An Analysis of Corporate Leniency Programs and Lessons to Learn for US and EU Policies*

Eberhard Feess† and Markus Walzl‡

September 17, 2004

*We are grateful to Dominique Demougin, Lazlo Goerke, Uli Hege, Peter Jost, Gerd Mühlheusser, Hans-Bernd Schäfer, Horst Raff, Urs Schweizer and seminar participants in Auckland, Berlin, Evanston, Hamburg, Nancy, Maastricht and Zürich for helpful comments.

†Department of Economics, University of Aachen (RWTH), Templergraben 64, D-52062 Aachen, Germany. Email: feess@rwth-aachen.de.

‡Department of Economics and METEOR, University of Maastricht, P.O. Box 616, NL-6200 MD Maastricht, The Netherlands. Email: m.walzl@algec.unimaas.nl.

Abstract

We develop a model to compare leniency programs for cartel behavior as enacted, for instance, in the USA and the EU. Although all such programs are based on the idea that the expected fine ex ante can be increased by granting fine reductions for self-reporting firms, they differ considerably in how this basic idea is legally implemented. Differences include the fine reductions granted for first and second self-reporters, the role of the amount of evidence provided, and the impact of whether the case is already under investigation. We elaborate on the role of asymmetric information to derive the optimal degree of leniency, and we make use of our findings to compare the programs in the US and the EU.

JEL-Classification: D62, D82, H50, K42
Keywords: self-reporting, optimal law enforcement, criminal teams, leniency programs.

1 Introduction

Legal Situation and Motivation In February 2002, the Commission of the European Union has substantially revised its law enforcement against cartels. In particular, the new policy follows the path of the leniency program enacted in the USA in 1993 and stresses the opportunity of fine reductions for self-reporting cartel members. A leniency program for cartels and other illegal teams seems is appealing both from a theoretical and an empirical perspective. Theoretically, a reduction of the fine for the first self-reporter and high fines for all other team members induces an incentive to be the first one who comes forward (often described as "race to the courtroom" by legal scholars and closely related to the Prisoner's Dilemma). Given that all members are identical, each of them will win the race with probability $1/n$, but will pay a high fine with probability $\frac{n-1}{n}$. Hence, reducing the fines for self-reporting firms can increase the expected fine, and thereby deterrence. Even though the US program is successful in the sense that it has led to a tenfold increase in the number or self-reporting cartel members from 1995 to 2000, the European Commission did not adapt the widely accepted US program but developed a significantly different policy. The most striking differences between the US and the EU program are summarized in table 1:[1]

	US Program	*EU Program*
(i)	The first self-reporting firm gets *full amnesty.*	Usually only a *partial amnesty* is granted for the first firm.
(ii)	The fine reduction (thus) *does not depend* on the evidence provided.	Fine reductions *depend* on the evidence provided.
(iii)	The second self-reporting firm gets *no fine reduction* at all.	Fine reductions also for *other* self reporting firms.
(iv)	*Full immunity* can also be granted if the case is already under investigation.	Maximum fine reduction of *50 percent* if the case is already under investigation.

Table 1: Distinctive features of the two leniency programs.

[1]For both programs, leniency rules only apply if some requirements are fulfilled. For instance, the applying firm must not be the leader or initiator of the cartel. For details see European Community (2002) and US Department of Justice (2004).

Inspecting the differences listed in Table 1, the US-program appears much more radical in implementing a "winner-take-it-all-approach" - the first self-reporting firm can get full immunity even if the case is already under investigation, whereas the second firm is heavily penalized. This approach seems to have the advantage to set maximum incentives to be the first one who provides evidence (i.e. inducing maximum tension within the cartel), which is emphasized by Hammond, Director of Criminal Enforcement of the Antitrust Division: "If you are second, even if only by a matter of a few hours, which has happened on a number of occasions, the second firm and all of its culpable executives will be subject to full prosecution" (Hammond 2000, p.5). Hence, self-reporting incentives for a *given* crime rate should be higher in the US program which is in line with a superficial look at the data. In fact, while the average number of self-reporting cartel members in the US amounts to 3-4 per month, it turned out to be considerably smaller in the EU (approximately 1-2 per month).

The EU-program, however, pays more attention to the specific facts of a case at hand - fine reductions may be granted for all firms, and these fine reductions depend on the amount of evidence produced, and on whether the case is already investigated. Thereby, the heterogeneity among cartel members and the different values of their reports can be taken into account. This may help to reduce the incentive to violate the law ex ante, and to enhance the willingness to provide evidence even if one firm has already self-reported.

Other countries like Japan offer leniency programs similar to the US, while several EU member states as the Netherlands which enacted a program in May 2004 follow the approach taken by the EU. The German program can be interpreted as a mixture of the EU and the US approach because only the first self-reporter can get fine reductions (as in the US scheme), but full immunity is not automatically granted which resembles the EU system. In fact, the EU and US programs can be taken as polar cases of implemented policy designs. Hence, a careful comparison of the EU and US system will not only prove useful for a coordination of international cartel prosecution but also for a harmonization of the competition policy within the EU.

Against this background, the objective of our paper is twofold: First, we derive the optimal leniency program in a framework that takes important factors concerning the incentives to violate and to self-report into account. Second, our results will be used to compare the existing programs in the EU and in the USA.

Framework We develop a model that captures the differences of the two programs described above in the simplest possible way. We assume that there are only two firms which agree upon collusive behavior whenever their team-benefit is (weakly) above their aggregated expected fines.[2] The following ingredients are required to distinguish between the two programs:

(a) The two members of the team differ with respect to the degree of evidence they can provide about their accomplice. This assumption is necessary to account for part (i) and (ii) of the programs described above.

(b) There are two self-reporting stages: A criminal may self-report *before* his case is detected, or *after* he has been detected, but not yet convicted. This distinction is important for two reasons: first, it is required to compare parts (iv) of the two programs. Second, it will turn out that the strategic interactions between the two criminals differ considerably in the two stages. While it will turn out that the pre-detection stage can best be described as a race to the courtroom, the second stage is simply the classical prisoner's dilemma.

(c) If the authorities were perfectly informed, then the task of designing an optimal self-reporting scheme would be trivial. Obviously, the court should assign the maximum self-reporting fine to the first self-reporter that only just ensures self-reporting with certainty, and impose the maximum fine to the other firm(s). This would lead both to a self-reporting frequency of 100 percent and to maximum deterrence. To be realistic (and to analyze part (iii)), we need to introduce some kind of asymmetric information. In our model, the probability of being detected does not only depend on the authority's investigation effort, but also on team-specific attributes ("types") that are private information to the violators. We assume that these types are learned only *after* a violation has been committed, and we will refer to this as *ex post* asymmetric information. For instance, the team might learn after the violation that it has acted careless in some sense, or that a (potential) new market entrant will not be willing to join the cartel. Since self-reporting occurs only if the type-specific detection probability is known to be high, the possibility to self-report leads to an option value that ceteris paribus reduces the expected fine to be paid. This disadvantage of leniency programs has almost been completely neglected in the economic literature

[2]Qualitatively, our results would not change if we assumed instead that each member of the team must weakly benefit from collusion. Implicitly, our assumption allows for side payments ex ante.

(see the literature review below) but seems to be important to understand why a full immunity for the first self-reporting firm may not come without costs. Furthermore, the two self-reporting opportunities are not based on the same information set which introduces a non-trivial relation between pre-detection stage and conviction stage.

Relation to the literature Most fundamentally, our analysis builds on the self-reporting-literature in the single violator case pioneered by Kaplow and Shavell (1994). They have shown that, if the sanction for a reported violation is infinitesimally smaller than the expected fine from being detected, all violators prefer to self-report. The advantage is that investigation costs can be saved as only non-reporting individuals need to be examined. Furthermore, an early detection might reduce social costs since countermeasures can be taken right away (see Innes 1999). Innes (2000) and also Livernois and Mc Kenna (1999) in a somewhat different context, demonstrate that only partial self-reporting occurs if violators have different detection probabilities ex-ante. Feess and Heesen (2002) extend the analysis to ex post asymmetric information in the sense described above.

An important point to note is that, in the single violator case, self-reporting always (weakly) *decreases* deterrence. Hence, leniency can only be optimal if investigation costs can be saved or if an early detection is socially beneficial. The situation is quite different for criminal teams, because strategic interactions between the team members in the self-reporting stage can be used to *increase* expected fines. Starting with Motta and Polo (MP 2003), there is now a small literature on self-reporting schemes with strategic interaction. MP 2003 consider an infinitely repeated collusion game between firms. In their model, a collusion may break down since partners may cheat on each other (for instance by setting lower prices than agreed upon). However, a self-reporting scheme may also lead to more collusion, because low self-reporting fines provide credible threats to reveal the collusion in case the partner cheats. MP 2003 show that in the authority's optimal policy, the former effect always dominates the latter such that leniency programs (weakly) improve social welfare. Another result (that turns out to be different to our findings) is that in the authority's optimal policy no violator self-reports *before* an investigation has been started as violation and self-reporting decision are based on the same information set.

Buccirossi and Spagnolo (BS 2001) also focus on potential cheating and show that moderate fines for single self-reporters may then provide credible threats to unravel the deal in case the accomplice cheats. However, they also confirm the earlier finding by Spagnolo (2000) that collusion can completely be deterred if the authority subsidizes self-reporters highly enough. This is not surprising, because high enough subsidies induce full self-reporting even if the cartel has still some future benefits to expect from their illegal behavior. In fact, when comparing our paper to MP 2003 and to BS 2001 in greater detail in section 4, we will point out that some kind of asymmetric information is required to understand why subsidies may not be the holy gral and to see that existing programs do not only differ in a marginal way.

In a preceding paper (Feess and Walzl (2004)), we derive the optimal fine structure under the assumption that the team members can decide cooperatively upon self-reporting by incurring some transaction costs. We then show that a full amnesty for the first self-reporting firm may not be optimal as it enhances the incentive for cooperative behavior. While this introduces a drawback of leniency programs not analyzed in the previous literature, the contribution clearly suffers from the ad hoc-assumption of transaction costs which are not endogenously derived in an asymmetric information framework.

Summing up, while there is some literature that justifies leniency programs in general (but neglects its drawbacks), our paper seems to be the first one that combines the ingredients (a)-(c) described above in order to compare existing programs and to draw clear-cut conclusions for policy design.

The remainder of the paper is organized as follows: Section 2 develops the model. Following backwards induction, section 3 analyzes equilibrium strategies (including the authority's optimal policy). Section 4 discusses merits and shortcomings of the model and draws conclusions with respect to the existing leniency programs.

2 The Model

In our model, there are two firms H and L which form a cartel whenever their team benefit B is above the aggregated expected fine denoted F.[3] B is private information and distributed with continuous density $g(B)$. The difference between H and L refers to the amount of evidence that can be provided about the partner if a firm self-reports. Specifically, we assume that

[3]Hence, we assume that B is sunk. For the significance of this an other assumptions see section 4.

if firm H (L) self-reports, than firm L (H) can be convicted with probability h (l) where $h > l$. This covers feature (a) as discussed in the introduction. Types H and L are common knowledge. If none of the team members self-reports, the case is detected with probability $p\theta$ where $p \in (0, 1)$ is the percentage of cases investigated, and $\theta \in (0, 1)$ is a team specific parameter distributed with continuous density $z(\theta)$. θ is only learned by the team *after* the violation decision and is private information (see feature (c)).
If the case is detected, however, the amount of evidence is not necessarily sufficient to actually convict the team which happens only with probability $t \in (0, 1)$. After detection, violators face a second self-reporting opportunity. For simplicity, we assume that if a violator self-reports, the authority relies on the evidence provided and does not engage in further investigation.[4] Furthermore, we assume that $l > t$ (as self-reporting would otherwise reduce the conviction probability which would not make sense) and that both p and t are exogenously given. [5]
The maximum fine the authority can impose is denoted s. Clearly, this fine has to be paid by those violators who are convicted either due to the accomplices report or the governments effort, not convicted violators, however, do not pay anything. Hence, the authority is left with the assignment of fines for convicted violators for all types (H, L) and self-reporting opportunities (before and after detection) depending on wether they have self-reported or are convicted (both due to the evidence provided by a team member or the effort spent by the authority). The game considered can now be described as follows:

- Stage 1 ("Authority's policy"): Authority commits to a vector R of (self-reporting) fines (i.e. assigns fines to every final node of the game tree where a violator is indeed convicted).

- Stage 2 Nature determines and individuals learn their private benefits B from cartel behavior.

- Stage 3 ("Violation decision"): Individuals decide upon violating the law (i.e. formation of a criminal team if $B \geq F$).

[4] Instead, we could define τ as the *additional* evidence provided by self-reporting, and then derive l as $l = \tau + t - t\tau > t$.

[5] This has the technical advantage that the authority minimizes social costs simply by maximizing the expected fine F. It can be shown that our qualitative results carry over to endogenous monitoring (see Feess and Walzl (2003)).

- Stage 4 Nature determines and violators learn their team-specific detection probability θ.
- Stage 5 ("Pre-detection stage"): Violators decide separately and non-cooperatively upon self-reporting.
- Stage 6 Non-reported cases (i.e. teams) are detected with probability $p\theta$.
- Stage 7 ("Conviction stage"): Detected violators decide separately and non-cooperatively upon self-reporting.
- Stage 8 Detected, but non-reported cases lead to conviction with probability t.

In the following, we will derive (the) subgame perfect equilibrium (SPE) of the game considered by backwards induction. Thereby, we can restrict attention to stages 1, 3, 5 and 7 as all other stages are either nature's moves or part of the authority's policy committed to in stage 1. With respect to stage 1, recall that the authority's objective is to reduce the number of criminal acts given by $\pi = \int_{\widetilde{B}=F}^{\overline{B}} g(B)dB$ where $\widetilde{B}$ is the borderline type who only just violates. Hence, the authority only has to maximize expected fines F. Furthermore, stage 3 is obviously a simple decision problem as both B and F are known by the potential violators is stage 3. The self-reporting decisions in stage 5 and 7 result in non-cooperative stage-games between the members of the criminal team. We will discuss these games in detail in the next section. In general, these games will prove to have multiple equilibria (in pure strategies). As equilibrium selection criterion, we assume throughout that the team members play the (pure strategy) Nash Equilibrium (NE) that minimizes their aggregated expected fines (Kaldor-Hicks dominance). In fact, all of our conclusions would otherwise be reinforced, because the Kaldor-Hicks criterion *minimizes* the deterrence of the self-reporting scheme we propose, so that the policy will be even more successful when using other equilibrium selection concepts. Furthermore, we assume without loss of generality that every individual self-reports in case of indifference.

3 Equilibrium Analysis

3.1 Conviction stage (stage 7)

We can now start with the self-reporting decision in stage 7 under the assumption that the team has been detected (note that stage 7 will not be reached if the team has not been detected, or if a member has already self-reported). As team members are interrogated separately, let us define c_i^k, $i = H, L$; $k = 1, 2$ as he self-reporting fine in the conviction stage if the cartel member i self-reports, and if there are k self-reporters (i.e. the fines imposed on player i explicitly depend on his self-reporting decision *and* on the action of his counterpart). The respective normal form game for the members of the criminal team is depicted in Table 2 (where the first fine refers to the row-player H):[6]

	S_L^c	N_L^c
S_H^c	c_H^2/c_L^2	c_H^1/hs
N_H^c	ls/c_L^1	ts/ts

Table 2: Expected fines in the conviction stage

If both self-report, they pay fines c_H^2 and c_L^2, respectively. If only type $L(H)$ comes forward, he pays $c_{L(H)}^1$ and the evidence provided leads to conviction of the counterpart (and maximum payments of s) with probability h or l, respectively. If no-one self-reports, the game enters stage 8 and hence expected fines are ts. When turning to the authority's optimal policy in stage 1, we will demonstrate that in the SPE, there is a unique NE in stage 7 (the conviction stage), and we define f_i^c, $i = H, L$ as the respective expected individual fines. No decisions are to be made in stage 6, so that we can directly proceed to stage 5.

3.2 Pre-detection stage (stage 5)

We can now turn to the self-reporting decision before the team has been detected. There is an important difference to the situation in the conviction stage. In the conviction stage, both members of the cartel are already known

[6]The team member's actions at this stage are given by X_i^c with $i = H, L$ and $X \in \{S, N\}$ where $X = S$ if a player self-reports and $X = N$ if he does not come forward (superscript c denotes the conviction stage).

and can be interrogated separately. Hence, the authority can impose different self-reporting fines depending on the accomplice's behavior, i.e. differentiate between c_i^1 and c_i^2. In the pre-detection stage, this is not possible, because, even if both *want* to self-report, only one can win the race to the courtroom (see e.g. the comment made by US-official Hammond quoted above). Hence, team member's action S now refers to the *attempt* to self-report. Assuming that each of them wins the race to the courtroom with equal probability of 0.5, this leads to the expected fine structure shown in Table 3.[7]

	S_L^d	N_L^d
S_H^d	$\frac{1}{2}(r_H + ls) / \frac{1}{2}(r_L + hs)$	r_H/hs
N_H^d	ls/r_L	$p\theta f_H^c / p\theta f_L^c$

Table 3: Expected fines in the pre-detection stage

In table 3, r_i, $i = H, L$ is the self-reporting fine in the pre-detection stage, and f_i^c is the expected fine for detected violators. If only one, H, say, wants to self-report, he pays fine r_H. His accomplice will then be convicted with probability h and pays in this case the maximum fine s. This explains the expected outcome in the $\left(S_H^d/N_L^d\right)$-action combination, and analogously in the $\left(N_H^d/S_L^d\right)$-action combination. If both want to self-report, each of them pays his self-reporting fine r_i in case he wins. Otherwise, he is convicted by the partner's evidence with probability l or h, respectively, which explains the expected outcome in the $\left(S_H^d/S_L^d\right)$-combination. If none of them self-reports, the team is detected with probability $p\theta$. The game then enters stage 7, which leads to expected fines of f_i^c as explained above.
To analyze possible equilibria, first note that $\left(N_H^d, N_L^d\right)$ is a NE if and only if $r_i > p\theta f_i^c \ \forall i \in \{H, L\}$. Hence, player i deviates from $\left(N_H^d, N_L^d\right)$ if $\theta \geq \frac{r_i}{pf_i^c}$. Depending on r_i and on f_i^c (and hence on the authority's policy), we have $\frac{r_H}{pf_H^c} \lesseqgtr \frac{r_L}{pf_L^c}$. To exclude the action-profile $\left(N_H^d, N_L^d\right)$ as a (potential) NE, it suffices that $\widehat{\theta} \equiv \min\left(\frac{r_H}{pf_H^c}, \frac{r_L}{pf_L^c}\right) \leq \theta$. Throughout the paper we will refer to the type of player who destroys the (N_d^H, N_d^L)-NE as the pivotal player - more formally:

Definition 1. *Player $i \in \{H, L\}$ is pivotal for (N_d^H, N_d^L) being a NE if $\widehat{\theta} \equiv \min\left(\frac{r_H}{pf_H^c}, \frac{r_L}{pf_L^c}\right) = \frac{r_i}{pf_i^c}$.*

[7] The team member's actions at this stage are given by X_i^d with $i = H, L$ and $X \in \{S, N\}$ as defined above (superscript d denotes the (pre-)detection stage).

Then, we can describe the possible equilibria as follows:

- **Case 1**: $\theta < \widehat{\theta}$: Then, $\left(N_H^d, N_L^d\right)$ is a NE.
- **Case 2**: $\frac{r_H}{pf_H^c} \leq \theta < \frac{r_L}{pf_L^c}$. In this case, $\left(N_H^d, N_L^d\right)$ is no NE as player H has an incentive to self-report (S_H^d) (i.e. H is pivotal in the sense of Definition 1). Then, (S_H^d, S_L^d) is a NE if $\frac{1}{2}(r_L + hs) \leq hs$ or, equivalently, if $r_L \leq hs$. Otherwise, (S_H^d, N_L^d) is a NE.
- **Case 3**: $\frac{r_L}{pf_L^c} \leq \theta < \frac{r_H}{pf_H^c}$. In this case, $\left(N_H^d, N_L^d\right)$ is no NE as player L is pivotal and has an incentive to self-report (S_L^d). Then, (S_H^d, S_L^d) if $r_H \leq ls$. Otherwise, (N_H^d, S_L^d) is a NE.

The distinction between cases 1-3 will be used when deriving the authority's optimal policy. In stage 4, the violators learn their team-specific component θ of the detection probability θp. In stage 3, they from a cartel if their joint benefit B is weakly above their (joint) expected fine F. In stage 2, violators learn B. Hence, we can directly turn to the authority's optimal policy.

3.3 The Authority's Optimal Policy

Fine Structure in the Conviction Stage Since a violation takes place whenever $B \geq F$, the authority maximizes the expected aggregated fine F for any t and p given assigning the self-reporting scheme $R = (r_H, r_L, c_H^1, c_L^1, c_H^2, c_L^2)$. With respect to the fines imposed in stage 7, we get

Lemma 1. *(i) In the authority's optimal policy* $\min_{i=H,L}(c_i^1) \leq ts$, $c_H^2 = ls$ *and* $c_L^2 = hs$. *(ii) When the conviction stage is reached, both violators self-report with probability one, and the expected individual fines are* $f_H^c = ls$ *and* $f_L^c = hs$, *respectively.*

Proof. Recall that we adopt Kaldor-Hicks dominance as equilibrium selection criterion in case of multiple equilibria. If $N^c \equiv (N_H^c, N_L^c)$ is implemented as equilibrium, then the individual fine is $f_i^c(N^c) = ts \; \forall i \in \{H, L\}$. To destroy the N^c-equilibrium, the authority has to set $\min_{i=H,L}(c_i^1) \leq ts$. This given, $S^c \equiv S_H^c/S_L^c$ is the unique NE if $c_H^2 \leq ls$ and $c_L^2 \leq hs$ hold, so that imposing $c_H^2 = ls$ and $c_L^2 = hs$ is optimal. Aggregated fines are then $F^c(S^c) = s(l+h)$. If $c_H^2 > ls$ or $c_L^2 > hs$, then the asymmetric decisions S_H^c/N_L^c and S_L^c/N_H^c become NE, leading to maximum aggregated fines $F^c(S_H^c/N_L^c) = hs + ts$ and

$F^c(S_L^c/N_H^c) = ls + ts$, respectively (for higher fines, N^c will be played). From $t < l$, it follows that $F^c(\cdot)$ is maximized if the authority induces self-reporting by both parties, which allows for maximum fines hs and ls, respectively. ■

The intuition for Lemma 1 follows immediately from the fact that, in the conviction stage, the authority can resemble the prisoner's dilemma by differentiating fines according to the accomplice's behavior. By setting either c_H^1 or c_L^1 (weakly) smaller than ts, it is ensured that $N^c \equiv (N_H^c, N_L^c)$ can be no NE. This given, the authority prefers to implement $S^c \equiv (S_H^c/S_L^c)$ as the unique NE in stage 7, because the fines if only one self-reports need to be low to destroy N^c as a potential NE. And since the maximum fines which support S^c as a NE are $c_H^2 = ls$ and $c_L^2 = hs$, the Lemma follows.

Fine Structure in the Pre-Detection Stage Determining the optimal fine structure in the pre-detection stage turns out to be more complicated. However, the following two insights are helpful: First, we know from the conviction stage that $f_H^c = ls$ and $f_L^c = hs$ if the conviction stage is reached (see Lemma 1). This allows us to re-write Table 3 as

	S_L^d	N_L^d
S_H^d	$\frac{1}{2}(r_H + ls) / \frac{1}{2}(r_L + hs)$	r_H/hs
N_H^d	ls/r_L	$p\theta ls/p\theta hs$

Table 4: Sub-game perfect expected fines in the pre-detection stage.

which will be easier to handle. And second, the only way to eliminate the $N^d \equiv \left(N_H^d, N_L^d\right)$ *with certainty* is to set either r_H or r_L to zero, as we have otherwise a θ small enough such that $\theta < \widehat{\theta} = \min\left(\frac{r_H}{pf_H^c}, \frac{r_L}{pf_L^c}\right)$. This given, the authority has to make the following three decisions:

- **Decision 1:** It has to decide whether player H or player L shall be pivotal in in destroying the N^d equilibrium. Since we already know that $f_H^c = ls$ and $f_L^c = hs$, it follows that player H will be pivotal if and only if $\frac{r_H}{pls} < \frac{r_L}{phs}$ which can be simplified as $\frac{r_H}{l} < \frac{r_L}{h}$.[8] Furthermore, note that this implies that player L is pivotal if $r_H = r_L$ as $l < h$ as he faces higher expected fines in case of conviction.

[8]Note that this simplification was not possible in section 3 because $f_H^c = ls$ and $f_L^c = hs$, which is part of the authority's optimal policy in stage 1, had to be derived before.

- **Decision 2:** Once a pivotal player (i, say) has been chosen, the authority faces the following trade-off: the lower r_i, the higher the probability that the unwarranted non-reporting situation N^c is no Nash Equilibrium. But on the other hand, the lower r_i, the lower is ceteris paribus the expected aggregated fine F in case self-reporting occurs. This trade-off determines r_i.

- **Decision 3:** The self-reporting fine for the non-pivotal player (j, say) r_j will then determine whether S^d or an asymmetric equilibrium will be played in case that N^d is no NE.

To solve the problem, let us *assume* for the moment that player H is pivotal, i.e. that $\frac{r_H}{l} < \frac{r_L}{h}$ and that $\theta > \widehat{\theta}$. Then, (S_H^d, N_L^d) is an NE if and only if $r_L > hs$, otherwise S^d is a NE. *If* the authority prefers the S^d-NE, it will clearly set $r_L = hs$, i.e. the maximum fine that only just supports (S_H^d, S_L^d) as an equilibrium. Inspecting Table 4 above shows that the aggregated expected fine in S^d is thus $F(S^d) = \frac{1}{2}(r_H + ls) + \frac{1}{2}(r_L + hs) = \frac{1}{2}(r_H + ls) + \frac{1}{2}(hs + hs) = \frac{1}{2}(r_H + ls) + hs$. On the other hand, for any $r_L > hs$, the aggregated expected fine in the (S_H^d, N_L^d)-NE is $F(S_H^d, N_L^d) = r_H + hs$ (see again Table 4). Comparing the two outcomes shows that

$$F(S^d) - F(S_H^d, N_L^d) = \frac{1}{2}(r_H + ls) + hs - (r_H + hs) = \frac{1}{2}(ls - r_H) > 0$$

where $ls - \frac{1}{2}r_H > 0$ follows from the assumption that player H is pivotal.[9] Hence, *if* the authority chooses the fines such that player H is pivotal, either S^d or N^d, but no equilibria where only one wants to self-report are possible in the authority's optimal policy.
Analogously, we can show that $F(S^d) - F(N_H^d, S_L^d) > 0$ so that we can summarize our results as follows.

Lemma 2. *(i) If H is pivotal, then $r_L = hs$. (ii) If L is pivotal, then, $r_H = ls$. (iii) In Stage 5, the NE for the stage game is given by $S^d = (S_H^d, S_L^d)$ if $\theta > \widehat{\theta}$ and $N^d = (N_H^d, N_L^d)$ otherwise.*

Proof. Part (i). Has been proven above. *Part (ii).* Proceeds analogously. *Part (iii).* Is implied by (i) and (ii). ■

[9] From $\frac{r_H}{pls} < \frac{r_L}{phs}$ together with $r_L = hs$ in the (S_H^d, S_d^L)-NE we get $r_H < ls$.

The reason for Lemma 2 is that the incentive to self-report is higher when the partner self-reports, because the probability of being convicted then increases from θph or θpl to h and l, respectively. Hence, implementing S^d instead of an equilibrium where only one self-reports allows higher fines without establishing N^d as NE. Hence, we are left with the question which player will be pivotal in the authority's optimal policy (i.e. $\frac{r_H}{l} < \frac{r_L}{h}$ or not). Our findings with respect to this are summarized in

Lemma 3. *In the authority's optimal policy, the high evidence provider H is pivotal, i.e.* $\frac{r_H}{l} < \frac{r_L}{h}$.

Proof. Assume for the moment that player L were pivotal. Recall that $r_H = ls$ in this case (see Lemma 2). Define r_L^* as the optimal r_L if L were pivotal, so that the borderline type $\widetilde{\theta}_L$ who only just self-reports would be given by $\widetilde{\theta}_L = \frac{r_L^*}{phs}$. Taking Lemma 2 into account, it follows that the aggregated expected fine if L were pivotal would be given by[10] (see Table 4)

$$F_L^* = \left(\frac{1}{2}\left(r_L^* + hs\right) + ls\right) \int_{\widetilde{\theta}_L}^{1} z(\theta) d\theta + p\left(h + l\right) s \int_{0}^{\widetilde{\theta}_L} \theta z(\theta) d\theta \qquad (1)$$

Now assume that player H is pivotal and assume that the authority *sub-optimally* chooses r_H such that $\widetilde{\theta}_H = \widetilde{\theta}_L$ which requires $\frac{r_H}{l} = \frac{r_L^*}{h}$. Then, the expected aggregated fine amounts analogously to

$$F_H = \left(\frac{1}{2}\left(r_H + ls\right) + hs\right) \int_{\widetilde{\theta}_L}^{1} z(\theta) d\theta + p\left(h + l\right) s \int_{0}^{\widetilde{\theta}_L} \theta z(\theta) d\theta \qquad (2)$$

where $r_H = \frac{lr_L^*}{h}$ yields

$$F_H = \left(\frac{1}{2}\left(\frac{lr_L^*}{h} + ls\right) + hs\right) \int_{\widetilde{\theta}_L}^{1} z(\theta) d\theta + p\left(h + l\right) s \int_{0}^{\widetilde{\theta}_L} \theta z(\theta) d\theta. \qquad (3)$$

It follows that

$$F_H - F_L^* = \frac{1}{2} \int_{\widetilde{\theta}_L}^{1} z(\theta) d\theta \left(s\left(h - l\right) - r_L^* \left(\frac{h - l}{h}\right)\right] > 0 \qquad (4)$$

[10] F_L^* (F_H^*) is the maximum (and hence optimal) aggregated expected fine the authority can implement *if* player L (H) is pivotal.

which always holds as $r_L^* < hs$ because $\left(S_H^d, S_L^d\right)$ would otherwise be no NE. And by definition of optimality, it follows that $F_H^* \geq F_H > F_L^*$. ■

The fact that the high evidence provider H in general gets a substantial fine reduction and will (hence) be pivotal in the authority's optimal policy is not trivial because there are countervailing effects. On the one hand, if H is pivotal, the equilibrium fine assigned to his counterpart ($r_L = hs$) is higher than for a pivotal L (which only results in $r_H = ls$). This clearly increases expected (aggregated) fines. On the other hand, if H is pivotal, the unwarranted N^d-action combination is (ceteris paribus) more often an equilibrium (as $\frac{r}{hs} < \frac{r}{ls}$), which is obviously a disadvantage. However, as shown in the proof to Lemma 3, the first effect always dominates. Then, taking the three Lemmata together, we can summarize the authority's optimal policy in

Proposition 1. *In the authority's optimal policy, (i) $c_H^2 = ls$ and $c_L^2 = hs$. (ii) $r_L = hs$. (iii) $r_H = argmax_{r_H} F_H(\cdot)$. (iv) $r_H < ls$ (if $z(1) \neq 0$). (v) A sufficient condition for $r_H > 0$ is that $z(\theta)$ satisfies $z(0) = 0$. (v) A sufficient condition for $r_H = 0$ is that $z'(\theta) \leq 0$.*

Proof. Part (i). See Lemma 1. *Part (ii).* See Lemma 2 and Lemma 3. *Part (iii)* The authority chooses r_H as to maximize expected fines

$$F_H = \left(\frac{1}{2}(r_H + ls) + hs\right) \int_{\widetilde{\theta}_H}^{1} z(\theta)d\theta + p(h+l)s \int_0^{\widetilde{\theta}_H} \theta z(\theta)d\theta$$

where the optimization program that implicitly defines the optimal r_H is given in the Appendix. *Part (iv)* Note that $r_H = ls$ can never be optimal, as $\frac{\partial F_H}{\partial r_H} = 0$ and $\frac{\partial^2 F_H}{\partial (r_H)^2} > 0$ at $r_H = ls$. *Part (iv).* A sufficient condition for $r_H \in (0, 1)$ is $z(0) = 0$ as this implies that $\frac{\partial F}{\partial r_H} > 0$ for $r_H = 0$. *Part (v).* A sufficient condition for $r_H = 0$ is $z'(\theta) \leq 0$ as this implies that $\frac{\partial^2 F}{\partial (r^H)^2} > 0$ $\forall r_H \in (0, ls)$ which excludes an interior solution. ■

Part (i), which is already known from Lemma 1 says that if the team is detected without having self-reported, then the fines are equal to the maximum fine s, weighted with the probability that one can be convicted by the partner's testimony. Part (ii) expresses that the low evidence provider L pays the same fine hs if he self-reports already in the pre-detection stage, so that he does not benefit from self-reporting at all. Conversely, the high evidence provider H's fine derived by $\frac{\partial F_H}{\partial r_H} = 0$ (see Part (iii)) for early self-reporting is *strictly* lower than his fine for a report in the conviction stage where $r_H < ls$

is ensured by the fact that type H is pivotal. Parts (iv) and (v) of the Proposition refer to the question whether a full amnesty should be granted to the high evidence provider if he confesses already in the pre-detection stage. As expressed in the Proposition, this depends on the distribution of the team-specific detection probability θ. $z(0) = 0$ ensures that all violators know that they will be detected with positive probability, so that a full amnesty cannot be optimal as even positive fines ensure full self-reporting. On the other hand, a decreasing density ($z'(\theta) \leq 0$) means that there is a large number of violators that will only come forward if a full amnesty is granted. In contrast, full immunity can never be optimal for type L who will pay $r_L = hs$ if he wins the race to the courtroom, and this holds regardless of the density function $z(\theta)$.

Comparative Statics An important point with respect to the leniency programs in the USA and the EU is how the self-reporting fines depend on the evidence provided. We summarize our results as follows.

Corollary 1. *(i)*$\frac{dc_H^2}{dh} = \frac{dc_L^2}{dl} = 0$. *(ii)* $\frac{dc_H^2}{dl} = \frac{dc_L^2}{dh} = s > 0$. *(iii)* $\frac{dr_L}{dh} = s > 0$. *(iv)* $\frac{dr_L}{dl} = 0$ *(v)*$\frac{dr_H}{dh} \leq 0$. *(vi)* $\frac{dr_H}{dl} \geq 0$.

proof: see appendix.

In the conviction stage, the equilibrium fines are $c_H^2 = ls$ and $c_L^2 = hs$, so that the self-reporting fines are independent of the own evidence provided, but are increasing in the evidence provided by the partner. The reason is that the fines are chosen such that both types only just self-report, and the own incentive to self-report is (for t and s given) a monotone increasing function of the conviction probability if the partner self-reports.

In the pre-detection stage, we have $r_L = hs$, so that the fine for the low-evidence provider is again independent of his own evidence, and increasing in the evidence provided by his accomplice. Both results follow intuitively from the fact that type H is pivotal for the self-reporting decision, so that type L's self-reporting fine is independent of whether he reports before or after the team has been detected.

The fact that type H's self-reporting fine is decreasing in his evidence provided ($\frac{dr_H}{dh} < 0$) follows from the fact that his testimony is more valuable if h is high. The higher h, the higher is type L's expected fine hs if H self-reports, which will hence be made more attractive by reducing r_H. The higher l, however, the higher is ceteris paribus type H's incentive to self-report, because his expected fine when the team is detected without having self-reported increases (recall that $c_H^2 = ls$). Stated differently, if l increases, a higher r_H can be chosen without destroying H's self-reporting incentive. Hence, the pivotal type's fine depends (non-trivially) on the evidence provided by *both* types.

4 Discussion

Significance of assumptions In this section, we will first review the plausibility and the significance of our assumptions. Together with our findings, this will finally enable us to draw some conclusions for the existing leniency programs.

First, one might question that the detection probability p and the conviction probability t are exogenously given. In reality, p and t reflect how much effort is spend in law enforcement, and are hence choice variables. We analyze endogenous p and t in a working paper (Feess and Walzl (2003)). In the conviction stage, the result will then be that the authority sets $\min(c_i^1) = 0$ in order to ensure full self-reporting even for $t \to 0$ which follows simply from the prisoner's dilemma structure in the conviction stage. In the detection stage, the optimal p is derived analogously to r_H, but this does not lead to economically interesting insights. Hence, treating p and t exogenous seems to be justified for ease of exposition.

Second, one might wonder why we introduce *ex post* asymmetric information with respect to θ. Of course, some kind of asymmetric information (besides the benefit B) is necessary to introduce partial self-reporting (i.e. a non-trivial relation between different self-reporting stages). However, one might ask why we do not assume that θ is already known ex ante (as assumed by Innes (1999)) for the single violator case. Modelling ex post asymmetric information has three advantages: first, it seems realistic that the cartel gets new (private) information after the collusion has started. Second, knowing θ ex ante means that the violation decision depends on a *vector* of private signals (on B and θ) which drives the analysis more cumbersome, while, third, ex-ante asymmetric information implies that the team knows right from the beginning whether it self-reports which neglects the option value of a leniency program and, as a consequence, biases the analysis.

Third, we restrict the problem of private information to the benefits and to the detection probability, and do not extend it to the conviction stage by

introducing an individual component of the conviction probability (like θ in the pre-detection stage). However, asymmetric information in the second stage would not alter the result at all, because one could still implement self-reporting as a dominant action by setting $\min(c_i^1) = 0$.
Fourth, we implicitly allow for cooperative behavior in the violation stage by assuming that only *aggregated* fines and benefits matter for the violation decision. Conversely, we assume non-cooperative behavior in the two self-reporting stages. Non-cooperative behavior in the violation stage would lead to less offenses, since both *individual* expected fines would need to be above the individual benefit. But this would not change the optimal scheme qualitatively, while the analysis of the violation decision becomes more cumbersome. If, on the other hand, all teams behave cooperatively in the self-reporting stage, then we are back in the single violator case (i.e. we loose the special appeal of leniency programs to exploit the strategic interaction between criminals).
Fifth, our first self-reporting stage is characterized as a race to the courtroom, while the second one is a prisoner's dilemma. Deterrence could considerably be increased if the pre-detection stage could also be designed as a prisoner's dilemma by assigning low fines for a single self-reporter, but high fines if both self-report. Self-reporting of both members would then require to define a time window (of two or three days, say) where self-reporting would be interpreted as "simultaneous self-reporting". Although tempting from a theoretical viewpoint, legal scholars convinced us that making the fine for a self-reporter in the pre-detection stage contingent on whether the second one self reports or not is legally unacceptable and impracticable, while such a practice is quite common in the conviction stage (i.e. once potential violators are interrogated separately).
Finally and most importantly, to analyze cartel behavior, a repeated game seems a natural assumption (see Motta and Polo (MP 2003), and Buccirossi and Spagnolo (BS 2001)). The key aspect of repeated interaction is that the team has some future benefits if it is not detected, and this reduces the incentive to self-report. This aspect could also be modelled in a reduced form by assuming that B is non-sunk. Then, without ex post asymmetric information, the optimal self-reporting fine would be negative (i.e. a self-reporting firm would be subsidized) to induce full self-reporting. This is exactly what is concluded in BS 2001. In our model, the optimal fine would still depend on the degree of evidence provided and on $z(\theta)$. Moreover, fines could still

be positive, but would definitely be lower.[11] The reason is that the foregone benefit creates opportunity costs of self-reporting, so that the incentive to self-report decreases, and the authority would partially offset this by assigning lower fines. This must be kept in mind when comparing our model to reality. More specifically, our model does not provide point recommendations of fine reductions for a given environment but tries to illuminate the comparative statics of fine reductions with respect to evidence provision and the date of report submission (which is exactly where existing programs mostly differ).

Finally, assuming that B is non-sunk, would have high costs in our model. The problem is that both the violation decision and the self-reporting decision in the first self-reporting stage would then depend on B, so that the authority would have to update its initial information after self-reporting has taken place. This would lead to a signaling game that could hardly be analyzed in a model with asymmetric information on θ, with two self-reporting stages and with different degrees of evidence provision. But as these elements are essential for a *comparison* of the programs, and because the additional effects caused by repeated interaction have been pointed out by MP 2003 and BS 2001, we follow the mainstream of the self-reporting literature by assuming that B is sunk.[12]

Conclusions for the two leniency programs At a very general level, our analysis has shown that self reporting schemes for criminal teams are much more promising than those for single violators. Since strategic interactions between team members can be used to increase expected fines and to reduce the frequency of violations, leniency is part of the authority's optimal policy even without any additional benefit from self-reporting like cost savings or early countermeasures. Hence, we emphasize that these programs should be extended to other fields. In Germany, for instance, all attempts to implement leniency programs for corruption have yet failed according to legal and moral considerations. Incentive arguments, however, are hardly taken into account. Let us now get back to the four differences between the corporate leniency

[11] The fact that optimal fines may still be positive seems to be *the* important feature of ex post asymmetric information.

[12] From a more practical point of view, there is an additional point to note. In reality, the maximum fine that can be imposed depends on the severity of crime which is not independent of the expected future benefits from crime. Hence, s would need to be made endogenous, and contingent on B.

programs in the USA and in the EU identified in the Introduction.

- Ad (i). In the USA, the first self-reporting firm *always* gets full immunity if some requirements are fulfilled, and if self-reporting occurs already in the pre-detection stage. In our model, full immunity can never be optimal for the low evidence provider. For the high evidence provider, it depends on $z(\theta)$, and hence on the case at hand, whether full amnesty is optimal or not. As a consequence, granting always full fine reductions is suboptimal. Recall that the only distinctive feature between the German and EU leniency programs is exactly the leniency for the second self-reporter. Hence, our model suggests a harmonization of EU competition law very much along the lines of the EU system (as already practiced in the Netherlands). However, as pointed out, a full immunity becomes more likely to be the optimal solution (for the high evidence provider) if future benefits from cartel behavior are taken into account. In other words, treating benefits as sunk leads to a bias towards the European system in that respect.

- Ad (ii). In the EU-program, it is explicitly stated that fine reduction increases in the amount of evidence provided. Although this seems straightforward, our findings are somewhat more subtle. In the conviction stage, both fines do *not* depend on the own evidence provided, but only on the probability of being convicted by the partner's testimony (this is driven by the prisoner's dilemma structure of the conviction stage). This is also true for the low evidence provider L in the pre-detection stage. For the high evidence provider H, however, we find indeed that his fine should be decreasing in his own evidence provision. In any case, we conclude that the fine structure should not be independent of the evidence which supports the European solution (i.e. favors a case-based analysis against a winner-take-all approach).

- Ad (iii). In the model presented, we did not explicitly analyze what happens with one violator, L say, if the other one (H) has self-reported. Then, L can be convicted with probability h and pays fine s. Of course, whenever convicting L with probability h requires some investigation effort (which is likely to be the case), it would again be optimal to offer a self-reporting fine of ls to induce self-reporting by the second team member. This makes the extensive form of the game somewhat more complicated and has thus only been considered in Feess and Walzl

(2003). But following the logic of the model, a corner solution where no fine reduction is granted for a second self-reporter is clearly suboptimal.

- Ad (iv). Finally, full immunity can be granted in the USA also if the case is already under investigation. According to our model, the EU-approach where the maximum fine reduction is limited, is more appealing, since the fines in the conviction stage should be higher than in the pre-detection stage for the one who provides more evidence. In contrast to both leniency programs, we find that the fine for the one who provides less evidence (type L) should be the same in both self-reporting stages ($c_2^L = r^L$). The reason is that it is unwarranted that type L wins the race to the courtroom, which explains the high fine $r^L = hs$.

In sum, our model shows the limits of the winner-take-all approach followed by the US authority and highlights the impact of heterogeneity among cartel members with respect to the evidence they can provide about their counterparts on the optimal fine structure. Moreover, we have shown how the nature of the strategic interaction between members of the criminal team (i.e. the differences between races to the court room and prisoner's dilemma situations) influences the optimal enforcement scheme. Both features seem to be underdeveloped in the design of existing leniency programs.

Appendix

Optimization Program. With

$$F_H = \left(\frac{1}{2}(r_H + ls) + hs\right)\int_{\widetilde{\theta}_H}^{1} z(\theta)d\theta + p(h+l)s\int_0^{\widetilde{\theta}_H}\theta z(\theta)d\theta$$

it follows that

$$\begin{aligned}
\frac{\partial F_H}{\partial r_H} &= \frac{z(\widetilde{\theta}_H)}{pls}\left(\frac{1}{2}r_H - \frac{1}{2}ls + r_H\frac{h}{l} - hs\right) + \frac{1}{2}\int_{\widetilde{\theta}_H}^{1} z(\theta)d\theta \\
\frac{\partial^2 F_H}{\partial (r_H)^2} &= \frac{z'(\widetilde{\theta}_H)}{pls}\left(\frac{1}{2}r_H - \frac{1}{2}ls + r_H\frac{h}{l} - hs\right) + \frac{z(\widetilde{\theta}_H)}{2pls}
\end{aligned}$$

Necessary and sufficient conditions for $r_H \in (0, ls)$ being optimal are $\frac{\partial F_H}{\partial r_H} = 0$ and $\frac{\partial^2 F_H}{\partial r_H^2} < 0$, respectively.

Proof of Corollary 1. (i)-(iv) obvious. (v) Comparative statics of r^H w.r.t. h and l are given by

$$\frac{dr_H}{di} = -\frac{\partial^2 F_H/\partial r_H \partial i}{\partial^2 F_H/\partial r_H^2}\ \forall i = h, l$$

where the second order condition requires $\frac{\partial^2 F_H}{\partial r_H^2} < 0$ for every interior solution. To prove (v) note that

$$\frac{\partial^2 F_H}{\partial r_H \partial h} = \frac{z(\widetilde{\theta}_H)}{pls}\left(\frac{r_H}{l} - s\right) < 0$$

(recall that $r_H < ls$).
(vi) Note that the first term in

$$\begin{aligned}\frac{\partial^2 F_H}{\partial r_H \partial l} \quad = \quad & -\frac{z'(\widetilde{\theta}_H) r_H}{(pls)^2 l}\left(\frac{1}{2}r_H - \frac{1}{2}ls + r_H\frac{h}{l} - hs\right) \\ & -\frac{z(\widetilde{\theta}_H)}{pl^2 s}\left(\frac{1}{2}r_H - \frac{1}{2}ls + r_H\frac{h}{l} - hs\right) \\ & +\frac{z(\widetilde{\theta}_H)}{pls}\left(\frac{r_H h}{l^2} - \frac{s}{2}\right) + \frac{1}{2}z(\widetilde{\theta}_H)\frac{r_H}{pl^2 s} \end{aligned} \tag{5}$$

can be approximated using the second order condition $\left(\frac{\partial^2 F_H}{\partial r_H^2} < 0\right)$ such that the sign of Eqn. (5) can be determined by

$$\frac{\partial^2 F_H}{\partial r_H \partial l} > \frac{z(\widetilde{\theta}_H)h}{pl^2}(1 + p) > 0$$

References

Buccirossi, P., and Spagnolo, G. (2001). The effects of Leniency on Illegal Transactions: How (Not) to Fight Corruption, Lear - Laboratory of Economics, working paper.

European Union (2002), Commission notice on immunity from fines and reduction of fines in cartel cases, EC Offcial Journal, 2002/C 45/03.

Feess, E., and E. Heesen (2002). Self-Reporting and Ex Post Asymmetric Information, Journal of Economics, 77, 141-153.

Feess, E. and M. Walzl (2003). Corporate Leniency Programs in the USA and the EU, German Working Papers in Law and Economics Vol. 2003, Art. 24.

Feess, E., and M.Walzl (2004). Self-reporting in optimal law enforcement when there are criminal teams, Economica 71, 333-348.

Hammond, S.(2000). Detecting and Deterring Cartel Activity through an effective Leniency Program, Speach of the Director of Criminal Enforcement, US Departement of Justice, presented at the International Workshop on Cartels, Brighton, November 2000.

Innes, R. (1999). Remediation and self-reporting in optimal law enforcement. Journal of Public Economics 72, 379-393.

Innes, R. (2000). Self-reporting in optimal law enforcement when violators have heterogeneous probabilities of apprehension. Journal of Legal Studies XXIX, 287-300.

Kaplow, L., and S. Shavell (1994). Optimal law enforcement with self-reporting of behavior. Journal of Political Economy 102(3), 583-605.

Livernois, J., and C.J. McKenna (1999). Truth or consequences: Enforcing pollution standards with self-reporting. Journal of Public Economics 71, 415-440.

Motta, M., and Polo, M. (2003). Leniency Programs and Cartel Prosecution, International Journal of Industrial Organization 21, 347-379.

Spagnolo, G. (2000). Optimal Leniency Programs. Fondazione Eni Enrico Mattei, Nota di Lavoro 42.2000.

US Department of Justice (2004), Official Giudelines
(http://www.usdoj.gov/atr/public/guidelines/0091.htm).