The effect of whistle–blowing incentives on collusion: an experimental study of leniency programmes

Nick Feltovich*
Monash University
nicholas.feltovich@monash.edu

Yasuyo Hamaguchi
Nagoya City University
yhamagu@econ.nagoya-cu.ac.jp

Abstract

Policy makers are increasingly using whistle–blowing incentives aimed at curtailing illegal or unethical behaviour. We theoretically and experimentally investigate one version of whistle–blowing incentive: *leniency programmes* aimed at curbing anti–competitive activities by firms, by reducing the punishment faced by a cartel member who reports the cartel's behaviour. The theoretical model captures the two important effects of whistle–blowing incentives: the *direct effect*, a reduction in the stability of cartels, and the counterproductive *indirect* effect, an increase in the incentives to form cartels in the first place by lowering the cost of exiting them. As these point in opposite directions, the net theoretical effect is indeterminate. Our laboratory experiment compares two leniency programmes – full immunity from fines and partial immunity – against a baseline with no whistle–blowing incentives in place. We find evidence of the direct effect but not the indirect effect, and thus both programmes reduce the extent of price fixing and the damage associated with it.

Journal of Economic Literature classifications: L41, K42, D43, C73, D03.
Keywords: leniency programme; whistle–blowing; antitrust policy; oligopoly; collusion.

*Corresponding author. Some of this research took place while Feltovich was at University of Aberdeen. Financial support from the UK Office of Fair Trading, the Japan Society for the Promotion of Science (Grants–in–Aid for Young Scientists (B), #21730231) and the Nomura Foundation for Social Science is gratefully acknowledged, though the views expressed are solely those of the authors. We thank Shuya Hayashi, Stephen King, Erika Seki, participants at several conferences and seminars, and two anonymous referees for helpful suggestions and comments.

1 Background

Many forms of illegal or unethical behaviour – from cheating on taxes, to bribery of officials, to firms' collusion – are difficult for outsiders to detect. The use of "whistle–blowers" is invaluable in curbing these kinds of activities and prosecuting those involved. Whistle–blowing often entails substantial personal cost, through self–incrimination or fear of reprisal. In order to overcome these costs, governments are increasingly instituting incentives designed to reduce the cost to the whistle–blower. These incentives take many forms, such as workplace protections for employees blowing the whistle on their bosses, a share of tax receipts for citizens reporting tax cheats, or reduced fines and punishments for collusive firms reporting their activity.

One important aim of these incentives is to increase whistle–blowing when deterrence has failed. We refer to this as the *direct effect*. This benefit may be offset by a perverse *indirect effect*: as an incentive reduces the cost of exit from a corrupt arrangement, such arrangements may become more likely to form in the first place. This latter effect is unlikely in some settings (e.g., paying people to report tax cheats may not increase tax evasion) but potentially important in others. Experimental studies of corruption sometimes find that the possibility of whistle–blowing leads to more, not fewer, corrupt transactions (Abbink 2006; Lambsdorff and Frank 2010), and theoretical and empirical analyses of leniency programmes show that while they can make cartels less stable, they can also make them more likely to form in the first place (Motta and Polo 2003; Chen and Harrington 2007; Marvão 2016).

In this paper, we investigate these two effects. We concentrate attention on the setting of an oligopolistic industry, where the misbehaviour in question is price fixing by the firms, and the whistle–blowing incentive is a *leniency programme*. This setting is worthy of our focus for several reasons, most notably because of the extent of anti–competitive behaviour by real firms, and the corresponding scale of damage to consumers and to social welfare.[1] Additionally, the increasing prevalence of leniency programmes (see Hamaguchi et al. 2009 for some statistics) makes it important to understand their effects. In particular, indirect effects are more likely to be seen here than in other settings of this kind, meaning that the overall effects are especially difficult to determine based on theory alone.

Attempts to assess leniency programmes using field data can suffer from other difficulties. Foremost is that the researcher may have good information about collusion that is detected by the com-

[1]Estimates of aggregate harm to consumers from anti–competitive behaviour are difficult to find, but Laitenberger and Smuda (2015) estimate that a single cartel (detergent), lasting for three years in eight countries, caused 315 million Euros in consumer damage. Baker (2003) estimates the deadweight loss from anti–competitive behaviour in the US to be at least 1 percent of GDP, which is likely a conservative estimate of harm to consumers since pure transfers from consumer to producer resulting from higher prices are not directly counted (though may be indirectly included if these transfers lead to rent seeking by would–be monopolists).

petition authority, but less about undetected cartels. As a result of this selection bias, identifying the entire population of collusive episodes – and hence reaching general conclusions about how leniency programmes affect collusion – can be problematic. Additionally, many of the relevant variables governing behaviour (such as firms' costs and the shape of the demand curve) are largely unobservable to the researcher. Even when these obstacles can be surmounted, natural experiments (pairs of policy environments that are identical except for the leniency programme) are very rare, so it is difficult to accomplish direct tests of policy variables (see, however, Miller 2009, and Brenner 2009).

We therefore turn to the lab. The previous experimental literature on leniency programmes (see Section 2.4) shows a reasonable consensus on the "direct effect": increasing the incentive to whistle–blow nearly always leads to more whistle–blowing. However, the effects on the overall level of collusion and on prices vary from negative to positive, indicating that the "indirect effect" varies in strength from negligible to powerful. The range of conclusions in this literature suggests a need for additional research.

Subjects in our experiment play the role of firms in a repeated duopoly with communication before the first period. Prices are chosen simultaneously, and in the event that both firms choose high prices, they run the risk of being punished by a competition authority. In two "leniency" treatments, firms can avoid the fine – either partly or completely – by blowing the whistle on the rival firm. We compare these results to a baseline where reporting and leniency are not possible.

Our results should inspire optimism about leniency programmes and whistle–blowing incentives in general. Our data indicate that indirect effects are negligible, with no increase in cartel formation under either partial or full leniency. On the other hand, there is plenty of evidence for the direct effect, as collusion is less stable in either leniency treatment than in the baseline, and this decrease in stability is due to firms reporting to the competition authority. The overall impact of these leniency programmes, while smaller than the incidence of whistle–blowing on its own would suggest (due to a countervailing reduction in price under–cutting), includes a decrease in the fraction of time firms spend in collusion and decreases in the harm done to consumers (measured by either prices or firms' excess profits).

2 The duopoly game and experimental design

Our experiment is based on the symmetric 3x3 "underlying game" shown in Figure 1. Two firms simultaneously choose prices: High, Medium or Low.[2] High and Medium on their own form a pris-

[2] We focus on a market with two firms, as the highly competitive nature of markets with even a small additional number of sellers is well recognised by economists. In the theoretical industrial organisation literature, for example, Selten (1973) shows that firms are substantially more likely to behave competitively in markets with five or more firms than in markets with fewer firms. In the empirical IO literature, Bresnahan and Reiss (1991) look at geographically separated markets

oners' dilemma: Medium signifies non–collusive competition while High represents an attempt to collude. The Low price is strictly dominated in the stage game, but its existence allows for a richer set of repeated–game Nash equilibria than would otherwise be possible.

[Figure 1 about here]

The stage game is infinitely repeated with fixed discount factor $\delta \in (0, 1)$. Since (Medium, Medium) is the unique Nash equilibrium of the stage game, there always exists a subgame perfect equilibrium of the repeated game in which both firms choose Medium in every round. However, when δ is sufficiently high, collusion can be achieved. For example, consider the "trigger strategy" in which a player chooses High as long as the opponent has never chosen Medium or Low, but punishes any choice of either of these by thence choosing Medium forever. Both players' choosing this trigger strategy constitutes a subgame perfect equilibrium of the repeated game as long as $\delta \geq \frac{1}{7}$.

2.1 The baseline game

To get the stage game used in our baseline treatment, we introduce competition law in the shape of a non–strategic competition authority. The competition authority discovers anti–competitive behaviour, which we define as a choice of High by both firms, with exogenous probability $p \in [0, 1]$. Upon discovery, two penalties are imposed: (i) an exogenous fine $F > 0$ in the current round, and (ii) a prohibition of choices of High in all future rounds.[3] Penalty (ii) reduces the stage game to the game shown in Figure 2, which we refer to as the game *with punishment in effect*. This reduced game also has (Medium, Medium) as the unique Nash equilibrium.

[Figure 2 about here]

Adding competition law has no effect on payoffs as long as the firms do not collude. So, there continues to be a subgame perfect equilibrium of the repeated game in which both firms always choose

in several industries, and find that three sellers in a market are typically enough to yield competitive prices. Isaac and Reynolds (2002) and Huck et al. (2004) find similar results in experiments, with four or more firms giving rise to a strong push toward competitive prices, collusion common when there are only two firms, and three firms yielding intermediate price levels. Based on these results from a wide variety of market settings, collusion seems to arise fairly easily when there are only 2 sellers, while 3–5 is typically enough to ensure approximately competitive outcomes.

[3]This latter penalty is meant to reflect the likelihood that once the competition authority has detected anti–competitive behaviour, it will have gained information about market conditions that will help it identify excessive prices more easily in future, and additionally it will keep these firms under increased scrutiny. Our assumption that this makes future collusion impossible, rather than just more difficult, is obviously a simplification of the real world (firms caught colluding are able to re–form cartels and sometimes do so), and it means that while we can study cartel formation and stability within our setting, we cannot investigate recidivism. However, this assumption is not unknown in the literature (see, e.g., Hinloopen 2006). Also, the alternative assumption typically made in this experimental literature – that the probability of a second cartel being detected is no higher than it was for the first cartel containing the same firms – is arguably also unrealistic.

Medium. As before, collusion can be supported for δ sufficiently high, though the threshold value will typically be higher than in the underlying game due to the possibility of getting caught and punished.[4]

2.2 The game with leniency

From the baseline game, we make one more modification to obtain our game with leniency: following a (High, High) pair, the competition authority allows each firm the opportunity to report this anti–competitive behaviour (see Figure 3). If either firm does so, then they are caught with certainty, and the restriction to Medium and Low prices goes into effect as before. The reporting firm receives a reduced fine of $F - R$ (with $R > 0$ the "reward" for reporting), while the firm not reporting receives the fine F. If both firms choose to report, both receive the reduced fine.

[Figure 3 about here]

Again, there is a subgame perfect equilibrium with Medium chosen by both firms in all rounds, and trigger strategies can support collusion for sufficiently high δ. There also exist subgame perfect equilibria in which both collusion and reporting take place. This cannot be achieved using trigger strategies, but one type of strategy that can support reporting is as follows. The firm chooses High in the first round, and if the other firm also chooses High, reports in that round and then chooses Medium in all future rounds. If either firm did not choose High in the first round, there is a j–round "punishment stage" in which both firms are meant to choose Low in each round. If after the punishment stage, both firms have complied with the punishment (i.e., chosen Low for j consecutive rounds), the firm continues by choosing Medium forever. If, however, either firm fails to comply with the punishment, then a new punishment stage of the same length begins immediately after the deviation.

This defines a family of "carrot and stick" strategies indexed by j, each of which (when played by both firms) has collusion and reporting along the path of observed play. To determine whether – for a given member of this family – this constitutes a subgame perfect equilibrium, we must check three conditions. First, we verify that the continuation after collusion or after punishment ends is subgame perfect; this follows from (Medium, Medium) being a stage–game equilibrium. Second, we check whether the punishment is credible: that is, both firms prefer to comply. Complying yields 0 for the first j rounds (starting from the beginning of punishment) and 4 thereafter, for a current value of $4\delta^j/(1-\delta)$, while choosing Medium or High in every round yields 2 in each round, for a current value of $2/(1-\delta)$, so complying is sequentially rational when $\delta \geq 1/2^j$. (A one–off choice of Medium or High followed by complying with the new punishment yields the same condition on δ.)

Third, we test whether firms prefer to choose High and report in the first round, assuming the other follows the strategy as specified. Clearly choosing High and not reporting is dominated; it earns

[4] For example, the trigger strategy described in the previous section can support collusion if $\delta \geq \frac{1+pF}{7(1-p)} \geq 1/7$.

less in the first round and the same from then on. Similarly, choosing Low is dominated by choosing Medium. So we just need to compare choosing High and reporting against choosing Medium and complying with the punishment. The former yields $10 - F + R$ in the first round and 4 in all later rounds, for a present value of $6 - F + R + 4/(1 - \delta)$. The latter yields 11 in the first round, 0 in the next j rounds and 4 after that, for a present value of $11 + 4\delta^{j+1}/(1 - \delta)$. So choosing High and reporting is weakly better if $6 - F + R + 4/(1 - \delta) \geq 11 + 4\delta^{j+1}/(1 - \delta)$, which simplifies to

$$\delta(1 + \delta + \cdots + \delta^{j-1}) \geq \frac{1 + F - R}{4}. \quad (1)$$

Therefore, if $\delta \geq 1/2^j$ and (1) holds for a given j, that member of the family of strategies supports collusion and reporting in a subgame perfect equilibrium.

Note that the left–hand side of (1) is increasing in δ. So, as R increases, the threshold value of δ to support collusion and reporting weakly decreases. That is, collusion and reporting become easier to support as the reward for reporting increases.

2.3 Experimental design and theoretical implications

In the experiment, the discount factor δ is 0.8, which is large enough to support many kinds of equilibrium but small enough to allow multiple supergames in an experimental session. Exogenous detection of collusion occurs with probability $p = 0.08$, and the fine is set at the stage–game gain from collusion over the competitive outcome ($F = 6$).[5] The reward for reporting anti–competitive behaviour R takes on one of two values: $R = 3$ or $R = 6$, corresponding to partial or full immunity from the fine. We refer to these as our "partial leniency" and "full leniency" treatments, to contrast with our baseline treatment, where reporting is not possible but detection by the competition authority is.

In all three treatments, there exist anti–competitive subgame perfect equilibria with the High price in all rounds (unless caught by the competition authority) and pro–competitive equilibria with Medium in all rounds. In our partial– and full–leniency treatments, there also exist subgame perfect equilibria in which reporting occurs, supported by the kinds of strategy described in the previous section. Namely, a one–round punishment stage will support collusion and reporting under full leniency as long as

[5] Both the probability of being caught in the absence of reporting, and the fine for being caught, are arguably lower than a regulator would choose given the power to do so. However, setting these on the low side reflects the reality of competition enforcement, where these variables are generally viewed as insufficient to deter collusion due to technological, legal and political constraints. As noted by Connor and Bolotova (2006, p. 1115), "punitive sanctions are the exception not the rule for illegal international price fixing" (p. 1115), leaving only a compensatory component (at most). Setting a low probability and fine for being caught serves an obvious additional purpose in our experimental setting: if collusion is rare without a leniency programme in effect, there is little power to detect a treatment effect from introducing leniency.

$\delta \geq 0.5$, but not under partial leniency for any discount factor. A two–round punishment stage will support collusion and reporting under both partial leniency and full leniency for any $\delta \geq 0.707$.[6]

This multiplicity of equilibria makes precise theoretical predictions impossible. However, the discussion above shows that as we move from no leniency to partial leniency and thence to full leniency, equilibria with collusion and reporting become easier to support, in the sense that a given equilibrium requires a lower discount factor, or alternatively for a given discount factor, a larger set of such equilibria exists. Hence, we hypothesise the following:

1. Reporting of collusive behaviour by firms is more likely under full than under partial leniency.

2. Collusion is more likely under full than under partial leniency, and more likely under partial leniency than in the baseline.

Note that these correspond to the direct and indirect effects of leniency programmes respectively, and that if both are true, then the net effects of leniency programmes on outcomes like prices, firm profits, and consumer surplus are theoretically ambiguous.

The final important aspect of our experimental design is the opportunity for cheap talk (costless, non–binding communication) between the two firms prior to the first round. We do not attempt to model cheap talk theoretically, but we note that experimental studies of cheap talk have tended to find that it raises the level of cooperation – even when such cooperation is not an equilibrium outcome.[7] We restrict cheap talk to the first round, so that the relevant solution concept remains subgame perfect equilibrium rather than some notion of renegotiation proofness (Farrell and Maskin, 1989).

2.4 Related literature

The experimental literature on leniency programmes dates back to Apesteguia et al. (2007, hereafter ADS). Several other studies (Hinloopen and Soetevent, 2008, hereafter HS08; Hamaguchi et al., 2009, hereafter HKS; Bigoni et al., 2012, hereafter BFLS12; Hinloopen and Soetevent, 2014, hereafter

[6]As the length of the punishment stage increases, it becomes easier to satisfy the condition that High price and reporting in the first stage out–earns Medium and complying with the punishment, but harder to satisfy the condition that complying out–earns ignoring the punishment. Hence the threshold discount factor δ^* can either rise or fall as punishment becomes more severe. For our parameters, δ^* falls under partial leniency as j increases from 1 to 2, but rises beyond $j = 2$, while δ^* always rises under full leniency. In particular, δ^* for a three–round punishment is approximately 0.794, just below the actual discount factor we use.

[7]For example, Fonseca and Normann (2012) report substantially higher prices in Bertrand oligopoly when communication is possible, as long as the number of firms is small (including the case of duopoly). See also Sally (1995) and Balliet (2010) for meta–analyses of communication in social dilemmas.

HS14; and Bigoni et al., 2015, hereafter BFLS15) have been conducted since then using oligopoly settings like ours. Two other studies (Hamaguchi et al., 2007, hereafter HIIKT; Hinloopen and Onderstal, 2014, hereafter HO) have used auction settings with collusion expressed as underbidding rather than overpricing; these are similar enough to be worth listing here. Besides oligopoly versus auction, the studies vary in several ways: number of agents (from 2 up to 7), probability of getting caught if no–one reports (from 0 up to 0.4), the definition of collusion (mutual choice of a high price or mutual agreement to communicate), specifics of the leniency programme (whether only current collusion can be reported or also past collusion, whether multiple agents can report, etc.), and so on.[8]

[Table 1 about here]

A summary of these experiments and their results is presented in Table 1. For each study and treatment, the table shows the effect of moving from no leniency programme to partial leniency, and from partial leniency to either full leniency or a bonus, for three outcome variables: average prices, collusion level (according to whatever definition the authors used), and the frequency of reporting.[9] As the table shows, there are few general results. There is substantial, though not universal, agreement that increasing leniency leads to increased reporting (our direct effect). The impact on prices or collusion (our indirect effect) is less clear, though the preponderance of the evidence suggests a non–monotonic effect: moving from no leniency to partial leniency tends to decrease both prices and collusion, while moving from partial to full leniency or a bonus tends to increase them.

Table 1 suggests two roles for additional studies such as ours. First, the lack of consensus suggests that even simply adding to the weight of evidence regarding the effects of leniency programmes is worthwhile. Second, while designing lab experiments inevitably involves a number of fairly arbitrary choices – where several options are reasonable but none is manifestly "correct" – the variety in designs shown in Table 1 suggest that this may be especially true for experimental tests of leniency

[8]The definition of collusion, in particular, has been the subject of a great deal of controversy. As noted already, we define collusion as both firms choosing the High price. This reflects our view that collusion means pricing above stage–game equilibrium levels, meaning that both tacit and explicit collusion can exist. However, prosecuting collusion without evidence of the "smoking gun" of a verbal or written agreement is often prohibitively difficult. This is why many previous studies have used agreement to communicate (but not the communication itself, or evidence of a verbal agreement) as their definition of collusion. Our view of collusion is shared by some (Posner, 2001; Mezzanotte, 2009; Kaplow, 2011; Kovacic et al., 2011; HKS; HS14), though we acknowledge that many others seem to view the communication itself as the offence.

[9]When aggregating multiple studies, some simplifications are necessary. For example, we define partial leniency as anything not guaranteeing full immunity from fines; this includes programmes that give full leniency to a single whistle–blower but less to at least one when there are multiple whistle–blowers. Also, to allow easy comparison between the two auction experiments and the others, "price" in the former is actually the corresponding bidder surplus (which is negatively related to the bid price, the variable actually chosen by subjects). Finally, we leave out degenerate results, such as the increase in reporting associated with moving from a no–leniency treatment with reporting impossible to a leniency treatment with reporting allowed.

programmes. It then becomes important to verify that observed results are not driven by the particular choices that were made. Our study differs from each of these previous studies in at least one notable way, meaning that our results add to the robustness of this literature.

2.5 Experimental procedures

In total, there were 23 experimental sessions with 384 subjects (see Table 2 below). The sessions were conducted in 2009 and 2010, and took place at the the University of Aberdeen's Scottish Experimental Economics Laboratory (SEEL) and Kyoto Sangyo University's Kyoto Experimental Economics Laboratory (KEEL); ethics approval was given by both universities. Subjects in both locations were primarily undergraduate students, recruited using ORSEE (Greiner 2015) from databases of people expressing interest in participating in economics experiments. No one took part more than once; there were no other exclusion conditions.

[Table 2 about here]

We implement infinite repetition with discounting via a fixed continuation probability of δ each round. So, the realised number of rounds varies across supergames, and the total number of rounds played varies across sessions. Each session comprises either five or six supergames (depending on how quickly the first five finished). In the first three supergames ("Part 1" of a session), there was never a leniency programme. In "Part 2", there was either no leniency programme, partial leniency, or full leniency, depending on the session. Having no reporting or leniency in Part 1 gave subjects an opportunity to become familiar with the setting; it will also allow a measure of control for unobservable subject characteristics in our later regression analysis (see Section 3).

At the beginning of a session, subjects were seated in a single room and, after signing consent forms, were given written instructions for Part 1.[10] The instructions stated that the experiment would comprise two parts, but details of the second part would not be announced until after the first part had ended. The instructions were read aloud to the subjects, in an attempt to make the rules common knowledge. Then, the first round began (there was no instructions quiz). After the third supergame was completed, new instructions were distributed and read aloud for Part 2 (even in the baseline treatment, where the game did not change), before the fourth supergame was played.

The experiment took place on networked computers, using z–Tree (Fischbacher 2007). Subjects were asked not to communicate except via the computer program. Subjects were randomly paired at the beginning of a supergame, but pairings were then fixed until the supergame ended. No identifying

[10]The instructions from the partial–leniency treatment conducted in the UK are in Appendix C. Instructions from the remaining treatments (including those conducted in Japan), and other experimental materials, are available from the corresponding author upon request.

information was given to subjects about their opponents (in an attempt to minimise incentives for reputation building and other supergame effects). Rather than using potentially biasing terms like "opponent" or "partner", we used more neutral terms such as "other firm" or "firm matched to you".

[Figure 4 about here]

Each round began with subjects being prompted to choose a price (High, Medium or Low). In the first round of each supergame, the screen also contained a "chat room" (see Figure 4 for the version used in the UK). Subjects were instructed not to send messages containing (a) personal or identifying information or (b) physical threats, but messages were otherwise unrestricted. Own and opponent messages were visible until a price was chosen or until the time available (75 seconds) ran out. Subjects could not observe other pairs' messages.

[Figure 5 about here]

Once all subjects had chosen prices, a feedback screen displayed the own and opponent prices. If a leniency programme was in effect, subjects reaching a (High, High) outcome were prompted to choose whether to report (see Figure 5). The next screen showed the complete result for the round, including whether either subject in the pair reported, whether collusion was discovered, and both profits for the round. Finally, subjects were notified whether the supergame would continue or end.

At the end of the session, one round from each supergame was randomly chosen for each subject, and the subject was paid his/her earnings in those rounds. Each "lab pound" was exchanged for £0.50 of real money in the UK, and for 150 yen in Japan (rounded to the nearest £0.01 or 1 yen); there was no show–up fee. Total earnings averaged about £15 in Aberdeen and 2400 yen in Kyoto, for a session typically lasting 60–75 minutes.

3 Experimental results

Throughout our analysis, unless stated otherwise, our unit of observation is a supergame–pair: an entire supergame played between a pair of subjects. This is more useful for our purposes than the typical analysis focussing on individual choices in each round, as it allows us to control for differences in supergame length within and across sessions (and treatments), and avoids excessively weighting results from the longer supergames.

3.1 Pricing, collusion and reporting

Table 3 reports some aggregates from the experiment, along with p–values from tests of pairwise significant differences between treatments in Part 2.[11] In computing these p–values, we use session–level data and two–sided rejection regions, even when our hypothesis is directional. This means that our non–parametric tests are very conservative (i.e., they under–state significance). To get a full picture of the significance of our treatment effects, these results should be viewed in conjunction with the regression results in later sections (which tend to over–state significance).

[Table 3 about here]

We define a "collusion attempt" as a High price choice by either subject in any round of a supergame–pair. So for example, the frequency of 0.919 in Part 2 of the baseline treatment means that in just under 92 percent of supergame–pairs, there was at least one High price choice, while in the other 8 percent, no–one chose High at any time. Similarly, "cartel formed" is a (High, High) outcome (which we also abbreviate as HH) occurring in one or more consecutive rounds at any time during the supergame for that pair of subjects; we also call this "successful collusion". This is distinct from "collusion extent", which is the fraction of rounds within a supergame–pair that the HH outcome occurs, and serves as one measure of harm to consumers.

Collusion can last until the supergame ends, or it can break down in one of three ways: reporting by one or both of the subjects in the pair, exogenous detection by the competition authority, or by "under–cutting", which we define as any choice by either subject of a Medium or Low price following a previous–round HH outcome that had not been detected. Under–cutting, unlike reporting or exogenous detection, allows the possibility of future collusion, so it is possible for a supergame–pair to successfully collude more than once.[12] Finally, since nearly all price choices were either Medium or High (Low prices represented fewer than 1 percent of price choices overall: 1.2 percent when punishment was in effect, and 0.9 percent when all three prices were available), the fraction of High choices will serve as our measure of "average price".[13]

In Part 1, most pairs attempt to collude, with about 85 percent having at least one High price choice by a firm. Just under half of pairs successfully collude at least once, and overall they collude in almost 30 percent of rounds. In Part 2 of the baseline sessions, both high prices and collusion become more

[11] See Siegel and Castellan (1988) for descriptions of the non–parametric tests used in this paper, and Feltovich (2005) for critical values for the robust rank–order test.

[12] To be precise, there are 478 supergame–pairs with exactly one successful collusion, while 19 pairs (3.7 percent) collude exactly twice, 3 (0.6 percent) collude three times and 1 pair (0.2 percent) colludes four times.

[13] The scarcity of Low price choices might be viewed as evidence against those equilibria we described in Section 2.2 that relied on Low price choices. However, it is important to remember that while Low price choices made up portions of these strategies, none of them required Low along the observed path of play: it only appeared in unreached subgames.

frequent, with nearly 60 percent of pairs colluding at least once and HH outcomes occurring just under half of the time. Since there is no difference in any strategic aspect of the game between Part 1 and Part 2 of the baseline, these increases seem to be due to increased cooperation between subjects as they become more experienced in this setting.

The table shows no evidence of the conjectured "indirect effect" of leniency programmes, as neither leniency treatment shows any increase in attempts to collude or successful cartel formation. There is, however, a significant decrease in prices in both leniency treatments compared to the baseline, as well as a decrease (though insignificant) in the fraction of rounds in which collusion occurred. These decreases are attributable to the "direct effect" of the leniency programme: just over 20 percent of pairs in the partial–leniency treatment, and nearly half in the full–leniency treatment, end collusion by reporting, with significantly more reporting under full leniency than under partial leniency. There is a corresponding decrease in payoffs under either leniency programme compared to the baseline, though this decrease (like the decrease in the extent of collusion) is smaller than it should be, because the increase in reporting is partly offset by a decrease in under–cutting. This suggests, that some – though not all – of reporting is done as a substitute for under–cutting, though it must be emphasised that the total frequency of endogenous cartel dissolution (reporting or under–cutting) does increase from no leniency to partial leniency to full leniency (though only the former increase is significant). Finally, even the reduced payoffs in the leniency treatments are well above the competitive level of 4.

[Figure 6 about here]

Figure 6 shows some additional statistics about collusion. Here we depart slightly from taking the supergame–pair as the unit of observation, and instead take each episode of successful collusion as an individual unit. The distinction exists since a small fraction of supergame–pairs successfully colludes more than once (see Note 12). The left panel shows that the vast majority of cartels (over 85 percent) are formed in the first round of a supergame. The fraction is even higher (nearly 90 percent) in Part 2 of the experiment, and does not vary substantially across treatments (90 percent in the baseline, 92 percent under partial leniency, 88 percent under full leniency). The right panel shows that collusion is often successful for only a single round; one–round cartels form a majority of all cartels overall, and nearly a majority (47 percent) in Part 2. There is some variation across treatments, consistent with leniency programmes' reducing cartel stability: one–round cartels comprise 33 percent of all Part–2 cartels in the baseline, 42 percent under partial leniency, and 59 percent under full leniency.

[Figure 7 about here]

Cartels that last for only one round not only predominate, they are different in nature from longer–lasting cartels. Figure 7 categorises all episodes of collusion according to whether they ended endogenously (due to subject decisions – either reporting or price under–cutting) or otherwise (exogenously due being detected by the competition authority without having been reported, or only by the supergame ending, which for present purposes we classify as an exogenous ending). Over all supergames and treatments, over half of the cartels lasting for exactly one round ended endogenously. In Part 2, 45 percent of cartels ended endogenously in the baseline, compared to 65 percent under partial leniency and 77 percent under full leniency. Also varying was the proportion of endogenous breakdowns due to reporting versus under–cutting: 73 percent of endogenous breakdowns were due to reporting in the partial–leniency treatment and 83 percent under full leniency, as compared to zero in the baseline. By contrast, once collusion survived into a second round, it was unlikely to break down due to decisions by the firms themselves: only about one–fifth of cartels lasting 2–4 rounds ended endogenously, and none of the cartels lasting for 5 rounds or longer did.

These properties of collusion, along with the descriptive statistics reported in Table 3, suggest that introducing a leniency programme does lead to changes in firm behaviour. It does not appreciably affect the likelihood of cartels forming, but it reduces their stability. A substantial fraction of cartels break down via reporting, though this fraction over–states the leniency programmes' effect on collusion, since there is a partially off–setting decrease in under–cutting.

We next present results from several regressions. We estimate eight models, each with a different left–hand–side variable, using either the entire Part–2 data or, when appropriate, the subset of supergame–pairs that reached an HH outcome (i.e., a successful collusion) at least once. The Part–1 data are used only to construct a variable to control for subjects' intrinsic proclivity toward collusiveness, as described below. Our first model is meant to shed light on the incentives to collude induced by the leniency programmes; the dependent variable is "HH indicator", equal to one if the pair achieves an HH outcome in any round. Our second model looks at the overall prevalence of collusion; the dependent variable is "HH fraction", the fraction of rounds in which collusion took place. Our third model assesses the effects on prices, using the fraction of High price choices as a measure of the average price. Our next three models look at the effects of the leniency programmes on the two endogenous ways of breaking cartels individually (using indicators for reporting and under–cutting), and for either kind of endogenous breakdown (an indicator with value 1 if either subject reported or under–cut). Our last two models look at the leniency programmes' effects on gross and net firm profits.

We used similar right–hand–side variables in each regression. Our main explanatory variables are indicators for the partial– and full–leniency treatments (except for the "Report" regression, which does not use the data from the baseline treatment, and hence we dropped the partial–leniency treatment indicator). We include the supergame number to control for changing behaviour over time, an indicator for the session taking place in Kyoto to control for any subject–pool effects, and a constant term. Finally, we included an index of the pair's collusive or cooperative tendencies, called "collusiveness";

this was simply the fraction of High choices within the pair in the first round of each supergame in Part 1 of the session (thus taking on values that are whole–number multiples of one–sixth). We use Tobit models for the two profit variables, H fraction and HH fraction, and probits for the others. All of the models were estimated using Stata (version 12) with robust standard errors.

[Table 4 about here]

The results, displayed in Table 4, largely confirm what was observed in the descriptive statistics.[14] The insignificant and near–zero marginal effects of the partial– and full–leniency variables on the HH indicator confirm that despite potentially providing incentives for firms to collude, these leniency programmes do not make cartels more likely to form. Moreover, they lead to lower average prices, as shown by their negative effect on the H fraction, and they make cartels less stable, as shown by their negative effect on the HH fraction. As with the aggregate data, we can see that these effects on prices and the extent of collusion are driven primarily by reporting, which is significantly more likely under the full–leniency programme than under partial leniency. Under–cutting is significantly less likely under full leniency than in the baseline, while there is no significant difference between partial leniency and the baseline (though there is also no significant difference between partial and full leniency). Endogenous breakdowns are more likely under full leniency than either the baseline or under partial leniency ($p \approx 0.012$ for the latter comparison). Both leniency programmes have negative effects on both gross and net profits, but these effects are smaller than they might have been due to the decrease in under–cutting mentioned in the discussion of aggregate results.

3.2 Analysis of pre–play messages

The preceding analysis focuses on measures of the *implementation* of anti–competitive behaviour and its effects. However, the data also allow us to examine the impact of our treatments on the kinds of messages that were exchanged – and in particular *verbal agreements to collude* – at the beginning of each supergame, and the relationship between agreements and subjects' subsequent behaviour. We hired research assistants (RAs) to classify each individual "conversation" – and therefore each supergame–pair – according to what events took place during it. The events we focussed on were:

- Any non–blank message;

[14] We are primarily interested in the effects of the partial– and full–leniency treatments, but some of the other variables' effects are worth a brief mention. The strong positive effects of our collusiveness variable on prices, collusion and payoffs (and weak negative effects on the breakdown variables) suggests that there may be intrinsic subject heterogeneity in propensity to collude. The Kyoto–session variable has no significant effects, suggesting that subject–pool effects were limited. Finally, the supergame number has no significant effects, suggesting minimal time variation.

- Both firms in a pair sending a non–blank message (this corresponds most closely to the "mutual agreement to communicate" definition of collusion used in many previous experiments);

- Any message specifically about prices (suggestion, factual statement, question, etc.);

- Any suggestion to collude (a proposal of a joint pricing scheme that included a positive frequency of HH pairs, irrespective of whether and how it was responded to);

- An agreement on any pricing scheme, collusive or otherwise (a proposal made by one member of the pair, assented to by the other, with no subsequent questioning or disagreement); and

- Agreement to collude (agreement on a pricing scheme with a positive frequency of HH pairs).

After this was done, a subset of the original classifiers was asked to partition this last group (agreements to collude) into those agreeing also to report versus those not making such an agreement (those agreeing not to report, those not discussing reporting at all, etc.). All of these variables were coded as binary (e.g., either a conversation mentioned prices or it did not; we were not interested in how many mentions there were after the first). Also recorded were the total number of messages, and the total number of characters over all messages, for each conversation.

For the Japanese data, only one RA was available, so our analysis uses the classifications directly from that RA; this was also the case for "agreement to collude and report" from the UK data. For the rest of the variables based on the UK data, we used three RAs, so we treated an event (e.g., suggestion to collude) has having occurred if at least two of the three RAs classified it as having occurred.

[Table 5 about here]

The top block of Table 5 shows the frequencies of these events. Overall, just under half of the supergames are preceded by a verbal agreement to collude, and an additional quarter have both subjects sending messages but no verbal agreement. Verbal agreements to both collude and report are uncommon, but significantly more likely under full leniency than under partial leniency (robust rank–order test, $p < 0.001$) and of course non–existent in the baseline.

The middle block of the table shows the frequencies of our outcome measures conditional on three communication events: (i) agreement to collude, (ii) both communicate but without reaching an agreement to collude, and (iii) at least one does not communicate. While reaching a verbal agreement is neither necessary nor sufficient for actual collusion, there is a strong positive association between the two. Collusion attempts and cartel formation are more likely, and collusion and high prices occur a greater fraction of the time, following a verbal agreement (case (i)) than without one (cases (ii) and (iii)), according to a Wilcoxon signed–ranks test ($p \approx 0.006$ for collusion attempts, $p < 0.001$ for the other three statistics); for example, collusion occurs in 76 percent of supergames after a verbal agreement to collude but only 27 percent of supergames with no such agreement. There is even weak evidence that cartels are more stable following a verbal agreement, with endogenous breakdowns borderline significantly less likely ($p \approx 0.067$), though there are only insignificant decreases in under–cutting ($p > 0.20$) and reporting ($p \approx 0.164$) separately. Finally, both gross and net profits are significantly higher following a verbal agreement ($p < 0.001$). By contrast, comparison of cases (ii) and (iii) highlights that when there is no verbal agreement to collude, mere agreement to communicate has little systematic effect on any of the measures of collusive behaviour or breakdown.[15]

The bottom block of the table shows outcome frequencies conditional on two subsets of (i): (iv) agreeing to both collude and report, (v) agreeing to collude but not to report. There are only negligible differences in either collusion attempts or successful collusion between these two cases. There is a substantial difference in reporting frequencies, with reporting occurring in more than half of cartels when an agreement to report was made, and less than 15 percent of the time without such an agreement. The increase in reporting is reflected in an increase in endogenous breakdowns, leading to significantly lower fractions of H choices and HH outcomes after an agreement to report (Wilcoxon signed–rank test, $p < 0.001$ for H choices, $p < 0.01$ for HH outcomes).

This last result suggests that the reduction in average prices and extent of collusion in our leniency treatments may be due to subject pairs making verbal agreements to report, then honouring these agreements. Indeed, a comparison between agreements to collude and report in our leniency treatments and agreements to collude in the baseline finds significantly lower fractions of H choices and HH outcomes in the former (robust rank–order test, $p \approx 0.036$, $p \approx 0.003$ and $p < 0.001$ for H choices in baseline versus partial leniency, full leniency, and pooled partial and full leniency respectively, $p \approx 0.072$, $p \approx 0.014$ and $p \approx 0.010$ for HH outcomes in baseline versus partial, full, and pooled leniency treatments respectively). However, there is weak evidence that even without an agreement to collude and report, fractions of H choices and HH outcomes are lower in the leniency treatments than in the baseline ($p \approx 0.057$, $p \approx 0.071$ and $p \approx 0.041$ for H choices in baseline versus partial, full, and pooled leniency treatments respectively, $p > 0.20$, $p \approx 0.20$ and $p \approx 0.16$ for HH outcomes in baseline versus partial, full, and pooled leniency treatments respectively). So while a substantial portion of the benefit of these leniency programmes is due to their effect on verbal agreements to report; a small but non–negligible portion comes from reporting even when no such agreement was made.

[15] The cases where the difference between (ii) and (iii) appears largest arise from small samples (e.g., the difference in reporting under partial leniency between (ii) and (iii) is based on one observation and five observations in the two samples).

4 Summary and discussion

In this paper, we have examined leniency programmes aimed at price–fixing firms, using an experiment with three treatments: one with no leniency programme, one where leniency involves a reduced fine (partial leniency) and one where it involves complete immunity from the fine (full leniency). Neither theoretical analysis nor review of the previous literature yields a clear prediction for leniency programmes' effects. Our main results are as follows.

Result 1 *There is no increase (significant or otherwise) in the fraction of subject pairs successfully colluding (reaching at least one (High, High) outcome) from the baseline to either leniency treatment.*

Result 2 *The overall extent of collusion (i.e., frequency of (High, High) outcomes) is significantly lower under full leniency than in the baseline. The frequency is also lower under partial leniency than in the baseline, but the difference is not significant.*

Result 3 *Average prices (i.e., the frequency of High price choices) are significantly lower under either full or partial leniency than in the baseline.*

Result 4 *The frequency of reporting, conditional on successful collusion, is significantly higher under full leniency than under partial leniency.*

Result 5 *Average profits are significantly lower under partial or full leniency than in the baseline.*

Also, (a) the increase in reporting observed under either leniency programme compared to the baseline is partly off–set by a decrease in under–cutting; (b) when cartels form, they typically form immediately; and (c) both reporting and under–cutting tend to involve relatively new cartels (both fall off quickly after the first round of collusion, and neither occurs after five rounds of collusion). Subsequent analysis of subjects' pre–play conversations suggests a high correlation between verbal agreements to collude and actual collusion, and between verbal agreements to report collusion and actual reporting. Importantly, those verbal agreements to report collusion occur much more often under full leniency than under partial leniency.

As always, caution should be taken in drawing conclusions about the outside world based on the outcome of an experiment. But with that caveat acknowledged, our results might be interpreted as follows. The effects of introducing a leniency programme, or increasing its generosity, are positive on balance: the direct effect of increased break–down of collusion via greater whistle–blowing is fairly strong, and the potential indirect effect of reduced deterrence is not observed (suggesting that this may be at worst a second–order effect). The net impact includes less time spent under collusion, lower average prices, and firm profits closer to competitive levels. However, several qualifications apply. First, some of the increased reporting comes from firms that would have deviated from collusion anyway (that is, they report instead of under–cutting the rival's price); this crowding–out implies that the leniency programme's visible effects will be smaller than the reporting frequency on its own would indicate. Second, most of the gains from introducing a leniency programme can be achieved with partial leniency; the additional benefits from full leniency are smaller. Finally, neither leniency programme has much ability to break collusion that has become well established.

On balance, our results should inspire optimism about the potential for leniency programmes to reduce anti–competitive behaviour. Some of our findings – such as the decreases in prices and collusion under partial leniency compared to no leniency, the increased reporting to the competition authority under partial leniency, and the further increase in reporting under full leniency – are in line with results seen elsewhere in the experimental literature (see Table 1). Given the differences between our experimental design and these other designs, this suggests that the broad effects of leniency programmes are robust to how collusion and competition policy are implemented in the lab.

By contrast, our findings of further decreases in prices and collusion under full leniency (compared to partial leniency) are at odds with several other studies findings of *increases* when moving from partial to full leniency or to bonuses. We do not view our results as necessarily contradicting theirs, especially given that the differences found in these studies have typically not been significant (and ours was also insignificant). At most, these results illustrate that the effects of moving from a less generous to a more generous leniency programme may be more ambiguous in general, and more sensitive to assumptions about the details of the programme and underlying setting, compared to the effects of implementing a new leniency programme (relative to no programme at all). Clearly, further research is still needed to improve our understanding of how leniency programmes should be designed in order to maximise their benefits.

References

Abbink, K. (2006), "Laboratory experiments on corruption", in S. Rose-Ackerman, ed., *International Handbook on the Economics of Corruption*, Cheltenham: Edward Elgar.

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

Baker, J.B. (2003), "The case for antitrust enforcement", *Journal of Economic Perspectives* 17, pp. 27–50.

Balliet, D. (2010), “Communication and cooperation in social dilemmas: a meta–analytic review”, *Journal of Conflict Resolution* 54, pp. 39–57.

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

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

Brenner, S. (2009), “An empirical study of the European corporate leniency program”, *International Journal of Industrial Organization* 27, pp. 639–645.

Bresnahan, T.F. and P.C. Reiss (1991), “Entry and competition in concentrated markets”, *Journal of Political Economy* 99, pp. 977–1009.

Chen, J. and J.E. Harrington (2007), “The impact of the corporate leniency program on cartel formation and the cartel price path”, in V. Ghosal and J. Stennek, eds., *The Political Economy of Antitrust (Contributions to Economic Analysis)*, Amsterdam, Elsevier, pp. 59–80.

Cowling, K. and D.C. Mueller (1978), “The social costs of monopoly power”, *Economic Journal* 88, pp. 727–748.

Farrell, J. and E. Maskin (1989), “Renegotiation in repeated games”, *Games and Economic Behavior* 1, pp. 327–360.

Feltovich, N. (2005), “Critical values for the robust rank–order test”, *Communications in Statistics – Simulation and Computation* 34, pp. 525–547.

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

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

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

Hamaguchi, Y., T. Kawagoe and A. Shibata (2009), “Group size effects on cartel formation and the enforcement power of leniency programs”, *International Journal of Industrial Organization* 27, pp. 145–165.

Hinloopen, J. (2006), “Internal cartel stability with time–dependent detection probabilities”, *International Journal of Industrial Organization* 24, pp. 1213–1229.

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, pp. 317–336.

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

Hinloopen, J. and A.R. Soetevent (2014), “Exploitation and induced tacit collusion: a classroom experiment of corporate leniency programs”, in M. Peitz and Y. Spiegel, eds., *Analysis of Competition Policy and Sectoral Regulation*, Boston, Now Publishers, pp. 193–212.

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

Isaac, R.M. and S.S. Reynolds (2002), “Two or four firms: does it matter”, in C.A. Holt and R.M. Isaac, eds., *Research in Experimental Economics, v. 9*, Amsterdam, JAI Press, pp. 95–119.

Kaplow, L. (2011), “Direct versus communications–based prohibitions on price fixing”, *Journal of Legal Analysis* 3, pp. 449–538.

Kovacic, W.E., R.C. Marshall, L.M. Marx and H.L. White (2011), “Plus factors and agreement in antitrust law”, *Michigan Law Review* 110, pp. 393–436.

Laitenberger, U. and F. Smuda (2015), “Estimating consumer damages in cartel cases”, *Journal of Competition Law & Economics* 11, pp. 955–973.

Lambsdorff, J. Graf and B. Frank (2010), “Bribing versus gift–giving – an experiment”, *Journal of Economic Psychology* 31, pp. 347–357.

Marvão, C. (2016), “The EU leniency programme and recidivism”, *Review of International Organization* 48, pp. 1–27.

Mezzanotte, F.E. (2009), “Can the Commission use Article 82EC to combat tacit collusion?” CCP working paper 09–5, University of East Anglia.

Miller, N.H. (2009), “Strategic leniency and cartel enforcement”, *American Economic Review* 99, pp. 750–768.

Motta, M. and M. Polo (2003), “Leniency programs and cartel prosecution”, *International Journal of Industrial Organization* 21, pp. 346–379.

Posner, R. (2001), *Antitrust law, 2nd ed.*, Chicago, University of Chicago Press.

Sally, D. (1995), “Conversation and cooperation in social dilemmas: a meta–analysis of experiments from 1958 to 1992”, *Rationality and Society* 7, pp. 58–92.

Selten, R. (1973), “A simple model of imperfect competition, where 4 are few and 6 are many”, *International Journal of Game Theory* 2), pp. 141–201.

Serra, D. and Wantchekon, L., eds. (2012), *New Advances of Experimental Research on Corruption: Research in Experimental Economics, Vol. 15*, Bingley, UK: Emerald Group Publishing.

Siegel, S. and N.J. Castellan, Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, New York, McGraw–Hill.

Table 1: Selected results from previous leniency–programme experiments

Study	Treatment	(N, p)	Collusion definition	Change in degree of leniency	Average price	Collusion level	Reporting frequency
ADS		$(3, 0)$	All choose to	none → part.	– –	–	+
			communicate	part. → full/bonus	+ + (bonus)	+ (bonus)	+ (bonus)
HS08		$(3, 0.15)$	All choose to	none → part.	– –	– –	N/A
			communicate				
	duopoly	$(2, 0.1)$		none → part.	N/A	– –	N/A
HKS			All choose	part. → full/bonus	N/A	+ (full), – – (bonus)	– (full),+ + (bonus)
	7–firm	$(7, 0.1)$	high price	none → part.	N/A	– –	N/A
				part. → full/bonus	N/A	+ (full), – (bonus)	+ + (full, bonus)
BFLS12		$(2, 0.1)$	All choose to	none → part.	– –	– –	+ +
			communicate	part. → full/bonus	– (bonus)	+ (bonus)	+ + (bonus)
HS14		$(2, 0.4)$	All choose	none → part.	+	+	N/A
			high price	part. → full/bonus	+	+	+
	High fine	$(2, 0.02)$	All choose to		– –	–	N/A
BFLS15	Low fine	$(2, 0.1)$	communicate	none → part.	– –	–	N/A
	No detection	$(2, 0)$			– –	–	N/A
HIIKT	FP auct.	$(5, 0.15)$	≥ 2 choose	none → part.	+	–	N/A
			to communicate				
HO	Eng. auct.	$(3, 0.15)$	All choose	none → part.	–	–	N/A
	FP auct.		to collude		+	–	N/A

Key: N: number of firms in each market. p: exogenous probability of detection of collusion. + (–): positive (negative) but insignificant or significance not assessed. + + (– –): positive (negative) and significant. N/A: not applicable (treatment or variable not part of design, variable not measured or not reported in a treatment, etc.).

Table 2: Session information

Location	Treatment	Sessions	Subjects
Aberdeen	No leniency	3	38
(SEEL)	Partial leniency	6	70
	Full leniency	7	90
Kyoto	No leniency	2	50
(KEEL)	Partial leniency	2	54
	Full leniency	3	82

Table 3: Descriptive statistics, (supergame–pair)–level data

	Part	Part 2			pairwise p–values		
	1	baseline	partial leniency	full leniency	base/ part	base/ full	part/ full
Unconditional							
At least one H choice (collusion attempt)	.851	.919	.862	.914	$n.s.$	$n.s.$	$n.s.$
At least one HH outcome (cartel formed)	.424	.586	.538	.555	$n.s.$	$n.s.$	$n.s.$
Fraction of L choices	.009	.010	.011	.004	$n.s.$	$n.s.$	$n.s.$
Fraction of H choices (average price)	.444	.603	.440	.450	0.047	0.085	$n.s.$
Fraction of HH outcomes (collusion extent)	.283	.463	.354	.322	$n.s.$	$n.s.$	$n.s.$
Gross average payoff	6.430	7.576	6.487	6.550	0.047	0.085	$n.s.$
Net average payoff	6.294	7.299	6.076	6.284	0.040	0.077	$n.s.$
Conditional on at least one HH outcome							
Reporting (by at least one firm)	—	—	.218	.450	—	—	0.006
Under–cutting (by at least one firm)	.347	.241	.141	.101	$n.s.$	0.040	$n.s.$
Endogenous breakdown (report, under–cut)	.347	.241	.359	.541	$n.s.$	0.008	$n.s.$

p–values from two–sided robust rank–order tests on session–level data (n.s.: $p > 0.20$)

Table 4: Regression results, Part–2 data – marginal effects at means, std. errors in parentheses

Dependent variable:	HH indicator	HH fraction	H fraction	Report	Undercut	Endog. break.	Gross profit	Net profit
Subsample:	Full sample			At least one HH outcome			Full sample	
partial leniency	–0.073	-0.352^{**}	-0.252^{***}		–0.076	0.129	-1.018^{***}	-1.238^{***}
	(0.068)	(0.180)	(0.072)		(0.049)	(0.093)	(0.276)	(0.277)
full leniency	–0.099	-0.493^{***}	-0.275^{***}	0.318^{***}	-0.117^{**}	0.311^{***}	-1.126^{***}	-1.193^{***}
	(0.064)	(0.168)	(0.065)	(0.057)	(0.054)	(0.082)	(0.241)	(0.244)
Kyoto session	–0.013	–0.000	0.049	–0.071	0.026	–0.026	0.077	0.125
	(0.050)	(0.126)	(0.051)	(0.057)	(0.047)	(0.066)	(0.203)	(0.202)
supergame #	0.016	0.042	0.009	–0.090	0.009	–0.089	0.056	0.078
	(0.036)	(0.090)	(0.037)	(0.040)	(0.031)	(0.045)	(0.147)	(0.146)
collusiveness	0.751^{***}	2.135^{***}	0.854^{***}	–0.130	–0.130	-0.292^{*}	3.276^{***}	3.015^{***}
	(0.105)	(0.300)	(0.105)	(0.128)	(0.104)	(0.154)	(0.363)	(0.367)
N	442	442	442	245	245	245	442	442
$\|ln(L)\|$	276.12	433.92	356.10	122.59	98.38	154.55	950.17	943.10

* (**,***): *Coefficient significantly different from zero at the 10% (5%, 1%) level. Endog. break = endogenous breakdown (report or undercut)*

Table 5: Communication outcomes – descriptive statistics

	Part 1	Part 2		
		baseline	partial leniency	full leniency
Frequency of...				
any non–blank message	.851	.758	.876	.871
both send a non–blank message	.753	.717	.814	.844
any message about prices	.740	.687	.834	.860
suggestion to collude	.498	.535	.745	.688
agreement on any pricing strategy	.495	.535	.703	.715
agreement to collude	.367	.444	.641	.559
agreement to collude and report	.000	.000	.021	.177
Frequency, given (i) agreement to collude; (ii) both send a non–blank message but no agreement; (iii) neither				
	(i)/(ii)/(iii)	(i)/(ii)/(iii)	(i)/(ii)/(iii)	(i)/(ii)/(iii)
Collusion attempt	.971/.815/.725	1.000/.893/.821	.946/.769/.667	.981/.846/.862
Cartel formed	.765/.241/.239	.767/.429/.464	.761/.115/.185	.724/.288/.448
Fraction of H choices	.640/.368/.306	.769/.470/.482	.572/.281/.140	.544/.328/.417
Fraction of HH outcomes	.552/.143/.135	.703/.235/.323	.528/.046/.060	.477/.090/.244
Reporting (given cartel)	.000/.000/.000	.000/.000/.000	.130/1.000/.400	.267/.733/.385
Under–cutting (given cartel)	.288/.442/.485	.093/.417/.385	.098/.000/.400	.057/.133/.231
Endogenous breakdown (given cartel)	.288/.442/.485	.093/.417/.385	.228/1.000/.800	.314/.867/.615
Frequency, given (iv) agreement to collude and report; (v) agreement to collude but no agreement to report				
	(iv)/(v)	(iv)/(v)	(iv)/(v)	(iv)/(v)
Collusion attempt	— /.971	— /1.000	1.000/.944	1.000/.973
Cartel formed	— /.765	— /.767	.667/.764	.727/.730
Fraction of H choices	— /.640	— /.769	.292/.582	.393/.599
Fraction of HH outcomes	— /.552	— /.703	.250/.537	.344/.525
Reporting (given cartel)	— /.000	— /.000	.667/.112	.576/.149
Under–cutting (given cartel)	— /.288	— /.093	.000/.101	.061/.054
Endogenous breakdown (given cartel)	— /.288	— /.093	.667/.213	.636/.189

Table 6: Correlations between alternative definitions of collusion (Part 2 supergames)

	(a)	(b)	(c)	(d)	(e)
(a) HH outcome	1.000				
(b) Both send non–blank message	.190	1.000			
(c) Verbal agreement to collude	.442	.577	1.000		
(d) Both send non–blank message and HH outcome	.849	.499	.601	1.000	
(e) Verbal agreement to collude and HH outcome	.743	.437	.757	.875	1.000

Table 7: Probit results (marginal effects at means, std. errors in parentheses), supergame–level data

	[9]	[10]	[11]	[12]	[13]	[14]
	Dep. variable: verbal agreement to collude				Both send non–blank message	
	All supergames		Part 2 only		All supergames	Part 2 only
partial leniency			0.208^{***}	0.152^{***}		0.089^{***}
			(0.061)	(0.055)		(0.044)
full leniency			0.090	0.073		0.106^{***}
			(0.062)	(0.055)		(0.044)
Kyoto session	0.043	–0.015	0.031	0.259^{***}	0.043	0.034
	(0.032)	(0.061)	(0.049)	(0.094)	(0.032)	(0.036)
supergame #	0.073^{***}	0.059^{***}	-0.077^{**}	–0.048	0.073^{***}	-0.112^{***}
	(0.010)	(0.011)	(0.036)	(0.033)	(0.010)	(0.028)
collusiveness			0.452^{***}	0.381^{***}		0.206^{***}
			(0.104)	(0.096)		(0.076)
number of messages		0.029^{***}		0.012		
		(0.006)		(0.008)		
total length of messages		0.000		0.003^{***}		
		(0.001)		(0.001)		
N	988	988	430	430	988	430
$\|ln(L)\|$	655.07	614.43	278.66	256.92	655.07	197.25

* (**,***): *Coefficient significantly different from zero at the 10% (5%, 1%) level. Message variables also interacted with Kyoto–session dummy.*

Figure 1: The underlying stage game

		Firm 2		
		High	Medium	Low
	High	10, 10	2, 11	2, 0
Firm 1	Medium	11, 2	4, 4	2, 0
	Low	0, 2	0, 2	0, 0

Figure 2: The stage game when punishment is in effect

		Firm 2	
		Medium	Low
Firm 1	Medium	4, 4	2, 0
	Low	0, 2	0, 0

Figure 3: Sequence of decisions in the stage game with leniency

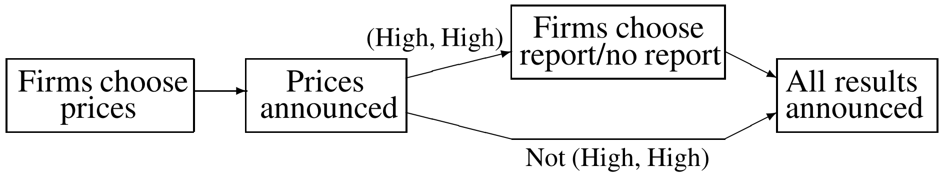


Figure 4: Screen–shot from first round of a supergame

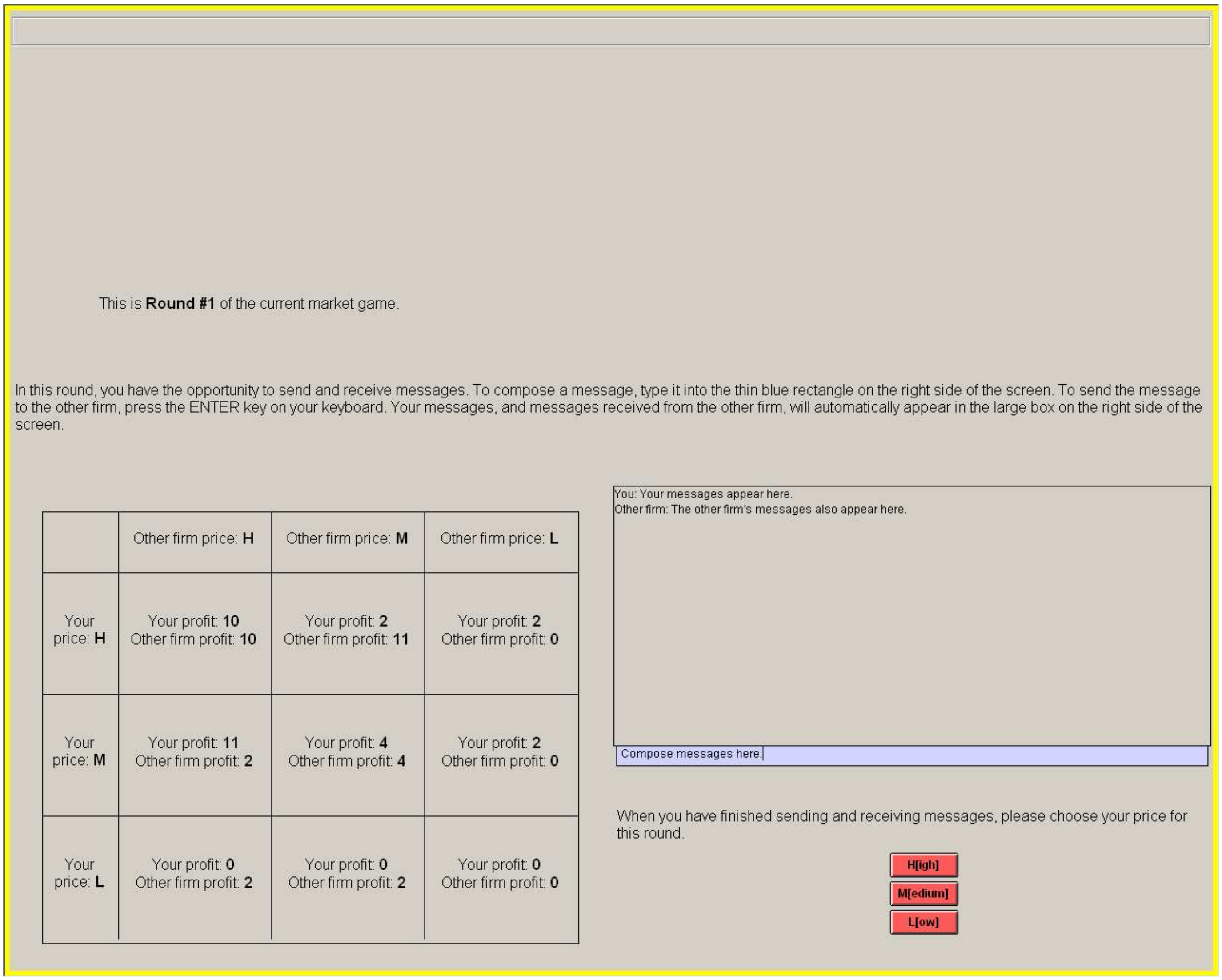


Figure 5: Screen–shot from reporting component following (High, High) outcome

Figure 6: Cartel formation and persistence – pooled treatments and supergames

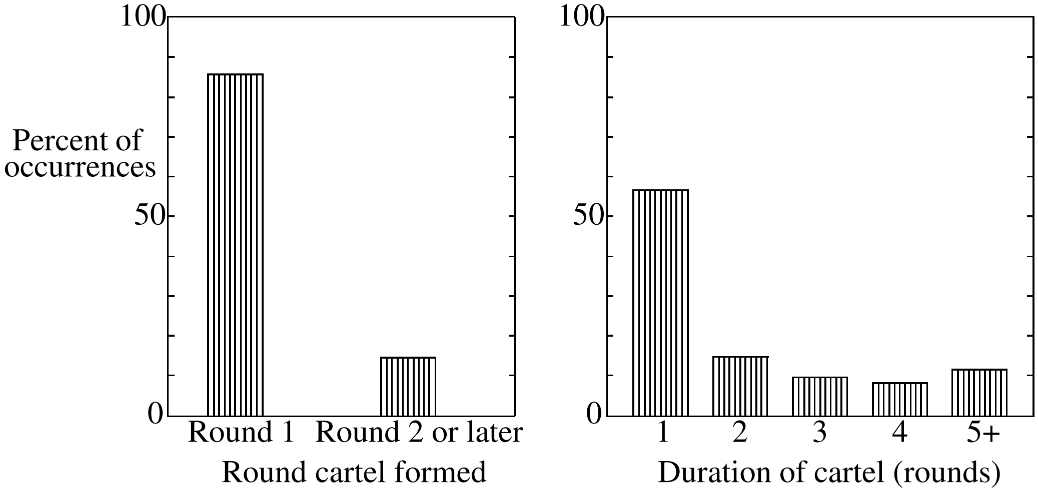
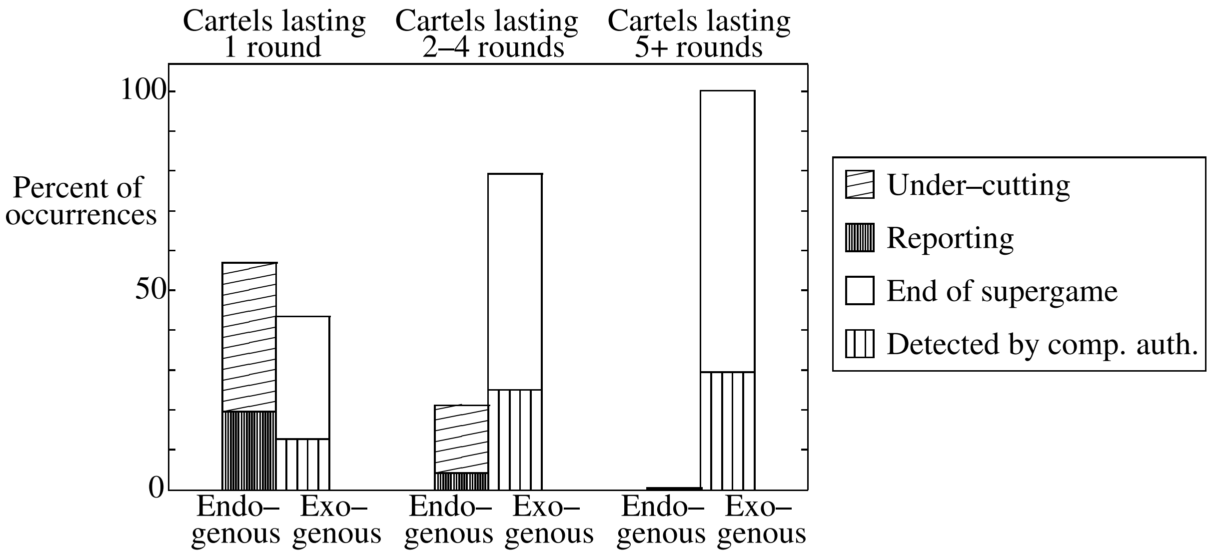


Figure 7: Exogenous and endogenous cartel dissolutions – all treatments and supergames



A Additional analysis of chat messages

Table 6 shows the extent to which various ways of defining collusion are correlated with each other. The first definition we use is (a) the "harm to consumers" one used in the main text: at least one HH outcome during the supergame. The second definition (b) is the "agreement to communicate" one used in previous experiments: both firms choosing to send a non–blank message. We add three more definitions: (c) a verbal agreement to collude; (d) both sending a non–blank message followed by an HH outcome (i.e., (a) plus (b)); and (e) a verbal agreement followed by an HH outcome (i.e., (a) plus (c)). Each can be justified by some combination of competition law, policy and de facto enforcement.

[Table 6 about here]

Although all of the correlations in the table are positive (and significant with $p < 0.001$), the weakest correlation is between our definition and the "agreement to communicate" definition used in previous experiments, with both more highly correlated with definitions (c), (d) and (e) than with each other. Interestingly, the "agreement to communicate" definition is somewhat more strongly correlated with verbal agreements alone (though perhaps not surprisingly, since the former is a necessary condition for the latter), but it is much less strongly correlated with definitions (d) and (e) than our "harm to consumers" definition is, even though *both* are necessary conditions for (d) and (e). Since (e) is arguably the first–best definition of collusion, these correlations suggest that our use of (a) gives us a better proxy for (e) than (b) would, and thus (a) is second–best.

Finally, we examine the effects of our leniency programmes on verbal agreements to collude and on agreements to communicate, by estimating six additional probit models, each with the appropriate indicator as the dependent variable. Models 9–12 concern verbal agreements to collude. Model 11 uses the same set of right–hand–side variables and same sample (all Part–2 data) as in the probit for collusion (Table 4), for easy comparison. Model 12 is nearly the same, but includes some additional variables that might be associated with a verbal agreement: the total number and length of messages in the conversation, and the product of these with the Kyoto–session indicator (sentences in English and Japanese have different lengths, which might affect the number and total length of messages needed to agree on collusion). Two more models (9 and 10) use the entire data–set but exclude the leniency–treatment variables and the collusiveness variable (which was based on outcomes in Part 1). Models 13 and 14 involve agreements to communicate (both send a non–blank message), and correspond to Models 9 and 11 for collusive agreements. (We leave out the message–number and –length variables, due to high correlation with the dependent variable.) As before, these models were estimated using Stata (version 12) with robust standard errors. The results are displayed in Table 7.

[Table 7 about here]

The most obvious difference between these results and those in Table 4 is that here, the leniency programmes have a positive effect on both verbal agreements to collude and on agreements to communicate – and these effects are significant in the case of partial leniency – while we saw previously that neither programme had a significant effect on actual collusion (indeed, the sign was actually negative). Thus, if instead of using the HH indicator as our measure of a successful collusion attempt, we had used either a verbal agreement to collude or the "agreement to communicate" of previous competition policy experiments, we would have reached a different and potentially misleading conclusion: we would have thought that our leniency programmes *do* in fact increase collusion, so that combined with the reduced cartel stability we also observed, we might mistakenly have concluded that the net effect of the leniency programmes was ambiguous.

B: Sample instructions (from partial-leniency treatment, UK sessions)

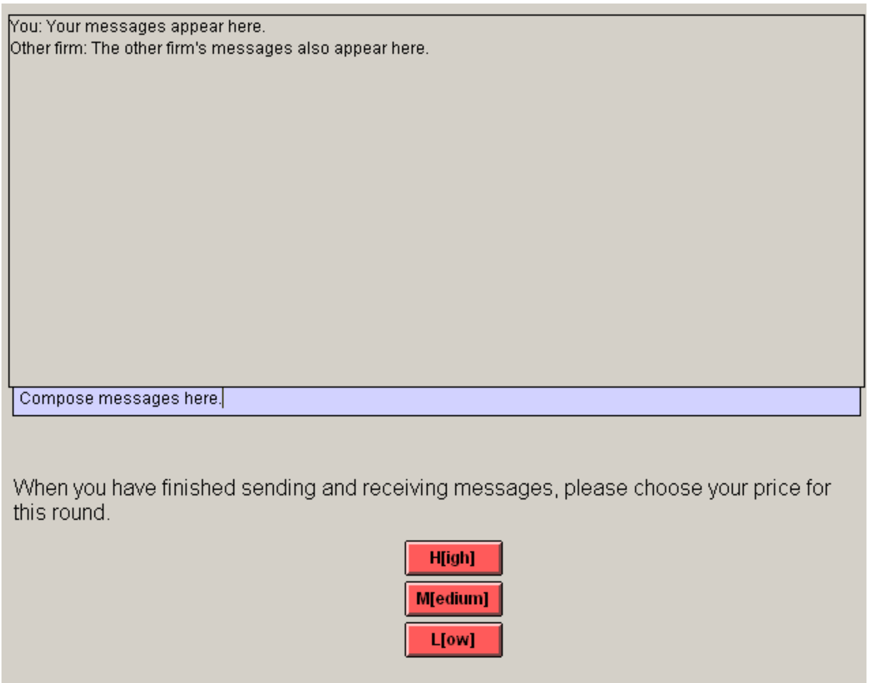
Instructions: first part of experiment

You are about to participate in a study of decision making. Please read these instructions carefully, as the amount of money you earn may depend on how well you understand them. If you have any questions, please feel free to ask the experimenter. We ask that you not talk with the other participants during the experiment.

These instructions are for the first part of the experiment. You will receive instructions for the second part after this part is finished. This first part is made up of several market games. You will be playing the role of a firm. At the beginning of each market game, you will be randomly matched to another participant, who also plays the role of a firm. You will be matched to the same firm for an entire market game, which will last for a number of rounds. In each round, you will choose the price of your product: High, Medium or Low. At the same time, the other firm will choose the price of its product. Your price and the other firm's price, together, determine your profits from the market, as shown in the table.

		Other firm price		
		High	Medium	Low
Your price	High	Your profit: £10 Other firm profit: £10	Your profit: £2 Other firm profit: £11	Your profit: £2 Other firm profit: £0
	Medium	Your profit: £11 Other firm profit: £2	Your profit: £4 Other firm profit: £4	Your profit: £2 Other firm profit: £0
	Low	Your profit: £0 Other firm profit: £2	Your profit: £0 Other firm profit: £2	Your profit: £0 Other firm profit: £0

Chat: In the first round of each market game, whilst deciding on your price, you have the opportunity to send and receive messages with the other firm. The chat portion of the computer screen is shown below.



To write a message, make sure your cursor is active in the narrow rectangle (where "Compose messages here" appears), and type normally using the keyboard. To send, press the Enter key on your keyboard. You may send as many or as few messages as you wish. However, we ask that you NOT send messages containing:

(a) personal or identifying information about yourself;
(b) physical threats.

Sent messages will appear in the box above the narrow rectangle, on your screen and on the screen of the other firm. If the other firm sends you a message, this will also appear in the box on your screen. Other participants in the experiment will not be able to see your messages, and you will not be able to see theirs. Once either you or the other firm has chosen a price, it will be impossible to send additional messages, but you can still view messages until you've chosen your price or the time available for messages runs out.

Competition law: The government has ruled that choice by both firms of the High price is anti-competitive pricing. It has also established a competition authority to discover such behaviour. In any round in which both you and the other firm choose a High price, there is an 8% chance that the competition authority will discover this. If your anti-competitive behaviour is discovered, the punishment is:

> (i) Both you and the other firm are fined £6 in the current round; this fine is subtracted from the profit you would have earned.
>
> (ii) Both you and the other firm will have your prices restricted in all remaining rounds of this market game (only Medium and Low prices will be available).

If your anti-competitive behaviour is not discovered, or if either you or the other firm (or both) does not choose the High price, there is no punishment.

Continuing or ending the market game: At the end of each round, there is a 20% chance that the market game will end, and an 80% chance that it continues for at least another round. If the game ends, a new market game will begin, and you will again be randomly matched to another participant. Also, punishments for anti-competitive behaviour disappear when a new market game begins. If the market game continues, you play another round, matched to the same other firm.

Sequence of play in a round: The sequence of play in a round is as follows.
(1) Sellers choose their prices. You will be able to send and receive messages at this time, if it is the first round of a market game.
(2) You are informed of the other firm's price.
(3) You are informed of your profit for the round, including any fine incurred. Also, if you and the other firm chose High prices, you are told whether your anti-competitive behaviour was discovered.
(4) The computer randomly determines whether the game will continue or end.

Payments: At the end of the experimental session, you will be paid based on your results. The computer will randomly select one round from each market game – in this part and the next part. You will be paid the total of your scores in those rounds, translated into real money at an exchange rate of 2 "lab-pounds" = £1. Payments are made privately and in cash.

Instructions: second part of experiment

The procedure in this part of the experiment is very similar to that in the first part. You will play several market games, each for a variable number of rounds. The participant matched with you will still be chosen randomly at the beginning of a market game, and remain the same until the market game has ended.

Leniency programme: The difference from the first part of the experiment is that the competition authority has instituted a leniency programme. In any round where both you and the other firm have chosen the High price, you are each given the opportunity to report your anti-competitive behaviour to the competition authority. If either you or the other firm does report, then the competition authority will (100% of the time) discover your anti-competitive behaviour, and both firms' prices will be restricted (to Medium or Low) in all remaining rounds of this market game. However, a firm that reports incurs a reduced fine of £3:
- If you report, then you will incur the reduced fine for anti-competitive behaviour.
- If the other firm reports but you do not, then you will incur the same fine as in the first part of the experiment, while the other firm incurs the reduced fine.
- If both firms report, then both incur the reduced fine.

Sequence of play in a round: The new sequence of play in a round is as follows.
(1) Sellers choose their prices. You will be able to send and receive messages at this time, if it is the first round of a market game.
(2) You are informed of the other firm's price. If you and the other firm chose High prices, you choose whether to report your anti-competitive behaviour to the competition authority.
(3) You are informed of your profit for the round, including any fine incurred. Also, if you and the other firm chose High prices, you are told whether your anti-competitive behaviour was discovered.
(4) The computer randomly determines whether the game will continue or end.