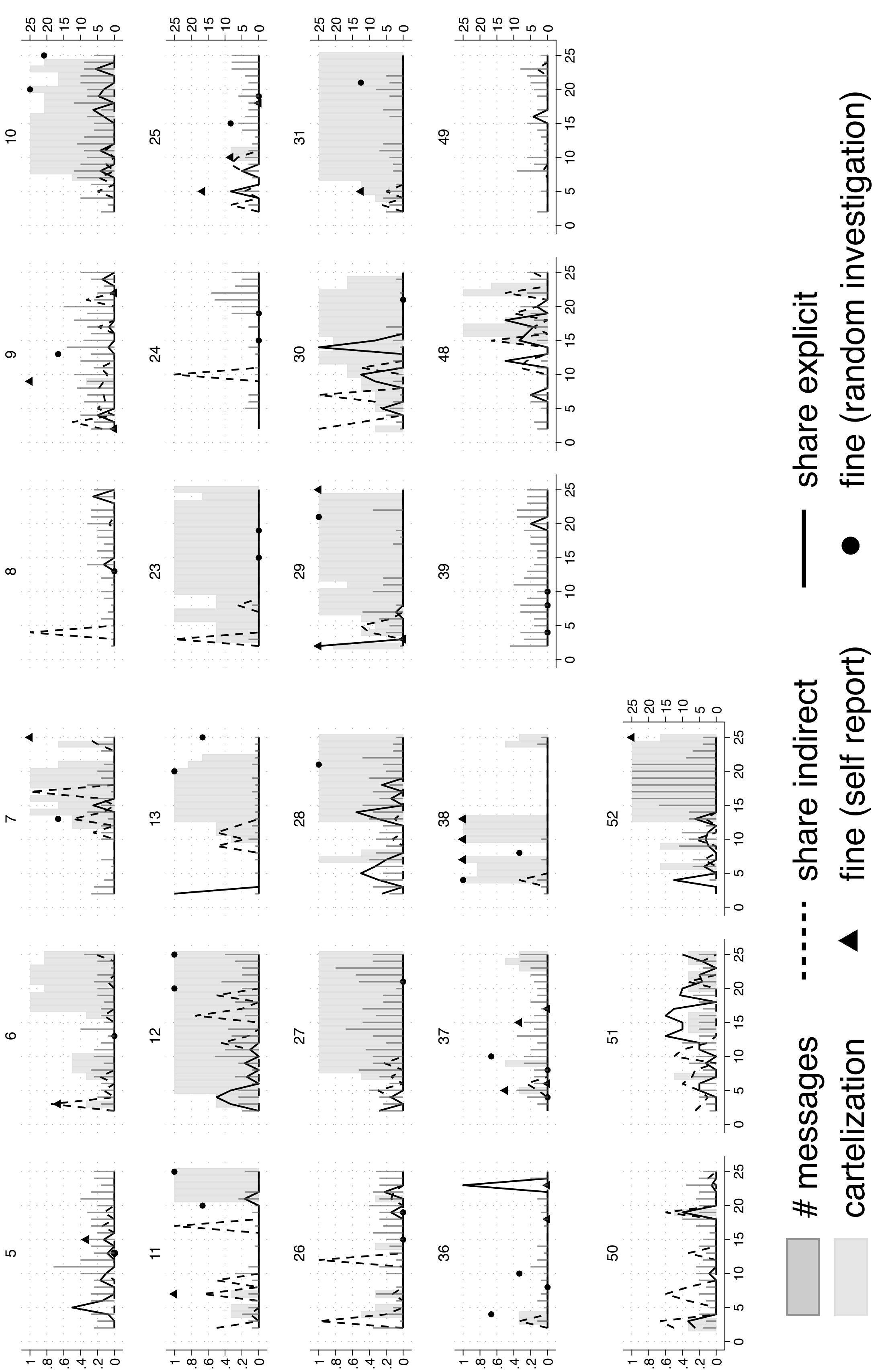


Figure 12: Communication and cartelization over time in Leniency treatment.

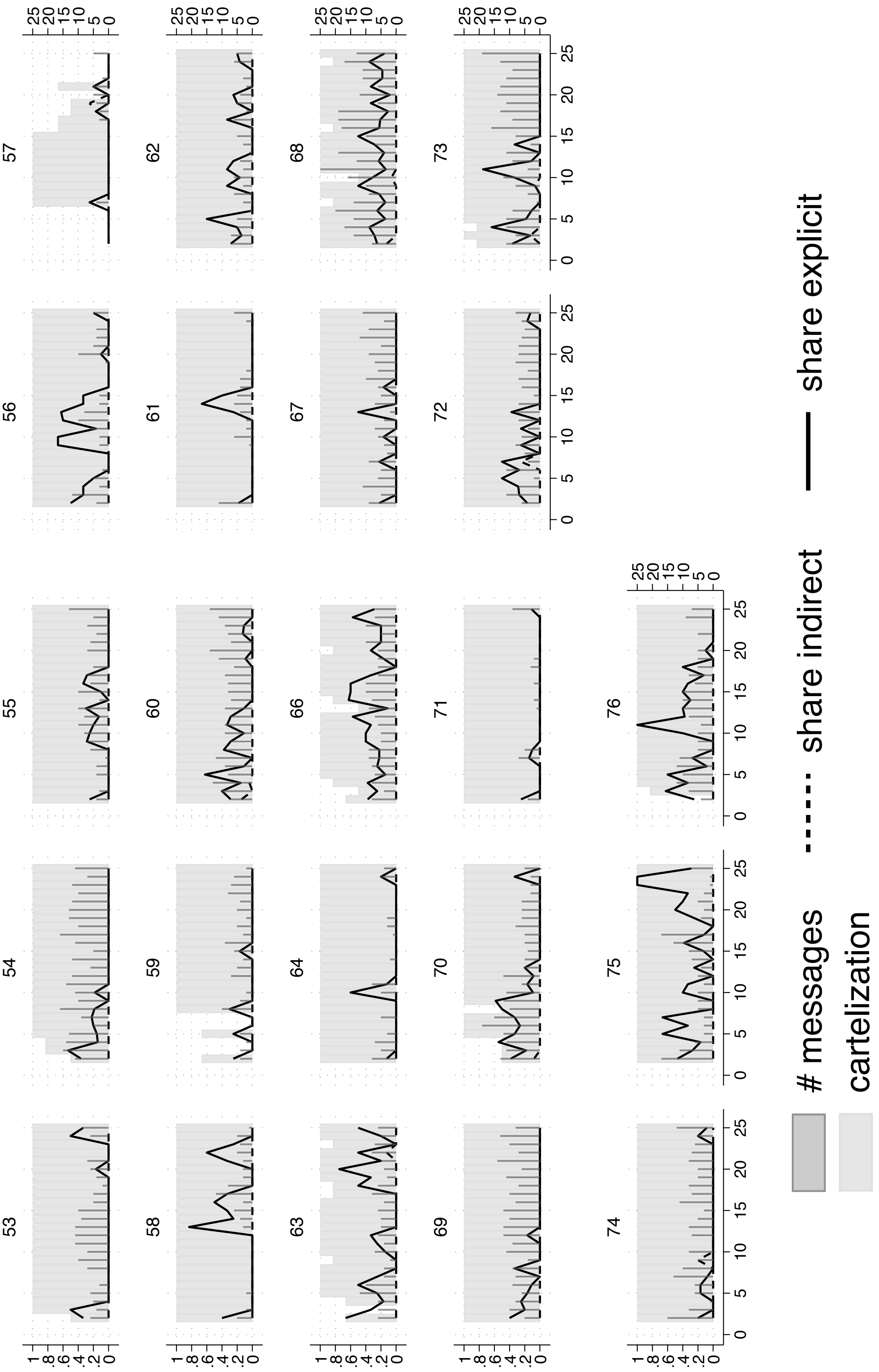


Figure 13: Communication and cartelization over time in NoSanction treatment.

E Instructions

In the following, we present our instructions for firms in Section E.1 and for authorities in Section E.2. Parts that appear only in the instructions of a particular treatment are clearly marked as such. Text in *italics* only appears in instructions for the LENIENCY treatment. The original instructions for the participants additionally included screen-shots of the different stages in the experiment.

E.1 Instructions for firms

Today you are participating in a decision-making experiment. If you read the following instructions carefully, you can earn money. The amount of money you receive depends on your decisions and the decisions of other participants.

For the entire duration of the experiment it is prohibited to communicate with other participants. Therefore, we ask you not to talk to each other. Violation of this rule will result in exclusion from the experiment and payment.

If there is something you do not understand, please have another look at these instructions or give us a hand signal. We will then come to your seat and answer your question personally.

During the experiment, we do not talk of euro but of points. The number of points you earn during the experiment will be converted into euro as follows:

125 Points = 1 euro

At the end of today's experiment, you will receive the points earned in the experiment converted into euro in **cash** plus 5 euro as basic endowment.

On the following pages we will explain the exact procedure of the experiment to you, starting with the general procedure. We will then familiarize you with the procedure on the screen. Then, you will have the opportunity to familiarize yourself on the computer screen with the calculation of profits in the experiment before the experiment begins.

The experiment

At the start of the experiment, you will be matched randomly into a group with two [**Fine and Leniency:** three] other participants. During the experiment, you will make decisions within this group of three [**Fine and Leniency:** four] persons in total. The composition of your group remains the same throughout the entire experiment. Neither you nor the other participants will be informed about the identity of the participants in the group – neither during nor after the experiment.

The experiment consists of at least 25 rounds. You will receive more information on the number of rounds on page 5 of this document.

[**NoSanction only:** Every participant in your group represents a firm. There are three firms (firm 1, 2 and 3). At the start of the experiment, you will be informed onscreen about which firm you are. You will be the same firm during the entire experiment.]

[**Fine and Leniency only:** Every participant in your group represents either a firm or the competition authority. There are three firms (firm 1, 2 and 3) and one competition authority. **In all rounds, you take the role of a firm.** At the start of the experiment, you will be informed onscreen about which firm you are. You will be the same firm during the entire experiment.]

The firms 1, 2 and 3 sell the same (fictional) good on the same market. Production of this good is costless for the firms. All firms decide simultaneously what price they want to charge for the good in a round. The price must be an integer between 0 and 12. If a firm does not enter its own price and clicks the OK button within 30 seconds (60 seconds in the first round only), a price of 0 is automatically set for this firm.

Your profit depends on your own price and the average price of the other two firms. Your profit is larger the higher the prices of the other two firms are. Your own price has two effects on your own profit: If you increase your own price, the quantity you sell decreases, but at the same time your earnings per unit sold increases. Depending on which effect is larger, your profit increases or decreases. The table on the following page shows your profit, depending on your own price and the averages prices of the other two firms. (This table is the same for all three firms, read from their perspective.)

		Average price of the other two firms												
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Your own price	**0**	**0**	0	0	0	0	0	0	0	0	0	0	0	0
	1	29	**38**	47	56	64	73	82	91	100	109	118	127	136
	2	36	53	**71**	89	107	124	142	160	178	196	213	231	249
	3	20	47	73	**100**	127	153	180	207	233	260	287	313	340
	4	0	18	53	89	**124**	160	196	231	267	302	338	373	409
	5	0	0	11	56	100	**144**	189	233	278	322	367	411	456
	6	0	0	0	0	53	107	**160**	213	267	320	373	427	480
	7	0	0	0	0	0	47	109	**171**	233	296	358	420	482
	8	0	0	0	0	0	0	36	107	**178**	249	320	391	462
	9	0	0	0	0	0	0	0	20	100	**180**	260	340	420
	10	0	0	0	0	0	0	0	0	0	89	**178**	267	356
	11	0	0	0	0	0	0	0	0	0	0	73	**171**	269
	12	0	0	0	0	0	0	0	0	0	0	0	53	**160**

From the second round on, you have the option to communicate with the other firms via chat messages at the beginning of each round. The duration of a chat cannot exceed 60 seconds in one round. In this chat, you can write anything you want with the exception that you are not allowed to reveal hints on your identity.

[**Fine and Leniency only:** §1 GWB of the Act against Restraints of Competition prohibits price agreements and the attempt of price agreements (for the exact wording, see the box).

§ 1 Prohibition of Agreements Restricting Competition
Agreements between undertakings, decisions by associations of undertakings and coordinated practices which have as their object or effect the prevention, restriction or distortion of competition are prohibited.

At the end of a round, the chat messages can be subject to an audit. In an audit, the competition authority judges whether the texts you and the other firms wrote in the chat are in accordance with §1 GWB. Such an audit can be initiated in two ways, by a random mechanism and by the firms:

- In each round, a random mechanism decides whether an audit takes place or not. This random mechanism is programmed so that an audit takes place with a probability of 10% (i.e. on average in 10 out of 100 cases).
- In addition, in each round the firms have the opportunity to initiate an audit themselves, both while setting their price and after they have learned the prices of the other firms. You can initiate an audit by clicking on a small white box at the bottom left of the screen. Initiating an audit cannot be undone. As soon as you click on the small white box, the box for that round disappears and an audit will definitely take place. The same applies to the other two firms in your group.

When an audit takes place, the competition authority has insight into all communication in the previous chats in your group as well as into the pricing since the first round. The competition authority imposes penalties on firms that have violated §1 GWB. It decides on the individual penalties for each of the three firms and for how long an agreement has been in place.

The penalty may be 0%, 50% or 100% of a firm's accumulated pecuniary profit during the agreement. 0% (no penalty) means that the firm has acted in accordance with §1 GWB, 100% means a clear, serious violation. 50% should be chosen for less serious violations.

The pecuniary profit is measured according to your profit that you have earned and the duration of the agreement. However, if the agreement has been in place for more than five rounds, the penalty will only be applied to the profits of the last five rounds. Previous rounds are included in the calculation of the penalty, but will not be punished themselves.

The competition authority has three minutes to reach its decision.]

[**Leniency only:** *The active initiation of an audit by a firm leads to the possibility that that firm is exempted from punishment. If only one firm has initiated the audit, that firm will automatically receive full amnesty. If two or three firms have initiated an audit, the penalty will only be waived for the firm that first initiated the audit.*]

[**NoSanction only:** After each round, the firms are informed about their own price and their profit. In addition, each firm is informed about the prices set by the other two firms in the current round.]

[**Fine and Leniency only:** After each round, the firms are informed about their own price, their profit and, if applicable, their penalty. In addition, each firm is informed about the prices set by the other two firms in the current round and, if applicable, their penalties. [Leniency only: *You will also be informed on whether a firm has initialized an audit by the competition authority and has thus received an exemption of its penalty.*]]

From the 25th round on, a random mechanism decides in each round whether the experiment ends with the last round completed. With a probability of 33.3% (i.e. in an average of 1 out of 3 cases) the experiment ends with the last round completed. With a probability of 66.7% (i.e. in 2 out of 3 cases) another round takes place. In addition, it is ensured that the experiment does not last longer than 2 hours and 30 minutes.

After the last round, you will see an overview screen showing you how many points you have earned in total. You will receive all points converted into euro directly after the experiment.

If something is not clear to you, please give a clear hand signal. We will then come to your seat.

After the experiment we will ask you to fill out a short questionnaire on the computer. You will then receive your payment.

E.2 Instructions for authorities (Fine and Leniency only)

Today you are participating in a decision-making experiment. If you read the following instructions carefully, you can earn money. The amount of money you receive depends on your decisions.

For the entire duration of the experiment it is prohibited to communicate with other participants. Therefore, we ask you not to talk to each other. Violation of this rule will result in exclusion from the experiment and payment.

If there is something you do not understand, please have another look at these instructions or give us a hand signal. We will then come to your seat and answer your question personally.

During the experiment, we do not talk of Euro but of points. The number of points you earn during the experiment will be converted into Euro as follows:

125 Points = 1 euro

As an exception, this time you will not receive your payment for today's experiment in cash at the end of the experiment, but in about 2-3 weeks via bank transfer. You will receive more information on the bank transfer on page 6 of these instructions. In addition to your other earnings in this experiment, you will receive 10 euro in cash.

On the following pages we will explain the exact procedure of the experiment to you, starting with the general procedure. We will then familiarize you with the procedure on the screen. Then, you will have the opportunity to familiarize yourself on the computer screen with your task in the experiment before the experiment begins.

The experiment

At the start of the experiment, you will be matched randomly into a group with three other participants. During the experiment, you will make decisions within this group of four persons in total. The composition of your group remains the same throughout the entire experiment. Neither you nor the other participants will be informed about the identity of the participants in the group – neither during nor after the experiment.

The experiment consists of at least 25 rounds. You will receive more information on the number of rounds on page 6 of this document.

Every participant in your group represents either a firm or the competition authority. There are three firms (firm 1, 2 and 3) and one competition authority. **In all rounds, you take the role of the competition authority.**

The firms 1, 2 and 3 sell the same (fictional) good on the same market. Production of this good is costless for the firms. All firms decide simultaneously what price they want to charge for the good in a round. The price must be an integer between 0 and 12. If a firm does not enter its own price and clicks the OK button within 30 seconds, a price of 0 is automatically set for this firm.

The profit of a firm depends on its own price and the average price of the other two firms. The profit is larger the higher the prices of the other two firms are. The own price has two effects on the profit of a firm. If the own price increases, the quantity sold by this firm decreases, but at the same time the earnings per unit sold increases. Depending on which effect is larger, a firm's profit increases or decreases. The table on the following page shows the profit of a firm, depending on its own price and the averages prices of the other two firms. (This table is the same for all three firms.)

		Average price of the other two firms												
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Your own price	**0**	**0**	0	0	0	0	0	0	0	0	0	0	0	0
	1	29	**38**	47	56	64	73	82	91	100	109	118	127	136
	2	36	53	**71**	89	107	124	142	160	178	196	213	231	249
	3	20	47	73	**100**	127	153	180	207	233	260	287	313	340
	4	0	18	53	89	**124**	160	196	231	267	302	338	373	409
	5	0	0	11	56	100	**144**	189	233	278	322	367	411	456
	6	0	0	0	0	53	107	**160**	213	267	320	373	427	480
	7	0	0	0	0	0	47	109	**171**	233	296	358	420	482
	8	0	0	0	0	0	0	36	107	**178**	249	320	391	462
	9	0	0	0	0	0	0	0	20	100	**180**	260	340	420
	10	0	0	0	0	0	0	0	0	0	89	**178**	267	356
	11	0	0	0	0	0	0	0	0	0	0	73	**171**	269
	12	0	0	0	0	0	0	0	0	0	0	0	53	**160**

From the second round on, the firms have the option to communicate via chat messages. The duration of chat cannot exceed 60 seconds.

§1 GWB of the Act against Restraints of Competition prohibits price agreements and the attempt of price agreements (for the exact wording, see the box).

> § 1 Prohibition of Agreements Restricting Competition
> Agreements between undertakings, decisions by associations of undertakings and coordinated practices which have as their object or effect the prevention, restriction or distortion of competition are prohibited.

At the end of a round, the chat messages can be subject to an audit. In an audit, you as the competition authority judge whether the texts the firms wrote in the chat are in accordance with §1 GWB. Such an audit can be initiated in two ways, by a random mechanism and by the firms:

- In each round, a random mechanism decides whether an audit takes place or not. This random mechanism is programmed so that an audit takes place with a probability of 10% (i.e. on average in 10 out of 100 cases).

- In addition, in each round the firms have the opportunity to initiate an audit themselves, both while setting their price and after they have learned the prices of the other firms. A firm can initiate an audit by clicking on a small box on the screen.

When an audit takes place, you will not be informed on how it was initiated. You have insight into all communication in the previous chats in your group as well as into the pricing since the first round. Your task is to impose penalties on firms that have violated §1 GWB. You decide on the individual penalties for each of the three firms and for how long an agreement has been in place. The duration is the number of all rounds since the last audit (or since the start of the experiment) in which, in your opinion, an agreement had a visible effect on the prices.

The penalty may be 0%, 50% or 100% of a firm's accumulated pecuniary profit during the agreement. 0% (no penalty) means that the firm has acted in accordance with §1 GWB, 100% means a clear, serious violation. 50% should be chosen for less serious violations.

The pecuniary profit is measured according to the profit of the respective firm and the duration of the agreement. However, if the agreement has been in place for more than five rounds, the penalty will only be applied to the profits of the last five rounds. Previous rounds are included in the calculation of the penalty, but will not be punished themselves. You, in the role of the competition authority, nevertheless enter the entire duration of the cartel; the computer program proportionally calculates the penalties for the last five rounds.

Your payment as an competition authority depends on the consistency of your penalty decisions with those of a real competition law expert. After today's experiment, in the same way as you do today, this expert (a licensed lawyer specialized in competition law) will see the chat messages and prices and will assess the extent to which they contain violations of §1 GWB. You will receive 900 points for each match between your decision and the expert's decision. You will also receive 900 points if you have correctly specified the duration of a possible agreement. Since you make four decisions for each penalty decision (one for each of the three firms and one for the total duration of the agreement), you can earn up to 3600 points. You will only receive points if you make exactly the same decision as the expert, otherwise (e.g. if you impose a 50% penalty on a firm and the expert would impose 100%) you will not receive any points for this partial decision. At the end, the **average** score of all rounds in which you were able to impose penalties is determined. This then determines your payment, which we will transfer to your bank account within 2 to 3 weeks. If there is no audit during the entire experiment, you will receive a fixed bank transfer of 15 euro in addition to your cash payment of 10 euro.

You have 3 minutes for each of your penalty decisions. If you do not specify the height of the penalty during this time, you will not receive any payment for your judgment and the computer program will assume for the calculation of the firms' profits that you have not imposed any penalties. **Please remember to submit your decision at the end by clicking the OK button.**

[**Leniency only:** *The active initiation of an audit by a firm leads to the possibility that that firm is exempted from its punishment. If only one firm has initiated the audit, that firm will automatically receive full amnesty. If two or three firms have initiated an audit, the penalty will only be waived for the firm that first initiated the audit. This exemption will also be automatically implemented by the computer program, if necessary, and will not be relevant to your penalty decisions.*]

After each round, the firms are informed about their own price, their profit and, if applicable, their penalty. In addition, each firm is informed about the prices set by the other two firms in the current round and, if applicable, their penalties. [**Leniency only:** *The firms will also be informed on whether a firm has initialized an audit by the competition authority and has thus received an exemption of its penalty.*]

From the 25th round on, a random mechanism decides in each round whether the experiment ends with the last round completed. With a probability of 33.3% (i.e. in an average of 1 out of 3 cases) the experiment ends with the last round completed. With a probability of 66.7% (i.e. in 2 out of 3 cases) another round takes place. In addition, it is ensured that the experiment does not last longer than 2 hours and 30 minutes.

Directly after the experiment you will receive 10 euro in cash. Your additional earnings from the experiment will be transferred to your bank account. Please enter your name and address as well as your bank details in the form and sign it. (You are welcome to fill in the form during the experiment, if you have nothing to do on the screen.)

If something is not clear to you, please give a clear hand signal. We will then come to your seat.

After the experiment we will ask you to fill out a short questionnaire on the computer. You will then receive your payment.

E.3 Assistance for participants in the role of a competition authority | How does the expert punish?

What counts as an agreement?

- If a firm explicitly suggest a price above 3 and then charges this price, the firm gets a 100% penalty.
- Convoluted descriptions of prices are punished in the same way as if the corresponding price was given as a number.
- Agreements on prices not higher than 3 do not distort competition and therefore do not count as an agreement.
- If a firm does not write anything in the chat (but of course can read what the others write) it can still be punished.[41] The amount of the penalty depends on the price and can be up to 100%, e.g. if the other two firms make a clear agreement and this firm sets exactly the price agreed by the other two firms over a long period of time.
- If the firms make an agreement that no one will abide by afterwards, there will be no penalty.
- Prices above 3, which have come about without any agreement, cannot be punished.

For determining the duration:

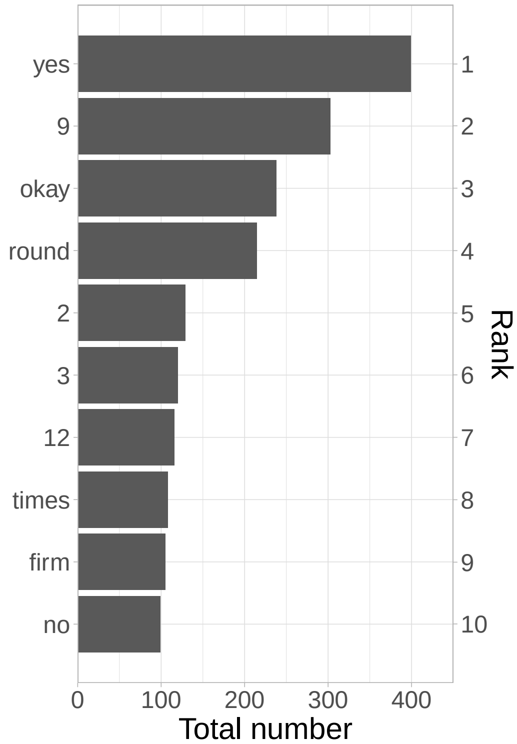
- For determining the duration of a cartel, all rounds in which the agreement was visibly effective in the prices count.
- If a company receives a 50% penalty for part of the total duration of the cartel and a 100% penalty for the remainder of the total duration, then the amount of the penalty that applies for a longer period will apply for the total duration (because the computer program does not allow for further gradation).
- If a firm joins an agreement already in place between the two other firms at a later round (or leaves the agreement earlier than the others), the longer overall duration of the cartel still applies to it. In order to prevent the fine from becoming unreasonably high, the amount of the fine can then be adjusted accordingly. (Example: Anyone who was involved in a 100% agreement in 5 out of 10 rounds receives a 50% penalty for the duration of 10 rounds.)

[41] Note that this rule follows the legal practice that a market participant who does not agree to take an expressed action but behaves as if she did, can be assumed to be part of the concerted practice (Albors-Llorens, 2006; European Union, 2019; Odudu, 2010; Whish and Bailey, 2015)

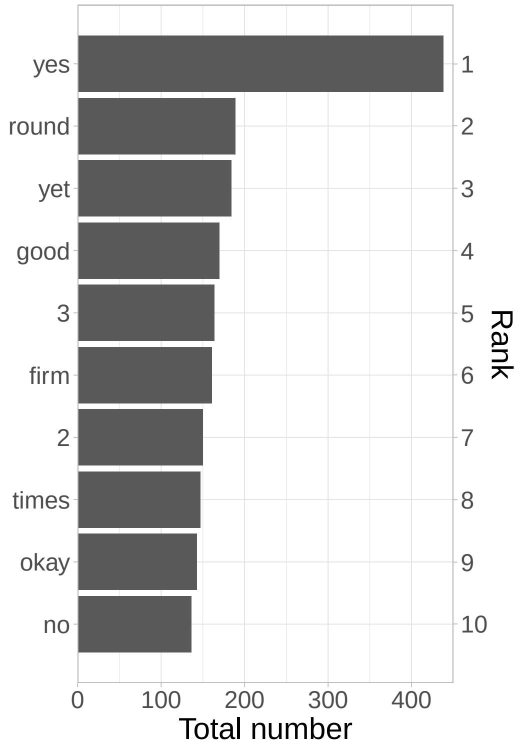
- If, after a penalty, prices remain at the same level as before the audit, a penalty may be imposed again at a later audit, even if there has been no new agreement.

F Text mining results

In this section, we present our text mining results. Figure 14 shows the token frequency per treatment. Figure 15 shows the tokens-per-topic distributions for all 25 topics and Figure 16 shows the average posterior distribution of the 25 topics by treatment using a LDA.



(a) NoSanction



(b) Sanction

Figure 14: Token frequency per treatment.

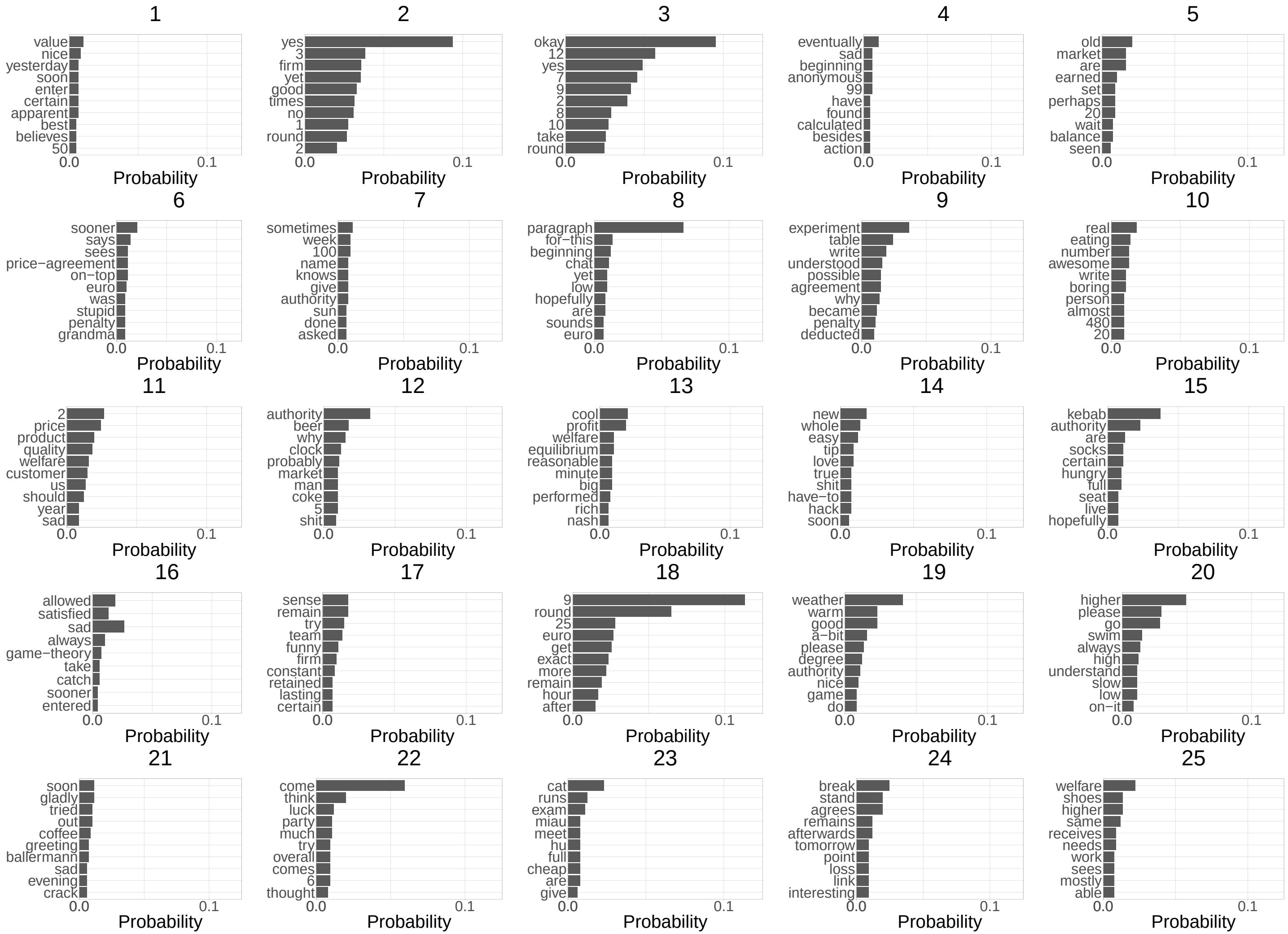


Figure 15: Token-per-Topic distributions of the top ten tokens for all 25 topics.

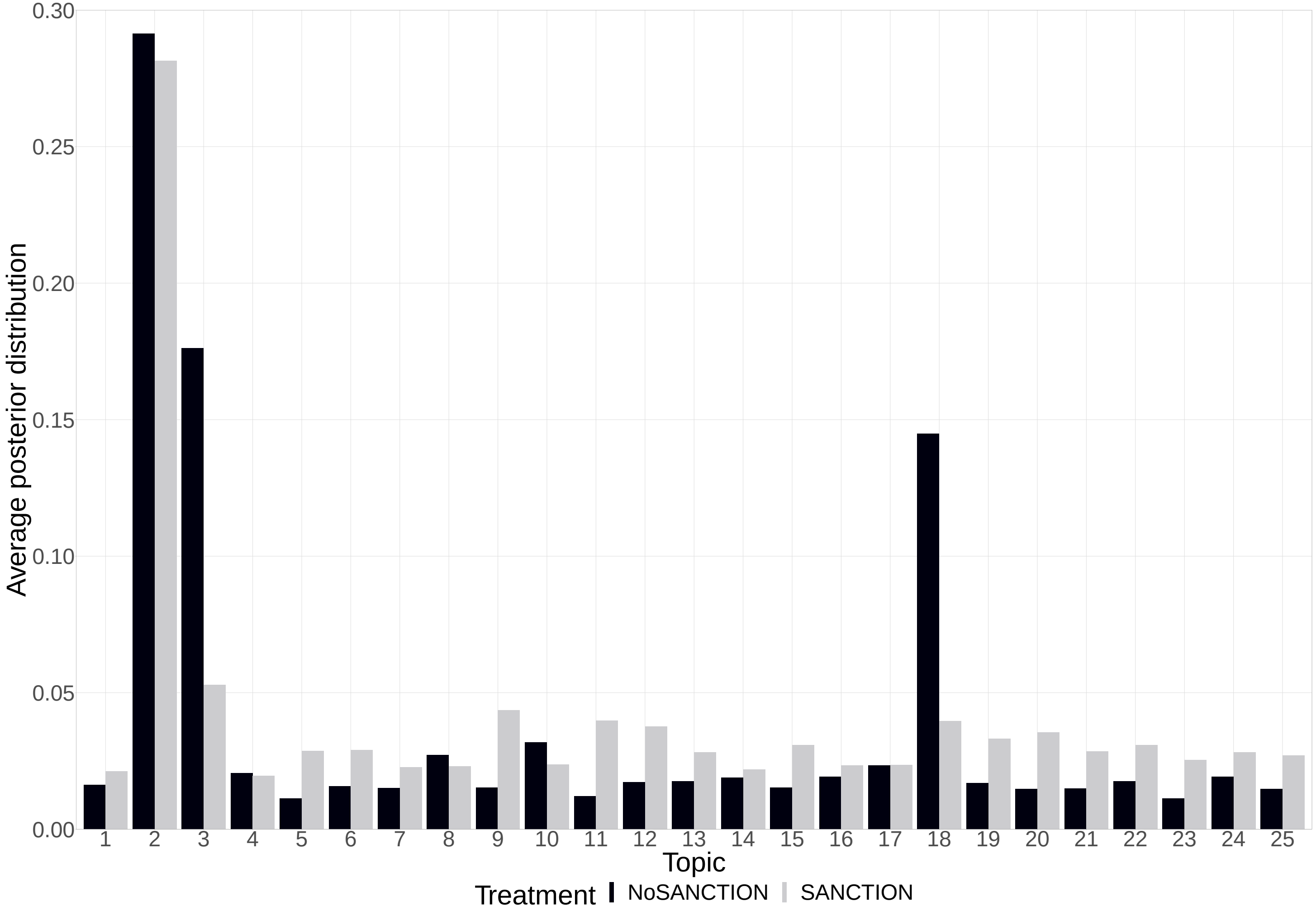
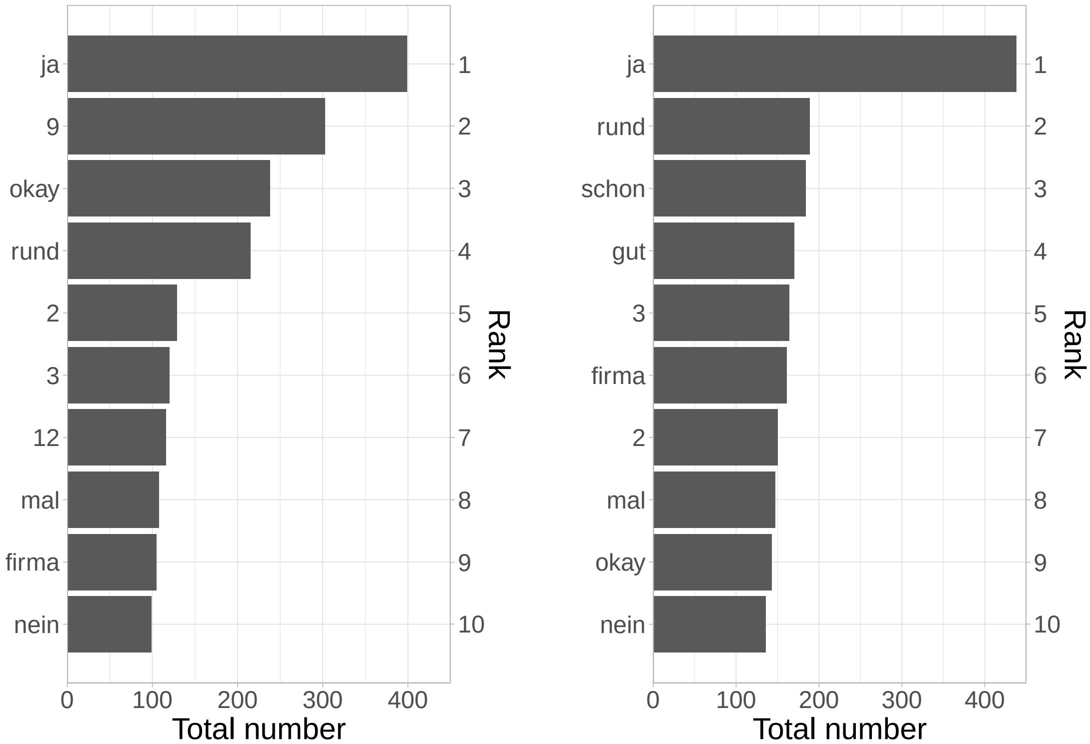


Figure 16: Average posterior distribution of the 25 topics by treatment.

G Original German tokens in their corresponding Figure

In the following, we present the original German tokens in their corresponding figure. We translated the tokens only after the analysis.[42]



(a) NoSanction (b) Sanction

Figure 17: Token frequency per treatment in German.

[42]Note for Figure 18: the German word "wohl", written in small letters, translates to "probably". Written with a capital letter, "Wohl", the word means "welfare". Both versions are used in the chats. However, the second translation better fits to the context of topic 25, *Joint Benefit*. Hence, we use the latter translation.

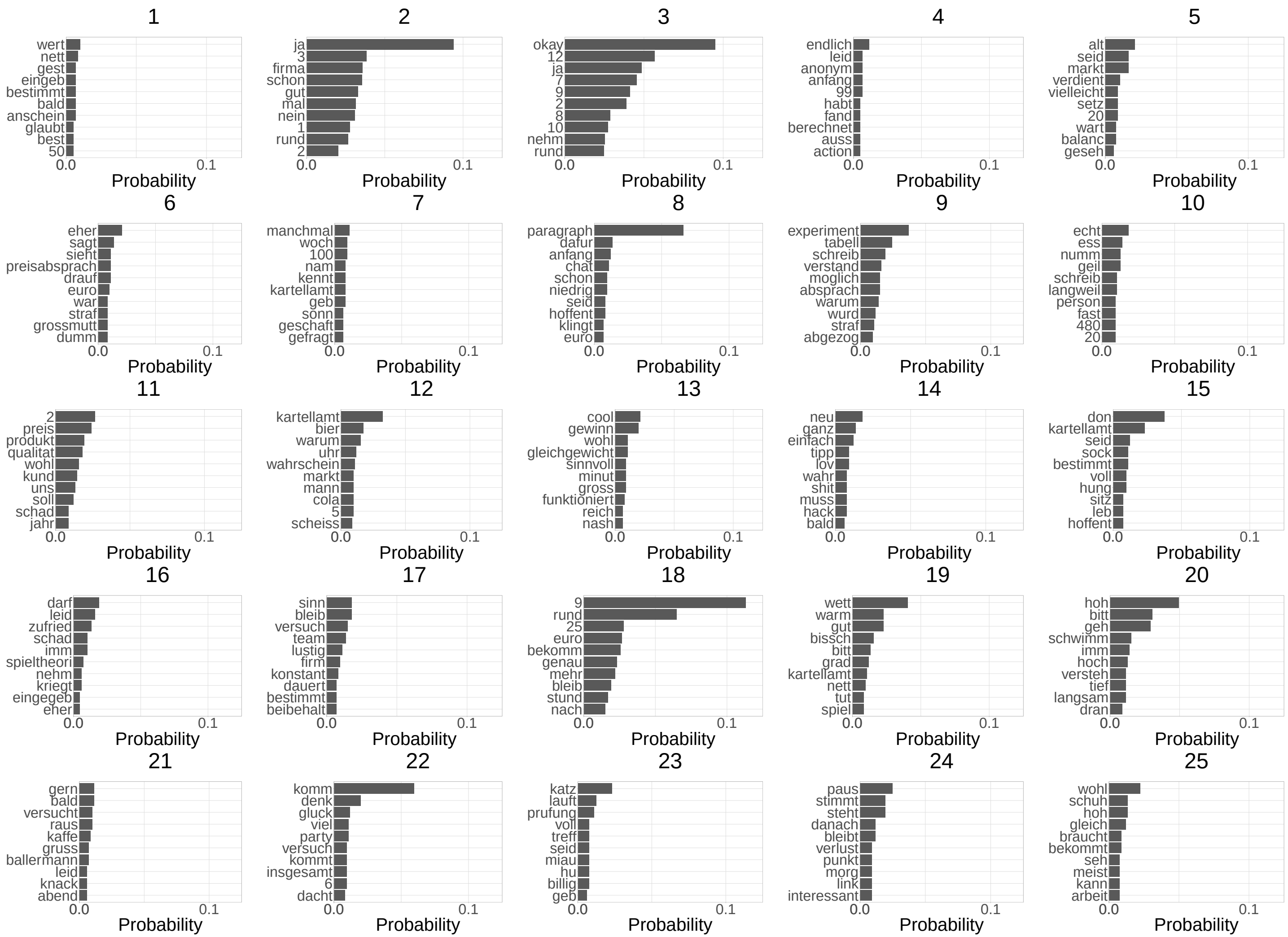


Figure 18: Token-per-Topic distributions of the top ten tokens for all 25 topics in German.

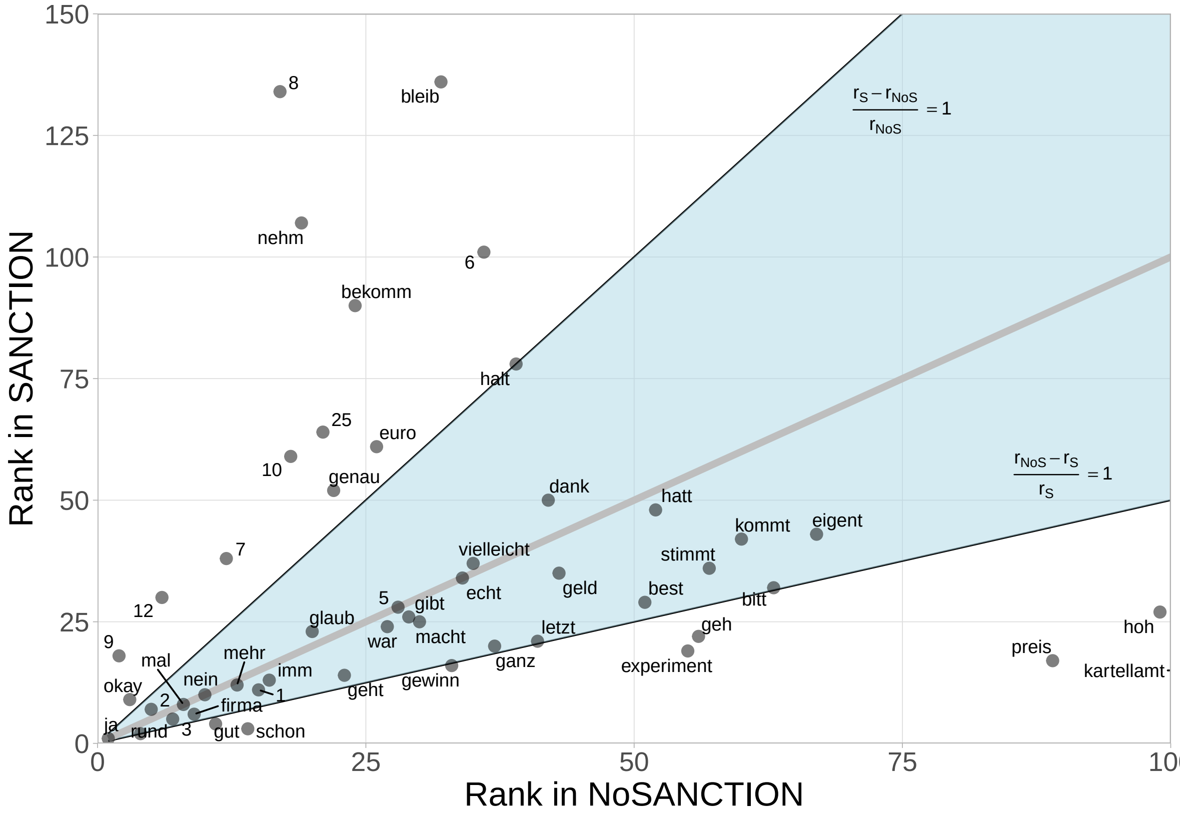


Figure 19: Frequency rankings of the 50 most used tokens in both treatments in German. Tokens that appear outside or at the border of the shaded area in Figure 19 have a relative rank differential weakly exceeding 1.