Leniency Programs and Cartel Prosecution

Massimo Motta
European University Institute, Florence
Universitat Pompeu Fabra, Barcelona

Michele Polo
Bocconi University and IGIER, Milan [*]

May 10, 1999

Abstract

We study the enforcement of competition policy against collusion under Leniency Programs, which give reduced fines to firms revealing information to the Antitrust Authority. Such programs give firms an incentive to break collusion, but may also have a pro-collusive effect, since they decrease the expected cost of misbehaviour. We analyze the optimal policy under alternative rules and with homogeneous and heterogeneous cartels, obtaining a ranking of the different schemes and showing when the use of reduced fines may improve antitrust enforcement.

1 Introduction

The enforcement of competition policy against collusion and price fixing agreements is one of the main fields of antitrust intervention. Recent developments show that the attention devoted by antitrust authorities to collusive agreements has not diminished over time. Recently, the DGIV, the Directorate-General of the European Union in charge of competition policy, has established a special investigation unit against cartels[1], and a similar pattern can be found in the US, where the Antitrust Division of the Department of Justice has reallocated and improved the resources of the Criminal section in charge for cartel prosecution[2].

In the design of the policy we find today richer and more complex mechanisms than those based simply on an increase in fines. Since 1978 the US Antitrust Division of the Department of Justice has allowed for the possibility of avoiding criminal sanctions if

[*]We benefited from discussions with Luis Cabral, Federico Ghezzi, Patrick Rey, Thomas Von Ungern-Sternberg and seminar participants at the University of Salerno and the European University Institute (Fiesole). Their comments are gratefully acknowledged.

[1]See Venit (1996), p.92, and European Union (1999, p.22).

[2]See Bingman and Spratling (1995).

speci¯c conditions occurred. In 1993 this policy has been redesigned in the Corporate Leniency Policy, which establishes that criminal sanctions can be avoided in two cases: either if a colluding ¯rm reveals information before an investigation is opened, as it was in the previous regime, or if the Division has not yet been able to prove collusion when a ¯rm decides to cooperate [3]. The new Leniency Policy has shown in the ¯rst years of application a signi¯cant success in terms of the number of cases that the Division has been able to open and successfully conclude.

The current EU system draws from the US experience. In order to reach a more e®ective deterrence of collusive practices, the DGIV has initially focused its enforcement policy on a sharp increase in ¯nes: the average ¯ne given to ¯rms involved in collusion cases, up to the mid Eighties has remained below 500.000 ECU while in the last decade it has reached an average of around 1.500.000 ECU[4]. However, if the ¯rms anticipate a low probability of having collusive practices discovered (and proved), ¯nes alone will be insu±cient to prevent ¯rms from establishing cartels. Although it is hard to quantify such expected probabilities, there seems to be a wide perception that the deterrence e®ects of the ¯nes has been relatively poor, and that various types of collusive practices are still widespread. This has pushed the European Union to introduce[5] a new regime in which reduced ¯nes can be given to ¯rms which cooperate with the antitrust authority by providing evidence of a collusive agreement in which they have been involved. A 75-100% reduction in ¯nes[6] can be given if ¯rms reveal information before an inquiry is opened, while a lower reduction (50-75%) can be granted if cooperation occurs after an investigation has started, but that investigation has failed to provide su±cient grounds for initiating a procedure leading to a decision. A 10-50% reduction in ¯nes can be given for partial cooperation, such as providing additional evidence or not contesting the facts on which the Commission bases its allegations. Moreover, only the ¯rst ¯rm which cooperates can obtain a reduction, provided that it is not the promoter and major partner of the cartel. It is too early to evaluate the e®ects of this new policy, although the US experience suggests that enforcement against cartels might become more e®ective.

In this paper we want to investigate the di®erent e®ects that the introduction of a Leniency Program[7] can have on both ¯rms' behaviour and deterrence. Our work is related to other papers on the optimal enforcement of law, speci¯cally those on pre-trial

[3]Some additional restrictions on the ¯rms entitled to bene¯t from this regime are introduced, as the fact that only the ¯rst can be given a ¯ne reduction, and that it must be a junior partner in the cartel.

[4]See Furse (1995), p.114.

[5]European Union (1996).

[6]Notice that while in the US the regime applies to criminal sanctions (which include both ¯nes and incarceration), in the EU reductions are referred only to monetary ¯nes. Criminal sanctions do not exist under EU competition law.

[7]The US Leniency Program involves both reductions of ¯nes and the elimination of the threat of incarceration. In this paper we focus on reduced monetary ¯nes. Hence, we use the term Leniency Programs in a broad sense.

negotiation and settlement[8] and on plea-bargaining[9], in which these alternative judicial procedures have been studied with a general reference to the US judicial system, although not explicitly to antitrust law.

There are however several important di®erences between our work and the existing literature. The papers on pre-trial negotiation have considered mainly the properties of these procedures in saving trial costs preventing wasteful litigation. In the plea bargaining literature the enforcer acts more explicitly on behalf of taxpayers, balancing the goal of condemning the guilty agents and not condemning the innocent ones with the minimization of resources devoted to enforcement. In both cases, the issue of deterrence is not really addressed: agents have (possibly) already committed a crime, and in most papers, whether the agent is innocent or guilty and how strong is the evidence against him (agent's type) is exogenous in the model, and it is not explained in terms of incentives to commit a crime. The e®ects of the legal procedures on preventing the crime or making it to cease are instead at the center of our analysis.

In our paper we are mainly concerned with the deterrence and desistence properties of negotiations between the Antitrust Authority and private ¯rms. The enforcer is motivated by the maximization of social welfare and aims at minimizing the occurrence of collusion among ¯rms by committing on a certain set of policy parameters [10]. In order to focus on deterrence, in our setting we exclude other ingredients already studied in the literature. First of all we do not consider (variable) litigation costs on either party, a central issue in the pre-trial negotiation literature. The enforcer's budget is set at the beginning of the game and enforcement costs are sunk, i.e. they are already allocated among the di®erent tasks of the organization, as general monitoring or prosecution. Secondly, we do not consider the possibility of wrong sentences, analyzed in the plea-bargaining papers: at the end of an investigation either a guilty ¯rm is condemned or no evidence is reached.

In this setting we consider several issues. First of all we analyze the reaction of ¯rms to di®erent policy regimes, i.e. on the incentive to collude and on the decision to reveal or not information to the Antitrust Authority. A perverse e®ect can arise under this respect: since a Leniency Program allows ¯rms to pay reduced ¯nes, it may have ex-ante a pro-collusive e®ect, decreasing the expected cost of anticompetitive behaviour. But we show that, if the Antitrust Authority has limited resources, and is therefore unable to prevent collusion ex-ante, the use of Leniency Programs can improve the e®ectiveness of the policy, by sharply increasing the probability of interrupting collusive practices. Hence, in a second best perspective, ¯ne reductions may be desirable because they allow to better implement ex-post desistence from collusion.

There is however a third component that operates in equilibrium: in order to induce ¯rms to reveal, a Leniency Program has to commit resources to guarantee a su±ciently

[8] Bebchuk (1984), Nalebu® (1987), Schweizer (1989), Shavell (1989).

[9] Grossman and Katz (1983), Reinganum (1988).

[10] Other papers that are related to our own are Kobayashi (1992) and Marshall, Muerer and Richard (1994).

high probability of independent prosecution. This is the implicit cost of a reduced fines regime, since those resources are subtracted from the general monitoring activity, which determines the frequency of "revelations" and successful inquiries. As the resources committed to prosecution become too costly, a Leniency Program loses its appeal, and a full fines regime may become more convenient again. The conditions under which these results hold will be identified in both a homogeneous and a heterogeneous cartel setting. The effects and desirability of alternative leniency rules will also be studied.

The paper is organized as follows. In section 2 we set up the basic model, in which every firm which decides to cooperate with the Antitrust Authority is given a fine reduction. In section 3 we consider alternative Leniency Programs, in which fine reductions can be granted only if cooperation occurs before an investigation is opened, or in which only the first comer, or a specific firm, is entitled to a reduced fine. Finally, in section 4 we extend the basic model to the heterogeneous cartels case. Section 5 concludes the paper.

2 The Model

Throughout the paper, we assume that the Antitrust Authority (AA from now on) aims at maximizing a utilitarian social welfare function and is able to commit to a certain set of policy parameters[11], which consist of full and reduced monetary fines and probabilities of enforcement. In the basic model of this section we consider a regime in which all firms which cooperate in the investigation even after this has been opened, and which simultaneously provide useful evidence to prove collusion [12], can benefit from a reduction in fines. In the following sections we shall consider alternative rules and compare them with this benchmark case.

The AA is (exogenously) endowed with a per-period budget B: in line with the literature, we assume that setting the fines at any level is not costly, while increasing the probability of enforcement requires resources. More precisely, we assume that the maximum fine that firms can receive if found guilty of collusion is exogenously given by law and equal to F, a fixed amount of money: then, being costless, it is always optimal to set the full fine at this maximum level. However, the AA can commit to a Leniency Program which allows for reduced fines $R \leq F$ to firms which reveal information useful to prove the existence of collusion. Indirectly, that is via the allocation of its given resources among different tasks, the AA determines the probability α of opening an investigation and the probability p of proving firms guilty. The former refers to the

[11] This is in line with actual experience, in which little discretion is left by the law to the Authority as to the conditions under which reductions can be given, and their amount.

[12] Throughout the paper, we assume that information given by a single firm is enough to prove that all the firms which have taken part in the collusion are guilty. This might be interpreted as the case where each firm has access to the minutes of the meetings which take place among all the colluding firms, or has copies of letters, faxes or e-mail messages which all the firms have used to coordinate on the collusive outcome. Since an important component in the working of cartels is the coordination of moves among participants, the access of each partner to some information regarding the others seems quite realistic.

preliminary activities (general monitoring) necessary to open an investigation such as collecting information about the ¯rms in the industry, interviewing ¯rms, suppliers and customers, collecting data from the di®erent sources; the latter (prosecution) involves collecting further more focused information on the case, ordering surprise "raids" in the ¯rms' headquarters, processing the information collected and preparing the case against the ¯rms according to the existing laws. The AA, allocating resources to these two groups of tasks can obtain a combinations of these probabilities according to their speci¯c production functions [13]. The budget constraint is then:

$$B = w_{®}® + w_p p \qquad (1)$$

where $w_{®}$ and w_p are the (constant) unit cost of monitoring and prosecution. We assume that ¯rms know the probabilities ® and p chosen by the AA and its budget constraint.

The AA objective function is a standard utilitarian welfare function, i.e. the sum of producers and consumers surplus. Fines, whether full or reduced, are pure transfers, i.e. they go to the general government budget and are redistributed to consumers without distortions, and cannot be used by the AA to increase its budget. The agency problem can therefore be described as choosing the incentive scheme (R; ®; p) in order to in°uence ¯rms' behaviour and maximize social welfare. The incentive compatibility constraints will be derived from the analysis of the subgame perfect equilibria in the supergame played by ¯rms once the policy parameters are set. After observing the policy parameters chosen by the AA, n identical ¯rms decide whether to collude or not, by correctly taking into account the probabilities (®; p) and by knowing whether a Leniency Program R is in place or not.

We follow the usual supergame literature and consider the incentive of each ¯rm to play an action which leads to the collusive outcome given that all other ¯rms take the collusive action. If a ¯rm deviates it earns a pro¯t $¦_D$ in the current period but it triggers the punishment of the other ¯rms, which will play the one-shot non-cooperative equilibrium action forever afterwards, by giving the deviating ¯rm a total discounted payo® of $¦_D + ± ¦_N$=(1 ¡ ±). If instead the ¯rm decides to take the collusive action, then it earns a payo® of $¦_M$ (with $¦_N < ¦_M < ¦_D$) in the current period.

We assume that the existence of a collusive outcome in the industry is perfectly observed by the antitrust agency, but this is not enough for collusion to be proved in courts. To be able to build a case against the ¯rms (which would otherwise win the appeal in a Court), the AA needs to ¯nd some "hard" information about collusion. Such information might consist of any document proving that ¯rms have agreed on prices or

[13] More precisely, let the AA budget constraint be $B = w(l_{®} + l_p)$; where B is the total budget available to the Authority; $l_{®}$ the number of hours allocated to general monitoring and l_p those devoted to prosecution, w the wage rate. In turn, the probabilites are determined given the resources according to the production functions $® = k_{®}l$, and $p = k_p l$, with ® and p 2 [0; 1], characterized for simplicity by positive and constant marginal productivity. Then the labor requirement to obtain ® and p are $l_{®}(®) = ®=k_{®}$ and $l_p(p) = p=k_p$ respectively and the total cost of implementing ® and p are $wl_{®}(®) = w®=k_{®} = w_{®}®$ and $wl_p(p) = wp=k_p = w_p p$. It will be clear in the analysis that assuming decreasing marginal productivity, which would imply a concave budget line and a convex budget set, would not alter all our conclusions.

have met to coordinate on the prices to be charged [14]. Perfect observability of collusive prices also implies that the antitrust agency will never open an investigation on firms which do not collude at equilibrium.

For simplicity we consider the case where firms decide once and for all at the initial period whether to collude or to deviate from the projected cartel [15]. From our discussion so far, the timing of the game, represented in Figure 1, is as follows:

$t = 0$ The Antitrust Authority determines the policy parameters $R; \alpha; p$, which are observed by all firms. The reduced fine R is granted to any firm cooperating even after the investigation is opened.

$t = 1$ Firms $i = 1; \dots; n$ decide whether to collude or deviate and realize the per-period associated payoff.

$t = 2$ The AA opens an investigation with probability $\alpha \in [0; 1]$. If the inquiry is not opened, each firm realizes the per-period profits associated to the previous choice. If the investigation starts, firms simultaneously decide whether to reveal information that the AA will find useful to prove collusion; if at least one firm reveals, the AA is able to prove them guilty. The firm(s) which cooperated with the AA pays RF while the others pay the full fine F. If no firm reveals, the AA is able to prove them guilty with probability $p \in [0; 1]$. If the AA has not been able to prove the firms guilty of collusion at the end of this inquiry, the firms will never be investigated again in the future. If proved guilty, they will behave non-cooperatively forever in the future.

$t > 2$ If up to the previous period the AA has not started an investigation, with probability α it opens an inquiry in t, firms decide whether to reveal, and so on.

Figure 1 about here

We can now solve for the equilibrium of this game. Our first step is to identify the incentive compatibility constraints, which requires to work out, for given policy

[14] To this purpose, note that in our model, like any repeated game with an infinite horizon, there exists a continuum of possible equilibria, and firms need some coordination to select the fully collusive outcome giving them the per-period profit π_M.

[15] In our setting, this is not a completely innocent assumption since the game becomes stationary only after the initial period, once firms have started colluding: considering the choice of deviating for any $t > 1$ is equivalent, since in this case firms, having participated for some periods to a cartel, pay an expected fine even if they deviate later on. When deviating initially, on the contrary, a firm can avoid the fine, since it never participated to the illegal agreement. However, notice that for this reason a deviation at the beginning is more attractive than breaking down the cartel later on, and the associated constraints are more stringent. Since the alternative case makes the analysis more complex but gives the same qualitative results, we have preferred to keep the simplest version where firms decide only at $t = 1$ whether to deviate or collude.

parameters, the subgame perfect equilibria of the game starting at $t = 1$, characterized by firms colluding or deviating and by the choice of revealing or not information to the AA. We first consider the "revelation game" which is played from $t = 2$ on if an investigation is opened by the AA. The following Lemma identifies the conditions for the existence of Nash equilibria in which firms cooperate or not with the AA.

Lemma 1 Let

$$1 - \frac{(1-p)(\pi_M - \pi_N)}{pF - R} \equiv \bar{\delta}(p;F;R) \tag{2}$$

Provided that an investigation has been opened, in the "revelation" game an equilibrium always exists in which all firms reveal information.
If

1) $pF < R$ or

2) $pF \geq R$ and $\delta \geq \bar{\delta}(p;F;R)$

an equilibrium exists in which no firm reveals. If this latter exists, it Pareto dominates the equilibrium outcome in which the firms reveal.

Proof: See Appendix. 2

Figure 2.a below illustrates the critical locus of points $\bar{\delta}$: To the right of this curve, firms reveal if an investigation has been opened by the AA. To the left of it, they do not. This curve, which always passes through the upper right corner of the picture, rotates to the left as the reward from revealing information increases (that is, the lower R) and the larger becomes the fine F to be paid if found guilty: in other words, revelation occurs for a wider set of parameters as the incentive to cooperate with the AA is sharpened.

We can now consider the initial decision to join the proposed agreement or deviate at t=1. Three possible outcomes can occur: firms might prefer not to collude (NC), since they expect an immediate deviation. Alternatively, collusion may start, followed by the decision not to reveal (CNR) or to reveal (CR) if an investigation is opened by the AA. To simplify the statement of the results, it is convenient to introduce the following expressions. $\delta_{NC}(\alpha;p;F)$ is the value which solves:

$$\frac{\pi_M\left(1 + \frac{\delta\alpha(1-p)}{1-\delta}\right) + \delta\alpha p\left(\frac{\pi_N}{1-\delta} - F\right)}{1-\delta(1-\alpha)} = \pi_D + \frac{\delta\pi_N}{1-\delta} \tag{3}$$

while $\delta_{CR}(\alpha;R)$ is:

$$\frac{\pi_D - \pi_M}{(1-\alpha)(\pi_D - \pi_N) - \alpha R} \equiv \delta_{CR}(\alpha;R): \tag{4}$$

The following proposition identifies the conditions on the discount factor δ for the three outcomes to occur.

Proposition 2 For given policy parameters $(F; R; \alpha; p)$:

- if $\delta_{CR}(\alpha; R) \quad \delta \quad \tilde{\delta}(p; F; R)$, firms collude and reveal if monitored (CR).
- if $\delta \geq \max\{\delta_{NC}(\alpha; p; F); \tilde{\delta}(p; F; R)\}$, firms collude and do not reveal if monitored (CNR).
- if $\delta < \min\{\delta_{NC}(\alpha; p; F); \delta_{CR}(\alpha; R)\}$ firms do not collude (NC).

Proof: See Appendix. $\square$

Figure 2.a below illustrates the line corresponding to δ_{CR}, for given values of α and R: this locus does not depend on p (it is flat) since in the region to the right of $\tilde{\delta}$ firms cooperate with the AA once an investigation is opened and p becomes irrelevant. Above the line, firms prefer to collude even though they anticipate that, if an investigation is opened, collusion would collapse because firms would reveal information to the AA. Below the line, firms, anticipating revelation, prefer to deviate, and the collusive outcome never occurs.

Consider now δ_{NC}, which identifies the regions where firms start colluding (above) or not (below). For $\alpha = 0$ of $p = 0$, we have $\delta_{NC} = \delta_{CR} = \frac{\pi_D - \pi_M}{\pi_D - \pi_N}$, and the condition for collusion amounts to the "textbook" critical discount factor, which is in fact derived under the condition of no antitrust enforcement. Positive values of α and p (and higher values of the full fine F) increase δ_{CR} and make the cartel harder to sustain, since the expected collusive profits are reduced.

Note also that the more generous the Leniency Program (the lower the reduced fine R) the lower δ_{CR}: if firms expect that in case an investigation is opened they have the possibility to reveal information and get away with a small fine, this will give an incentive to choose the collusive strategy. In other words, a generous Leniency Policy might stimulate ex-ante collusion. (We shall come back to this issue below.)

Figures 2.a and 2.b about here

The curves represented in Figure 2.a define, for a given α, the conditions that must hold for a collusive agreement to emerge, and those which induce revelation or not if an inquiry is opened by the AA. More precisely, if no Leniency Program is introduced ($R = F$) firms have no reason to reveal information to the Authority if an investigation is opened, and the equilibrium outcomes would be defined uniquely by the line δ_{NC}: Above the line, firms would collude (CNR); below, they would not (NC), because any proposed agreement would break down immediately. Reduced fines modify the situation: in the region to the left of $\tilde{\delta}$ firms don't reveal if monitored, and the same argument above still applies. To the right of that curve, firms anticipate that they reveal information if monitored: below δ_{CR} they prefer not to collude and above they initially collude and then reveal if monitored.

We can notice that the conditions for collusion are more demanding with respect to the standard case when no AA operates: the critical discount factor needed for a collusive outcome is always higher than $(\pi_D - \pi_M)/(\pi_D - \pi_N)$ when α and p are positive. When a firm considers whether to join a cartel or deviate, in fact, it evaluates the collusive profits taking into account that with a certain probability collusion will be detected, inducing a double loss: the fine to be paid and the lost collusive profits from there on. The higher the probability of these losses, the lower the collusive profits. Hence, we need a higher and higher discount factor to balance the temptation to deviate.

To understand the role of Leniency Programs on the sustainability of collusion, consider what happens when, starting with a situation in which no Leniency Program is used, we introduce reduced fines. This has two effects which are shown in Figure 2.a. On the one hand, the Leniency Program might have an adverse, pro-collusive effect. By reducing the expected value of the fine to be paid if an investigation is opened, the Leniency Program might give an incentive to collusion. This occurs in the area (1) included between the dotted part of the curve δ_{NC} and the line δ_{CR}. In this region, no collusion can be sustained in the industry if full fines are given (NC), but under a Leniency Program firms would engage in collusion and, if monitored, they would reveal (CR) and pay the reduced fine $R < F$.

On the other hand, there exists an area (2) where collusion will break down (because the firms reveal information) if the AA starts monitoring the industry (CR), whereas in the absence of a Leniency Program collusion could stop only after a successful complete investigation (CNR). This is the area comprised between the dotted part of the curve δ_{NC} and the curve $\tilde{\delta}$. [16].

We can now move to the analysis of the optimal policy, having identified the implementable allocations. So far we have expressed the conditions for the different equilibrium outcomes in the space $(p; \delta)$: this was useful because we obtained the conditions of cartel stability in terms of critical discount factors, thereby allowing a comparison with the modern theory of collusion. To proceed with the analysis of the optimal policy design, it is convenient to rewrite the critical loci found above in the space $(p; \alpha)$ of policy parameters.

Firms would reveal if monitored if:

$$p \geq \frac{\pi_M - \pi_N + R(1-\delta)}{\pi_M - \pi_N + F(1-\delta)} = \tilde{p}(\delta; R; F). \tag{5}$$

Firms would prefer to collude rather than deviate, when they anticipate that the opening of an investigation would result in collusion broken down by revelations, if:

[16] If the Leniency Program were unanticipated, firms would decide whether to collude or not on the basis of an expected fine $R = F$ and therefore would not cooperate unless $\delta \geq \delta_{NC}$. When the leniency program is introduced unexpectedly, collusion would break down in all the area below the curve $\tilde{\delta}$ (that is, (1) plus (2)), without any adverse effect arising.

$$\alpha \quad \frac{\pi_M - \pi_D + \delta(\pi_D - \pi_N)}{\delta(\pi_D - \pi_N + R)} = \alpha_{CR}(\delta; R): \tag{6}$$

Finally, collusion arises in the case where firms anticipate that no revelation would occur after the opening of an investigation, if:

$$\alpha \quad \frac{(1-\delta)[\pi_M - \pi_D + \delta(\pi_D - \pi_N)]}{\delta[pF(1-\delta) + p(\pi_M - \pi_N) + \pi_D(1-\delta) - \pi_M + \delta\pi_N]} = \alpha_{NC}(\delta; p; F): \tag{7}$$

The three loci above allow to define, in the space of policy parameters, three regions associated with different implementable allocations, in which firms do not collude (NC), collude and reveal if monitored (CR) and collude and do not reveal (CNR) :

$$A_{NC} = \{(\alpha; p) \in [0;1]^2 \mid \alpha \geq \max\{\alpha_{NC}(p); \alpha_{CR}\}\} \tag{8}$$
$$A_{CNR} = \{(\alpha; p) \in [0; \bar{p}] \times [0;1] \mid \alpha < \alpha_{NC}(p)\} \tag{9}$$
$$A_{CR} = \{(\alpha; p) \in [0; \alpha_{CR}] \times [\bar{p}; 1]\} \tag{10}$$

When no Leniency Program is introduced the only outcomes are NC, if $(\alpha; p)$ are above the α_{NC} curve, or CNR otherwise. If $R < F$ the threshold $\bar{p}$ becomes lower than 1 and CR is an outcome if $\alpha < \alpha_{CR}$ and $p > \bar{p}$. Notice that $\alpha_{NC}(\bar{p}) = \alpha_{CR}$, that is the upper left corner of the region associated to CR shifts up along the α_{NC} curve as R is reduced. When $R = 0$ we obtain the widest CR region.

We find also in the $(\alpha; p)$ space the same adverse effect of Leniency Programs already discussed: the intersection of A_{NC} when $R = F$ and A_{CR} when $R < F$ is non empty. That means that there are policy combinations which prevent collusion when full fines are given and that induce firms to collude and reveal once a Leniency Program is introduced.

Moreover, if $\delta < \frac{\pi_D - \pi_M}{\pi_D - \pi_N}$, where the latter term is the standard critical discount factor for collusion when no antitrust prosecution is considered, $\alpha_{NC} < 0$ and $\alpha_{CR} < 0$, i.e. the only admissible outcome for any value of the policy parameters is NC. Figure 2.b illustrates the equilibrium outcomes when $\delta > \frac{\pi_D - \pi_M}{\pi_D - \pi_N}$ and $R < F$, and it is the dual of figure 2.a - see above.

We summarize the subgame perfect equilibrium outcomes of the supergame played by firms for given policy parameters and discount factor δ in the following proposition, which is the dual of Proposition 2.

Proposition 3 Given the gains π_M and π_D specified above,

- If the policy combination $(\alpha; p) \in A_{NC}$ there is a unique subgame perfect equilibrium in which firms will abstain ex ante from collusion (NC).
- If $(\alpha; p) \in A_{CNR}$ there is a unique subgame perfect equilibrium in which firms collude and don't reveal if monitored (CNR).

• If $(\alpha; p) \in A_{CR}$ there is a unique subgame perfect equilibrium in which firms collude and reveal if monitored (CR). If $R = F$, A_{CR} is an empty set.

The AA chooses $(\alpha; p; R)$ given the incentive compatibility constraints, summarized in Figure 2.b and Proposition 3, in order to maximize a utilitarian welfare function in which fines are pure transfers. Let $K = DWL/(1 - \delta)$ be the discounted sum of the deadweight loss DWL, which can be thought of as the net social benefit from preventing collusion. Moreover, let W_j be the present value of the welfare gain if the policy induces the equilibrium outcome $j = NC; CR; CNR$. Then we have, for given $(\alpha; p)$, $W_{NC} = K > W_{CR} = \alpha K/(1 - \delta(1 - \alpha)) > W_{CNR} = \alpha p K/(1 - \delta(1 - \alpha))$.

It is useful to identify the (welfare) indifference curves for the policy problem in the $(\alpha; p)$ space: if we do not introduce fine reductions, in all the region A_{NC} we have full deterrence ex-ante and the associated welfare gains are K for all the policy parameters in the A_{NC} region. In the region A_{CNR} the indifference curve for a level of welfare gains $\overline{W}_{CNR}$ is $\alpha = \overline{W}_{CNR}(1 - \delta)/(pK - \delta \overline{W}_{CNR})$, i.e. it is a decreasing and convex curve in the $(\alpha; p)$ space: ex-post desistence in this case depends on both α and p according to the trade-off described by the curve.

Figure 3 about here

Moreover, it is easy to show that the indifference curves in the A_{CNR} region have a shape similar to the α_{NC} curve as defined in (7), which is the upper boundary of that region, and in the limit they overlap with that curve. In fact, if we consider the indifference curves for given $\overline{W}_{CNR}$ and the $\alpha_{NC}(p)$ curve which is the upper boundary of the A_{CNR} region and equate them we obtain after rearranging:

$$\frac{\overline{W}_{CNR}}{K} = \frac{\delta(\pi_D - \pi_N) - (\pi_D - \pi_M)}{\delta(F(1 - \delta) + \pi_M - \pi_N)} \equiv \tilde{p}$$

The right hand side expression corresponds to the upper intercept of the $\alpha_{NC}(p)$ curve at $\alpha = 1$, i.e. $\alpha_{NC}(\tilde{p}) = 1$. Hence, looking at the expression above, if $\overline{W}_{CNR} = \tilde{p}K$ the indifference curve overlaps with the upper boundary of the A_{CNR} region, that is with the $\alpha_{NC}(p)$. For $\overline{W}_{CNR} < \tilde{p}K$ the indifference curve in the A_{CNR} region shifts toward the origin.

When a Leniency Program is introduced, below the A_{NC} region we have the A_{CR} and A_{CNR} regions. The indifference curves across the region A_{CR} are $\alpha = (1 - \delta)\overline{W}_{CR}/(K - \delta \overline{W}_{CR})$: those curves are horizontal, since in the CR case ex-post desistence depends only on α. The same level of welfare in the A_{CNR} region can be obtained only if α is higher; that means that the indifference curve is discontinuous at $\tilde{p}$ and jumps up as we move from the A_{CR} to the A_{CNR} region [17].

[17] Notice that $W_{CNR} = W_{CR}$ for the same α when $p = 1$. Hence, if we extend the W_{CNR} indifference curve in the A_{CR} region up to $p = 1$ we find the level of α such that $W_{CR} = W_{CNR}$ and we are able to identify the level of the indifference curve in the A_{CR} region, as shown in figure 3.

The iso-welfare curves in the A_{CNR} and A_{CR} regions do not identify a convex set of policy parameters. We proceed therefore convexifying the indi®erence curves in the following way. Consider an indi®erence curve in the A_{CNR} and A_{CR} region; draw a line which passes through the point of discontinuity (® = (1 ¡ ±)W_{CR}=(K ¡ ±W_{CR}); p = $\bar{p}$) and which is tangent to the indi®erence curve in the A_{CNR} region. Let the tangency point be e(W_{CNR}); repeating this precedure for di®erent values of W_{CNR} an entire locus e(W_{CNR}) is obtained. De¯ne E_{CNR} the subset of A_{CNR} to the left of that locus, which is represented in Figure 3. Notice that, constructing E_{CNR}, we have excluded those points on the indi®erence curves in the A_{CNR} region which are dominated by a combination of policy parameters in (at the boundary of) the A_{CR} region, obtaining a convex set of policy parameters.

We can now analyze the optimal policies. According to the values of B; $w_®$ and w_p, i.e. the position of the budget constraint B = $w_®$® + w_pp in the (®; p) space, we can have di®erent solutions to the optimal policy problem.

Proposition 4 Consider the optimal policies given the budget constraint.

² If the budget constraint is above or on the $®_{NC}$(p) curve, the optimal policy implements NC at a tangency point between the budget constraint and the $®_{NC}$(p) curve, and the set of possible equilibrium outcomes includes all the curve, i.e. E_{NC} = f(®; p) j p 2 [0; 1]; ® = $®_{NC}$(p)g.

² If the budget constraint is below the $®_{NC}$(p) curve the optimal policy implements either CR or CNR.

  { In a CR equilibrium the optimal policy sets R = 0, p = $\bar{p}$ and ® along the budget constraint, and the policy combinations lie along the vertical line $\bar{p}$, i.e. in the set E_{CR} = f(®; p) j ® 2 [0; $®_{CR}$]; p = $\bar{p}$g.

  { In a CNR equilibrium the optimal policy combinations are at the tangency point between the budget constraint and the indi®erence curve.

² If the budget constraint is tangent to an indi®erence curve in the E_{CNR} region de¯ned above, the optimal policy implements a CNR outcome; otherwise CR is the equilibrium outcome.

Proof: See Appendix. 2

Proposition 4 gives the conditions which in general allow to identify the optimal policies for given budget constraint and it de¯nes three sets of policy parameters which correspond to the di®erent equilibrium outcomes, as represented in Figure 3. It is useful to consider the sequence of policy regimes that are associated with lower and lower budget constraints. Notice that two possibile sequences can arise, according to the way in which the budget constraint shrinks: either we move from a NC to a CNR regime, if the budget constraint is initially very steep and the tangency point on the $®_{NC}$(p) curve which implements the NC outcome lies in the neighborhood of the E_{CNR} region, or we have, for °atter budget constraints, a NC-CR-CNR sequence if the tangency point with the ®$_{NC}$(p) curve is in its lower part. This latter case seems quite interesting and allows to get the intuition of the pros and cons of the Leniency Programs.

Consider the optimal policies for parallel shifts of the budget constraint; for a relatively high total endowment a NC outcome can be implemented at a tangency point with the ®$_{NC}$(p) curve: in this case reduced ¯nes would be harmful, inducing collusion (and revelation) when otherwise the AA would be able to prevent collusion. When the budget constraint shifts downwards and lies below the A_{NC} curve, it is no longer possible to obtain ex-ante deterrence of collusion. In this case it is optimal to implement a CR outcome by granting maximum ¯ne discounts and setting the policy parameters along the $\bar{p}$ vertical locus: intuitively, when the AA is only able to implement ex-post desistence, reduced ¯nes become appealing as a less costly way of proving and interrupting collusion. The implicit cost of such a policy is the need to sink resources in order to make independent prosecution a credible threat which induces revelation. As a consequence, when the total endowment is further reduced (the budget constraint shifts further downwards), fewer and fewer resources are left for general monitoring, which in the end determines the likelihood of interrupting collusion and the desirability of such a policy. At some point, we ¯nd that the (low) budget constraint becomes tangent to the iso-welfare curve in the E_{CNR} region: it means that we obtain a higher expected welfare moving to the region where ¯rms collude and do not reveal, abandoning the Leniency Program and changing the mix of policy parameters in a more favourable way.

3 Alternative Leniency Rules

In this section we adapt the model to alternative Leniency Rules that have been adopted in the recent experience in the US and in the European Union. The ¯rst extension refers to the possibility of giving reduced ¯nes only if ¯rms reveal information before an inquiry is opened by the AA. Another regime assigns the reduction in ¯nes only to the ¯rst ¯rm which o®ers cooperation with the agency. Next, we suggest that if only one speci¯c ¯rm is entitled to bene¯t from a Leniency Program, this policy would be even more successful.

3.1 Fine reductions only before the inquiry is opened

As mentioned in the introduction, the initial Leniency Program introduced in the US in 1978 entitled ¯rms with a reduction in ¯nes only if the cooperation started before an inquiry was opened. On the same line, the actual regime chosen in the EU with the July 1996 Notice is more favourable for ¯rms who reveal information before the AA has opened an o±cial investigation. It is therefore interesting to analyze whether this rule can be justi¯ed in terms of enforcement e®ectiveness. We show that this is not the case.

Let us consider a "¯ne reductions only before an inquiry is opened" regime. The corresponding game structure is described for the general case of n firms in the following[18]:

$t = 0$: The AA sets the policy parameters α, p, R which are observed by the firms.

$t = 1$: Firms $i = 1, \ldots, n$ decide whether to collude or deviate and realize the associated payoffs.

$t = 2$: At the beginning of the period, firms simultaneously choose whether to reveal the existence of the cartel to the AA, benefitting of reduced fines, or not; if no firm reveals, the AA opens an investigation with probability $\alpha \in [0, 1]$, proving them guilty with probability $p \in [0, 1]$. Then, payoffs are realized.

$t > 2$: if up to the previous period the AA has not started an investigation, the game restarts as from $t = 2$, etc.[19]

Consider first the subgame starting at $t = 1$ after a decision to collude. To find the conditions under which not revealing is an equilibrium, we have to compare the payoff from revealing when the other firms do not reveal, namely $\frac{\pi_N}{1-\delta} - R$, with the payoff from not revealing when the other firms do not reveal. The latter is given by:

$$\pi_{nr} = \alpha[p(\frac{\pi_N}{1-\delta} - F) + (1-p)(\frac{\pi_M}{1-\delta})] + (1-\alpha)(\pi_M + \delta\pi_{nr});$$

whence:

$$\pi_{nr} = \frac{\alpha p(\frac{\pi_N}{1-\delta} - F) + [(1 - \delta + \alpha(\delta - p)]\frac{\pi_M}{1-\delta}}{1 - \delta(1-\alpha)}.$$

It is simple algebra to check that this payoff is higher than $(p(\frac{\pi_N}{1-\delta} - F) + (1-p)\frac{\pi_M}{1-\delta})$, the expected payoff from not revealing after the investigation has been opened, which was the relevant one under the rule analyzed in the previous section. Since the payoff from revealing is the same in both cases, it follows that the equilibrium where firms do not reveal is more likely to occur when the Leniency Program is applied only for revelations before the inquiry is opened. In other words, the curve $\bar{\delta}$ moves to the right and collusion is less likely to be broken by revelations in this regime. This is hardly surprising, because the probability of the event "being found guilty and thus fined" is lower before seeing if the industry will be monitored than after an investigation is actually opened.

We have now to consider if the Leniency Program might change the ex-ante incentives of firms to collude. It turns out that there would never be collusion in the industry when firms expect that there would be revelation of information to the AA in the following period: this implies that an equilibrium in which firms choose to collude and reveal does not exist. In fact, by colluding when expecting the cartel to be broken by information

[18] The payoffs in the different outcomes are similar to the model analysed above, and will be omitted here in the description of the game.

[19] Allowing firms to choose whether to reveal or not before an investigation is opened at any period would not change the results.

given to the AA, a firm would get $V_c = \pi_M + \delta(\pi_N/(1-\delta) - R)$. By deviating, it would get $V_d = \pi_D + \delta\pi_N/(1-\delta)$. Since $\pi_D > \pi_M$ and $R \geq 0$, it follows that $V_c < V_d$.

In the case, considered in the previous section, where firms were entitled to fine discounts after the opening of an investigation, the expected profit from collusion decreases when the event "opening of an investigation" realizes, leading firms to reveal information to the agency. In the case we are considering here, instead, nothing new happens between the moment when the firms decide on collusion and the moment when they are asked to cooperate with the authorities to break down the cartel.

Our analysis reveals that if Leniency Programs are to be effective in breaking down cartels, they should be extended to benefit firms which reveal after the industry is put under monitoring. Since proving firms guilty of collusion is a very lengthy and complex issue, which does not always end up with the firms being condemned, a great amount of resources can be saved and a final positive outcome guaranteed by ensuring that firms have the proper incentives to collaborate with the AA even after an investigation has been started.

This result is consistent with the US experience, where initially the Leniency Program was used only for firms which spontaneously offered evidence before the inquiry was opened by the AA. In this initial regime the program was quite ineffective while, once allowed in 1993 for reduced fines even after the inquiry was opened, the number of cases in which firms cooperate with the judges increased significantly. In the 1994 Annual Report of the Antitrust Division it is stressed that in the first year of the new regime "an average of one corporation per month come forward with information on unilateral conspiracies, compared to an average of one per year under the previous policy. The policy thus allowed the Division to extend the reach of its criminal enforcement activities with relatively little expenditure of resources"[20].

According to our results, the new regulation on Leniency Programs [21] adopted by the EU should be widened. The regulation states that firms which denounce a cartel before the Commission has opened an investigation are entitled to a reduction of 75-100% of the fines. Firms which denounce a cartel after that a "verification" has been opened are entitled to a 50-75% reduction of the fines, but only if those verifications had not been fruitful and had not led to the opening of a procedure. Basically, this means that Leniency Programs are opened only for firms operating in industries which are not under the scrutiny of the AA. This narrows too much the scope of the application of the regulation, and fails to provide the firms with enough incentives for revealing information which can be useful to break the cartel. [22]

[20] Antitrust Division (1994), p.6-7.

[21] See Official Journal of the European Communities, Series C, 207, 18-7-1998.

[22] Taken literally, our analysis would also suggest that when firms reveal they should always receive a zero fine ($R = 0$), since this would give them the greatest incentive to denounce the cartel. However, we are assuming that collaborating with the AA is a binary variable. Either one does not collaborate, or if it does it can give all the information necessary to prove the participation in the cartel of all the firms. In reality, the type of information that firms can provide would be more of a continuous variable, and tuning the fine reductions to the quality of the information revealed makes sense.

3.2 Only the first comer obtains a fine reduction

The criteria that determine which firms can receive the benefits of a reduced fine have been restricted in different ways both in the US and in the EU experience. An interesting case is where only the first firm which offers evidence is given a fine reduction, as it is the case in the EU regulation.

In this case the game structure is the same as in our initial model. The only difference is that if all firms decide to reveal information to the judges, as it happens in a subgame perfect equilibrium in which firms reveal if monitored, the expected payoff becomes:

$$\frac{\pi_N}{1-\delta} - \frac{R+(n-1)F}{n} \tag{11}$$

where n is the number of firms in the cartel: every firm is ex-ante the first one to disclose information to the AA with probability $1/n$. Notice, however, that when we check for the existence of an equilibrium in which no firm reveals, a deviating firm obtains the reduced fine R for sure, being the only one which cooperates with the judges. Hence, the condition for an equilibrium in which no firm reveals is $\delta \geq \bar{\delta}$, exactly as in the case treated above.

Moreover, it is easy to see that if an equilibrium exists in which no firm reveals if monitored, it also Pareto dominates the equilibrium in which all firms cooperate with the AA.

Consider now the decision of firms on collusion at $t = 1$: if $\delta \geq \bar{\delta}$ firms will not reveal if monitored and everything is as in the basic model. However, if $\delta < \bar{\delta}$, revelation will follow the opening of an investigation, but firms' incentives to collude are modified in the present regime, since the expected payoff if monitored is lower than in the previous case where all firms could benefit from the Leniency Program.

One can check that firms will abstain from collusion iff

$$\delta \geq \frac{\pi_D - \pi_M}{(1-\alpha)(\pi_D - \pi_N) - \alpha\left(\frac{R+(n-1)F}{n}\right)} \equiv \delta^I_{CR} \tag{12}$$

It is immediate to notice that $\delta_{CR} < \delta^I_{CR}$, that is, the region of parameters that induce firms to abstain from collusion is larger than in the previous "all firms get the reduction" regime - see Figure 4.

Figure 4 about here

The intuition of this result is as follows: in the more restrictive set of rules analyzed in this section, the expected reduction in fines is smaller when all firms choose to cooperate with the judges, although it is equivalent when we consider the incentive for a firm to cheat the partners when they do not reveal. Hence, when firms anticipate that they all will confess if monitored, they expect higher sanctions. Consequently, in some cases they find it less attractive to collude and reveal as an alternative to deviating from the beginning and avoiding the fine. The regime therefore is able to partially reduce the ex-ante incentive to collusion without reducing the power of the program in making firms denounce a cartel after an inquiry is opened.

This case suggests an alternative rule which might increase the effectiveness of a Leniency Program, by further reducing the ex-ante incentive of engaging in collusion induced by the expected reduction in fines.

3.3 Only a specific firm receives a fine reduction

[23]

As we have repeatedly emphasized, a Leniency Program influences firms in two ways. The first is that it stimulates ex-post breaking of the cartel via revelation of information to the AA; the second (adverse effect) is to increase the incentive of collusion via a reduction in the punishment in case of being found guilty. We have also seen that granting a reduction in fines only to the first firm which reveals works because it leaves unchanged the first effect but reduces the second. The effectiveness of the Leniency Program could be increased even further by increasing asymmetries in the industry and specifying ex-ante that only a specific firm could be entitled to the LP, no matter the way in which such a firm is selected. The way of interpreting this rule is that of defining ex-ante a set of parameters which allow all the participants in each specific situation to identify a single firm entitled to a reduced fine[24] : the firms involved in the cartel, applying the rule, are able to work out which one will be the firm selected. Denote this firm with a number, say 1. The conditions under which revelation occurs are the same as usual: If $\delta < \bar{\delta}$, the cartel would break because firm 1 denounces it. On the other hand, the conditions under which ex-ante collusion occurs will change. For the $n - 1$ firms which are not entitled to the Leniency Program, the condition for taking part in the collusion will be:

$$\delta \geq \frac{\pi_D - \pi_M}{(1-\alpha)(\pi_D - \pi_N) - \alpha F} \equiv \delta^F_{CR} \tag{13}$$

For firm 1, the condition is laxer:

$$\delta \geq \frac{\pi_D - \pi_M}{(1-\alpha)(\pi_D - \pi_N) - \alpha R} \tag{14}$$

However, since all the firms must find it convenient to take part in collusion, the latter condition does not play any role, while the former is binding and determines the existence of collusion. Also, notice that $\delta^F_{CR} > \delta^I_{CR} > \delta_{CR}$ - see Figure 4. In other words, if the authority targets the Leniency Program to a specific firm, it will be able to reduce the ex-ante adverse effect of it without decreasing the ex-post incentive to reveal information. Hence, collusion becomes less likely because the firms excluded from the program find it less appealing to engage in a cartel which includes a likely cheater[25].

[23] We thank P.Rey for suggesting this extension.

[24] For instance, it might be the firm located in the smallest city, or the last one in alphabetical order, etc.

[25] Of course, leniency rules which limit the applicability of the fine reduction to only one firm will result

4 Heterogeneous Cartels

So far we have considered homogeneous cartels, in which the payoffs in each possible outcome were the same across partners. Notice however that, in all our arguments, if the participants have heterogenous payoffs and they know the payoffs of the partners in each possible outcome, the equilibria are governed by the conditions of one of the firms, the one whose constraints bind. This decisive agent is the point of reference for the others, whether they expect such firm to deviate or to reveal information after colluding, and drives the equilibrium conditions of the entire cartel. Hence, in a sense, our previous analysis allows to consider heterogeneous firms within a cartel, but it assumes that in each cartel in the economy such a decisive agent is always the same. It is therefore interesting to consider the case in which the cartels are truly heterogeneous, in the sense that the participants may differ in payoffs and the decisive partner may be different across cartels.

We consider in this setting the design of an optimal enforcement policy which cannot be made conditional on cartel's type, due to informational and/or institutional restrictions. Hence, the AA has to design a single, general policy facing many different industries, characterized by heterogeneous market conditions and potential gains from collusion. In this case, the policy implemented will induce different effects in the various industries, reaching a more or less effective deterrence of collusion and inducing different types to choose different reactions: hence we might have some cartel types colluding and not revealing while others will prefer not to join the cartel; or we might have all types colluding, but only a subset of them revealing information when monitored, etc. Hence, the different effects that we have identified in the previous sections will be combined in a richer way once the AA faces heterogenous types.

From the previous analysis we already know that the incentive compatibility constraints for given policy parameters depend on two variables of cartel's type: $\pi_M - \pi_N$ and $\pi_D - \pi_N$. Hence, multiple types would require to deal with a bivariate distribution, related to the gains from collusion and from deviation. To maintain the analysis simple, we assume in this section Bertrand competition (with constant marginal costs) in the non-cooperative equilibrium: hence $\pi_N = 0$ and $\pi_D = n\pi_M$: the gains from collusion are now perfectly correlated to those from deviation, and we can consider a univariate distribution of types. Cartel types refer to the gains from collusion, due for example to different marginal cost levels, with $\pi_M \in [\underline{\pi}_M; \overline{\pi}_M]$; the AA does not observe cartel types but knows their distribution $g(\pi_M)$, and is not able to condition the policy chosen to some observable that can make it contingent on a message. In other words, the AA sets a single combination of policy parameters taking into account that there exist many cartel types in the economy.

Under the assumption of Bertrand competition the standard critical discount factor when antitrust is absent, $\delta^* = (\pi_D - \pi_M)/(\pi_D - \pi_N)$, is $(n-1)/n$. We can rewrite the

in a larger amount of money collected through fines. In a world where non-distortionary transfers are not available, this would be an additional advantage of such rules.

relevant loci as:

$$\alpha_{CR} = \frac{\gamma_M(1 - n + n\delta)}{\gamma_M n\delta} = (\delta - \delta^*)/\delta$$

which does not depend on cartel's type,

$$\alpha_{NC}(p) = \frac{(1 - \delta)n(\delta - \delta^*)\gamma_M}{\delta[pF(1 - \delta) + \gamma_M(p - n(\delta - \delta^*))]}$$

and

$$\bar{p} = \frac{\gamma_M}{(\gamma_M + F(1 - \delta))}$$

which are both increasing in γ_M. Moreover, α_{CR} is always above α_{NC} at $p = 1$. Hence, when $R < F$ we can distinguish 5 regions which are represented in figure 5.

Figure 5 about here

In region A all types choose CNR and the corresponding welfare is $W_A = \frac{\alpha p}{1 - \delta + \alpha\delta}E(K)$ where $E(K)$ is the expected value of the gains from deterrence given the distribution of types $g(\gamma_M)$. In region B all types choose CR with $W_B = \frac{\alpha}{1 - \delta + \alpha\delta}E(K)$ while in E all types abstain from collusion and welfare is $W_E = E(K)$. In region C some types choose CNR and others CR: let γ_M^C be the type whose $\bar{p}$ equals the p chosen by the AA in region C: all types lower than γ_M^C collude and reveal while the cartels more profitable collude and don't reveal. The expected welfare is therefore

$$W_C = \frac{\alpha}{1 - \delta + \alpha\delta}\int_{\underline{\gamma}_M}^{\gamma_M^C} K(\gamma_M)g(\gamma_M)d\gamma_M + \frac{\alpha p}{1 - \delta + \alpha\delta}\int_{\gamma_M^C}^{\overline{\gamma}_M} K(\gamma_M)g(\gamma_M)d\gamma_M$$

Analogously, in region D lower types abstain from collusion and higher types collude and don't reveal, with the threshold type γ_M^D such that the actual policy combination in D lies on that type's $\alpha_{NC}(p)$ curve. The expected welfare is then

$$W_D = \int_{\underline{\gamma}_M}^{\gamma_M^D} K(\gamma_M)g(\gamma_M)d\gamma_M + \frac{\alpha p}{1 - \delta + \alpha\delta}\int_{\gamma_M^D}^{\overline{\gamma}_M} K(\gamma_M)g(\gamma_M)d\gamma_M$$

When $R = F$ only regions A,D and E exist, defined by the set of $\alpha_{NC}(p)$ curves which extend up to $p = 1$.

The analysis of the optimal policy proceeds in three steps, which are developed analytically in the Appendix. First, the iso-welfare curves in each of the five regions A-E are derived; then, we check how the same welfare level is obtained passing (eventually) from one region to the neighbouring one, distinguishing whether fine reductions are given or not; finally, comparing the two cases, it is selected whether reduced fines R allow to save enforcement costs, defining a set of iso-welfare curves along which Leniency Programs are optimally used.

The result of this analysis is shown in ¯gure 6.a: the lower bold curve is the iso-welfare (cost minimizing) curve setting R = 0, which passes through the regions A-C-B. The upper bold curve passing in the D-C-B regions entains the use of reduced ¯nes only in a subset of the B and C regions. The policy combinations (R; ®; p) which minimize the cost of reaching the same expected welfare are summarized in a map of iso-welfare curves which are not convex: as before, we have to convexify them excluding from the set of possible equilibrium outcomes those policy combinations which belong to the non-convex portions of the indi®erence curves.

Given the map of indi®erence curves that minimize the cost of a given expected welfare, we exclude those portions which can never be selected given our linear budget constraint[26]. For the indi®erence curves in the A region we obtain a subset of points E_{CNR} analogous to the one obtained in the single type case already discussed. In region C we ¯nd a subset of points E^{CR}_{CNR} in which some types choose CR and higher types choose CNR. In region B we select only the boundary to the left, which corresponds to the E^{CR}_{CNR} case when all types opt for CR. A subset of D, E^{NC}_{CNR} is obtained where low type select NC and high types chooase CNR, and ¯nally from region E we select the lower bound. Once excluded the non-convex portions of the iso-welfare curves, the optimal policies for given budget constraint can be established along the same lines of Proposition 4's proof. We summarize the results in the following Proposition, which is respresented in ¯gure 6.b.

Proposition 5 Consider the optimal policy under asymmetric information given the budget constraint and the distribution of cartel types.

- If the budget constraint passes through region E, the optimal policy implements NC for all types at a tangency point between the budget constraint and the ®$_{NC}$(p) curve of the highest type.
- If the budget constraint passes through region D and is tangent to an indi®erence curve in E^{NC}_{CNR}, the optimal policy is at the tangency point with no ¯ne reduction, and implements a CNR-NC outcome according to the di®erent types.
- If the budget constraint passes through region A and is tangent to an indi®erence curve in E_{CNR}, the optimal policy is at the tangency point and implements CNR for all types.
- If the budget constraint passes through C and is tangent to an indi®erence curve in E^{CR}_{CNR}, that is the optimal policy and implements a CNR-CR outcome.
- In all the other cases the optimal policy implements CR for all types setting p equal to the $\bar{p}$ of the highest type along the budget constraint.

[26] Any convex budget set, as that obtained under the assumption of decreasing returns to enforcement, would allow a similar exercise.

Figure 6.a and 6.b about here

We can give an explanation of the result above considering the sequence of optimal policies when the budget constraint becomes steeper and steeper as a result of an increase in the cost of independent prosecution (higher w_p). For low values of w_p the policy implements full deterrence ex-ante for all types. As the budget constraint rotates toward the origin we initially move to a CNR-NC mixed outcome with no ¯ne reduction, in which the more pro¯table cartels are not deterred. Granting ¯ne discounts in this case would shift low types from NC to CR: the pro-collusive e®ect of Leniency Programs would dominate reducing welfare. However, when the fraction of low types which choose NC shrinks further, reduced ¯nes are introduced, inducing all types to collude and reveal. In this case the improvement in prosecution allowed by reduced ¯nes becomes predominant. A further increase in w_p moves the equilibrium outcome in the E^{CR}_{CNR} region with an increasing portion of high types that choose CNR while low types collude and reveal. The implicit cost of the Leniency Programs, which forces the AA to commit resources to independent prosecution to make it a credible threat, becomes heavier and heavier as the resources left to open inquiries decrease and as the fraction of types which are induced to reveal shrinks. In the end we move to the E_{CNR} region, abandoning the Leniency Program. Hence, the optimal policy is determined, in a sense, by the relative importance of the pro-welfare e®ect of Leniency Program, that allows to obtain more e®ective ex-post desistence, and the welfare decreasing e®ects of reduced ¯nes: the incentive to collude (and reveal) instead of abstaining from collusion, and the need to sink resources to make independent prosecution credible, which reduces the probability of opening an inquiry and of obtaining ex-post desistence.

5 Conclusions

In this paper we have analyzed the e®ects of Leniency Programs on the incentives of ¯rms to collude and to reveal information that helps the Antitrust Authority to prove illegal behaviour. The benchmark regime gives to any ¯rm a reduction in ¯nes even if revelation occurs after an investigation is opened. We show that reducing the expected ¯nes may induce a pro-collusive reaction: combinations of policy parameters which, without Leniency Programs, would prevent collusion, may induce ¯rms to collude (and reveal if monitored) when ¯ne reductions are given. Hence, if the resources available to the AA are su±cient to prevent collusion using full ¯nes, Leniency Programs should not be used.

However, when the AA has limited resources, Leniency Programs may be optimal in a second best perspective. Fine reductions, inducing ¯rms to reveal information once an investigation is opened, increase the probability of ex-post desistence and the expected welfare gains. The optimal scheme requires maximum ¯ne reductions and a shift of resources from prosecution to monitoring.

A ¯xed amount of resources, however, must be committed in any case to make independent prosecution a credible threat, since no firm would reveal if it expects that the AA is unable to prove them guilty. When independent prosecution is very costly, too few resources are left to general monitoring, which in the end determines the effectiveness of Leniency Programs. In this case it may become more convenient to shift back to a full fines regime with a more favourable mix of policy parameters.

We have then compared our benchmark regime with alternative sets of rules: the first allows to give fine reductions only to firms which cooperate with the Antitrust Authority before an inquiry is opened, as initially established in the US policy in 1978, and similar to the approach followed by the EC Notice on the non-imposition of fines, and we proved this regime to be inferior with respect to our benchmark case. We have then considered other rules which restrict the set of the firms that can benefit from a Leniency Program. We showed that by granting a fine reduction only to the first firm which cooperates with the AA the perverse pro-collusive effect of the Leniency Program would be reduced without softening the incentives to reveal information. Better still, the AA might target a specific firm and allow only this one to benefit from the reduction. The intuition for this result, which makes the Leniency Program even more effective, lies in the asymmetry that the policy introduces among otherwise identical firms, between the entitled firm and the excluded ones: the latter would more often prefer to abstain from collusion rather than join a cartel together with a likely cheater.

Finally, the case of multiple cartel types has been considered: the AA is assumed to be unable, for informational or institutional reasons, to implement Leniency Programs contingent on cartel's type, and therefore has to set general rules. For instance, the AA cannot shape the policy to the conditions of each specific industry, but has to choose a general rule of behaviour, obtaining different effects in different industries. Then, according to the position of the budget constraint in the set of policy combinations, we characterized the optimal policy: it turns out that the policy parameters and the regime of full or reduced fines are chosen according to the relative weight of the three effects described above, where the weights depend on the share of types which choose the different equilibrium outcomes (no collusion; collusion; collusion and revelation).

We believe that, despite the simple setting, our paper sheds some light on the desirable features of leniency programs, and suggests some changes in the EC leniency policy. First of all, if it is optimal to use a leniency program (as in the realistic case where the antitrust agency has limited resources), then the program should be as generous and certain as possible with the firms which provide fresh evidence that establishes the existence of a cartel. In contrast, the EC policy of keeping some degree of discretionality instead of granting automatic and total reduction of fines even to those firms which fulfil all the (strict) conditions laid down in the EC Notice undermines the success of the leniency program, as it does not give certainty to the prospective cooperating firm and reduces the incentive to break the cartel.

Likewise, some of the conditions required by the EC policy are too strict. For instance, a firm must "maintain continuous and complete co-operation throughout the investigation" to be entitled to have a very substantial reduction (more than 75%) of the ¯ne. This has led the Commission to give only a 50% reduction to a ¯rm, Tate & Lyle, which had spontaneously brought conclusive evidence of a cartel to the attention of the Commission (at a time when the Commission did not even suspect the existence of an agreement), but had later (partially) contested some of the allegations made by the Commission[27]. The strict wording and application of the Notice will reduce the incentive of the ¯rms to reveal information[28].

Furthermore, our analysis indicates that a leniency program should be equally applicable to information disclosed before and after an investigation has started, whereas the EC policy does not create enough incentives for post-investigation disclosure of information. It gives only 50-75% reduction of the ¯nes for cooperation after an investigation has been undertaken already but only if such an investigation has failed to provide su±cient material for initiating a procedure leading to a decision.

The US experience (where after the 1993 policy revision a corporation is granted leniency after an investigation has begun provided that "the Division, at the time the corporation comes in, does not yet have evidence against the company that is likely to result in a sustainable conviction" - point B.2.) clearly shows that extension of the leniency program to post-investigation amnesty (along with the automatic granting of the amnesty) is a crucial ingredient for success: "...under the old policy on average only one corporation per year applied for amnesty," (Spratling (1998, page 2) whereas under the revised policy, "Amnesty applications over the past year have been coming in at the rate of approximately two per month" (Spratling (1999, page 2).

So far, the leniency program of the EC has been applied to a very reduced number of cases, since its introduction in the end of 1996. There was no case in 1997 and only four in 1998 [29]. We believe that granting higher and automatic reductions of ¯nes and extending the leniency program to after-investigation cooperation would greatly increase the success of this policy.

[27] This is the case "British Sugar", EC Decision of 14 October 1998, published in the O±cial Journal of the EC, L76, 22 March 1999.

[28] Hornsby and Hunter (1997) also point out that the ¯rms do not have enough incentives to cooperate under the EC policy. Part of the problem is also due to the fact that the Notice cannot provide immunity from civil proceedings. Admission of an infringement leads to a formal Commission Decision on which an action for damages can be built, without the plainti® having to prove the infringement again. This problem does not exist in the US, where the cooperating ¯rms can resort to a consent decree.

[29] Information provided by an EC o±cial, Guerrin (1999). Of these, three regarded instances of minor cooperation (¯rms were given discounts for not having contested the Commission's allegations or for providing additional evidence which helped establishing the facts). These cases are "Alloy surcharge", EC Decision of 21 January 1998; "Pre-insulated pipes", EC Decision of 21 October 1998; "Greek Ferries", EC Decision of 9 December 1998. The fourth case was the "British Sugar" case reported in the previous footnote, which might be a discouraging precedent for ¯rms considering cooperation with the Competition Commission. According to Guerrin, in some further half a dozen current cases the Leniency Notice has been invoked. No further details were given for reasons of con¯dentiality.

References

[1] Antitrust Division (1994), Annual Report for Fiscal Year 1994.

[2] Bebchuk L. (1984), Litigation and Settlement under Imperfect Information, Rand Journal of Economics, 15, 404-413.

[3] Bingman A., Spratling G., (1995) Criminal Antitrust Enforcement, mimeo.

[4] European Union (1996), Notice on the Non-Imposition or Reduction of Fines in Cartel Cases, O±cial Journal, v.207, p.4.

[5] European Union (1999), Competition Policy Newsletter, No.1, February.

[6] Furse M., (1995), Article 15(2) of Regulation 17: Fines nad the Commission's Discretion, European Competition Law Review, p.110.

[7] Grossman G., Katz M. (1983), Plea-Bargaining and Social Welfare, American Economic Review, 7, 749-757.

[8] Guerrin M.(1999), Private communication. Electronic message of 5 March 1999.

[9] Hornsby S., Hunter, J. (1997), New incentives for "whistle-blowing": will the E.C. Commission's Notice bear fruit?, European Competition Law Review, 1, 38-41.

[10] Kobayashi B. (1992), Deterrents with Multiple Defendants: an Explanation for "Unfair Plea Bargains", Rand Journal of Economics, 22, 507-517.

[11] Marshall R., Muerer M., Richard J. (1994), Litigation, Settlement and Collusion, Quarterly Journal of Economics, 1, 211-239.

[12] Nalebu® B. (1987), Credible Pre-Tial Negotiation, Rand Journal of Economics, 18, 198-210.

[13] Reinganum J. (1988) Plea-Bargaining and Prosecutorial Discretion, American Economic Review, 78, 713-728.

[14] Reinganum J., Wilde L. (1987), Settlement, Litigation and the Allocation of Litigation Costs, Rand Journal of Economics, 17, 557-566.

[15] Schweizer U. (1989), Litigation and Settlement under Two-Sided Incomplete Information, Review of Economic Studies, 56, 163-178.

[16] Shavell S. (1989), Sharing Information Prior to Settlement or Litigation, Rand Journal of Economics, 20, 183-195.

[17] Spratling G.R. (1998), The corporate leniency policy: answers to recurring questions, speech of the Deputy Assistant Attorney General, presented at the Bar Association of the District of Columbia, February 16, 1999.

[18] Spratling G.R. (1999), Making companies an offer they shouldn't refuse, speech of the Deputy Assistant Attorney General, presented at the ABA Antitrust Section 1999 Spring Meeting, April 1, 1998.

[19] Venit J. (1996), EU Competition Law - Enforcement and Compliance: An Overview, Antitrust Law Journal, p.87.

Appendix

Proof of Lemma 1

If a firm reveals, it gets a payoff of $\pi_N/(1-\delta) - R$ independently of the action chosen by the other firms. If a firm does not reveal any information but at least one firm does, then the former firm receives a payoff of $\pi_N/(1-\delta) - F$. Hence, it is always (weakly) better to reveal if the other firms are expected to reveal, which establishes the existence of the "revelation" equilibrium . Finally, if no firm reveals any information, each firm receives an expected payoff

$$p\left(\frac{\pi_N}{1-\delta} - F\right) + (1-p)\frac{\pi_M}{1-\delta}. \qquad (15)$$

If a firm expects the others not to reveal, the best reply is trivially not to reveal if $pF < R$. If $pF \geq R$, when the other firms don't reveal, a firm prefers not to reveal as well if the payoff above is higher than $\pi_N/(1-\delta) - F$, which simplifies to $\delta \geq \bar{\delta}(p;F;R)$. Hence, in this case a "no revelation" equilibrium exists. Moreover, the same inequality implies that the "no revelation" equilibrium gives higher payoffs to all firms than the "revelation" equilibrium.

Proof of Proposition 2

We consider the decision to collude or deviate in both cases, when firm will decide to reveal if investigated, and when they will prefer not to cooperate with the AA.

- Case 1: $\delta < \bar{\delta}$. In this case firms reveal if an investigation is opened by the AA. Define π_R as the expected profit immediately before an investigation is opened. It is easy to see that:

  $$\pi_R = \alpha\left(\frac{\pi_N}{1-\delta} - R\right) + (1-\alpha)(\pi_M + \delta\pi_R)$$

  from which we obtain:

  $$\pi_R = \frac{(1-\alpha)\pi_M + \alpha\left(\frac{\pi_N}{1-\delta} - R\right)}{1-\delta(1-\alpha)}$$

If a firm decides to set the collusive price, then its expected discounted payoff will be:

$$V_{CR} = \pi_M + \delta\pi_R = \frac{\pi_M + \delta\alpha(\frac{\pi_N}{1-\delta} - R)}{1 - \delta(1-\alpha)}$$

If instead a firm decides to deviate from the collusive strategy, then its payoff is given by:

$$V_D = \pi_D + \frac{\delta\pi_N}{1-\delta}$$

Collusion can arise if $V_{CR} \geq V_D$, that is if the following condition is satisfied:

$$\delta \geq \frac{\pi_D - \pi_M}{(1-\alpha)(\pi_D - \pi_N) - \alpha R} \equiv \delta_{CR}(\alpha; R). \quad (16)$$

- Case 2: $\delta \geq \bar{\delta}$. In this case, firms anticipate that even if an investigation is started, no firm will reveal any information. Collusive outcome will be obtained unless the AA can prove the firms guilty of collusion.

  Write the expected profit immediately before knowing if an investigation is opened as:

  $$\pi_{NR} = \alpha[p(\frac{\pi_N}{1-\delta} - F) + (1-p)(\frac{\pi_M}{1-\delta})] + (1-\alpha)(\pi_M + \delta\pi_{NR})$$

  We can then obtain:

  $$\pi_{NR} = \frac{\alpha[p(\frac{\pi_N}{1-\delta} - F) + (1-p)(\frac{\pi_M}{1-\delta})] + (1-\alpha)\pi_M}{1 - \delta(1-\alpha)}$$

  If a firm follows the collusive strategy its expected discounted payoff is given by:

  $$V_{CNR} = \pi_M + \delta\pi_{NR} = \frac{\pi_M(1 + \frac{\delta\alpha(1-p)}{1-\delta}) + \delta\alpha p(\frac{\pi_N}{1-\delta} - F)}{1 - \delta(1-\alpha)}$$

  As before, a firm which deviates obtains a payoff:

  $$V_D = \pi_D + \frac{\delta\pi_N}{1-\delta}$$

  The inequality $V_{CNR} \geq V_D$ implicitly defines the locus of points $\delta_{NC} = \delta_{NC}(\alpha; p; F)$.

Proof of Proposition 4

We proceed in two steps: first we show, for each given outcome NC, CR, CNR, which is the associated optimal policy; second, we show the conditions under which a particular outcome is better than the others. If a NC outcome is implemented, we can have two cases: either the budget constraint is above the lower boundary of the A_{NC} region or it is tangent: in the latter case the tangency point with the α_{NC} curve is trivially the optimal solution; if the budget constraint is above that curve, the tangency point can still be suggested under a cost saving argument. In this case we set $R = F$ since granting fine discounts would shrink the A_{NC} region. If a CR outcome is implemented, the welfare gain depends only on α, which therefore must be maximized: we can therefore set $R = 0$, shifting to the left the $\bar{p}$ threshold, and setting $p = \bar{p}$; with p at its lowest level in the A_{CR} region we can set $\alpha = B/w_{\alpha} - (w_p/w_{\alpha})\bar{p}$ along the budget constraint. Finally if a CNR outcome is chosen, a tangency point between the budget constraint and the indifference curve in the A_{CNR} region must be chosen.

Consider now the choice among the three outcomes: since W_{NC} is always dominant for any set of policy parameters, if the budget constraint is not below the $\alpha_{NC}(p)$ curve, NC is the optimal outcome implemented. The choice between a CR and a CNR outcome is more complex, since both W_{CNR} and W_{CR} depend on the associated policy parameters, which, in turn, are different at the optimal points in the two regimes. Suppose the tangency point in the A_{CNR} region belongs to the subset E_{CNR}: from the definition of E_{CNR}, even if the budget constraint in its lower portion reaches the A_{CR} region, it passes through indifference curves lower than the initial one: hence, picking the tangency point in the E_{CNR} region and implementing a CNR outcome is the optimal policy. On the contrary, if the tangency point is in A_{CNR} but not in E_{CNR}, the budget constraint reaches the A_{CR} at a higher indifference curve, and a CR outcome is optimal.

The iso-welfare curves with heterogeneous cartel types

In the following three Lemmas we identify the iso-welfare curves when the AA faces heterogeneous cartels.

Lemma 6 The iso-welfare curves in each of the five regions have the following pattern:

- in E all the policy combination give the same welfare;
- in A and D they replicate the shape of the $\alpha_{NC}(p)$ curves;
- in B they are horizontal;
- in C they are decreasing;

Proof: Since W_E does not depend on the policy parameters, all the region correspond to the same expected welfare. In region A all types choose CNR. From our analysis of the single type case we already know that the iso-welfare curves when no type colludes have a shape similar to the α_{NC} curves (one for each type) and in the limit overlap with those. Hence, in region A the indifference curves replicate the α_{NC} curves shape. In region D high types choose CNR and lower types choose NC: as long as we move along the α_{NC} curve of type θ^{D}_{N}, the threshold type does not change and the first term in W_D is unchanged as well; moreover, we know that moving along a α_{NC} curve we keep the expected welfare for types choosing CNR constant. We conclude that the indi®erence curves in region D correspond to the $®_{NC}$ curves through it. In region B all types choose CR and the expected welfare depends only on ®, i.e. we have °at iso-welfare curves. Finally, in the C region high types select CNR and low types CR: since W_C increases when ® is higher (more frequent revelation) as well as when p increases (more e®ective prosecution and more types induced to reveal), the iso-welfare curve in the C region must be decreasing. 2

Notice that when no Leniency Program is used, only the regions A, D and E exist, and the result above states that all the curves in A (or D) never pass through another region. Hence, the three relevant sets of indi®erence curves are completely de¯ned. When $R < F$ all the ¯ve regions A-E exist; the Lemma above de¯nes the iso-welfare curves in each region, but now the iso-welfare curves in A (D) eventually continue through region C and B. Hence, we have to carefully check how the iso-welfare curves behave moving from one region to the other.

Lemma 7 Consider the case $R < F$. The iso-welfare curves passing through the regions A-C-B are continous and kinked at the boundaries of the A and B regions. The iso-welfare curves passing through the region D-C-B discountinously shift to the right passing from D to C.

Proof: We start by identifying the indi®erence curves that pass through the A-C-B regions. We already know, borrowing from the analysis of the single type case, that the iso-welfare curve jumps down from A to B, when all types choose CNR and then CR: however, from the de¯nitions of the expected welfare we can notice that W_C tends to W_A or W_B as the threshold type $¦^C_M$ tends to $\underline{¦}_M$ or $\overline{¦}_M$. Hence, the indi®erence curves are now continuous; it is easy to check also that they are kinked at the boundaries of the A and B regions, with the indi®erence curve steeper in C than in the other two regions[30].

Consider next the indi®erence curves passing through the D-C-B regions: we already established that in D the curves replicate the $®_{NC}$ curves shape, with some types choosing CNR and others NC; once moving into the C regions, some types still select CNR while others CR. Since the welfare associated to NC is higher than that when CR occurs, it must be that, moving from region D to region C along a iso-welfare curve, less types choose CNR. That requires a discontinuous jump to the right of the iso-welfare curve once entering in the C region. 2

Figure 6.a shows the two cases of indi®erence curves, one through A-C-B and the other through D-C-B. In the two Lemmas above we have completely characterized the iso-welfare curves when a Leniency Program is introduced and when it is not. Our next step is to verify when it is convenient to o®er reduced ¯nes in order to reach a certain expected welfare. This exercise correponds to comparing the iso-welfare curves in the two cases, selecting in the di®erent regions the lower one.

[30] The concavity or convexity of the indi®erence curve in C cannot be stated in general, since it depends on the distribution of types $g(¦_M)$. In what follows we consider the case of concave indi®erence curves, while the extension to the convex case is left to the reader.

Lemma 8 When the iso-welfare curve with R = 0 passes through the A-C-B regions, it is always optimal to use the leniency Programs. When the iso-welfare curve with R = 0 lies in the D-C-B region, the Leniency Programs are optimal only in a subset of the C and B regions.

Proof: Since R = 0 is the more e®ective way of inducing CR, we compare the case R = F and R = 0, selecting the lower of the two iso-welfare curves. Since in A and D the iso-welfare curves are the same in both regimes, our problem amounts to selecting the lower curve in regions C and B. This can be done by distinguishing the case in which the indi®erence curve with ¯ne reduction passes through the A-C-B region and that in which it lies in the D-C-B areas. Comparing the full and reduced ¯nes indi®erence curves is immediate for the A-C-B case: in the A region they overlap while in the C and B region the iso-welfare curve is lower when R = 0, as shown in ¯gure 6.a. Hence, the iso-welfare curve through A-C-B is that identi¯ed when the Leniency Program is used.

More complex is the comparison of the indi®erence curves with and without ¯ne reductions when the former passes through the D-C-B region. In this case, in fact, the iso-welfare curve with ¯ne reductions jumps to the right entering the C region, while with no Leniency Program the indi®erence curve, which is the same as before in the D area, goes on smoothly in the C region[31]. Hence, entering the C region from the top, the lower indi®erence curve is initially the one associated with no ¯ne reduction. It may happen, as shown in ¯gure 6.a, that continuing along it, the indi®erence curve with ¯ne reductions becomes the lower one for a while. Finally, moving further to the right, the indi®erence curve with full ¯nes lies again below that with ¯ne reductions. For higher levels of the expected welfare, the indi®erence curve with full ¯nes always dominates that with reduced ¯nes. Hence, when we consider the indi®erence curves for increasing value of the expected welfare, as long as we are in the A-C-B region we use Leniency Programs, while entering the D-C-B region we adopt reduced ¯nes only with a subset of policy parameters (®; p), as shown in ¯gure 6.a. 2

[31]Strictly speaking, with no Leniency Program no C region exists; hence we refer to the C region as those policy combinations de¯ned in case of ¯ne reductions.

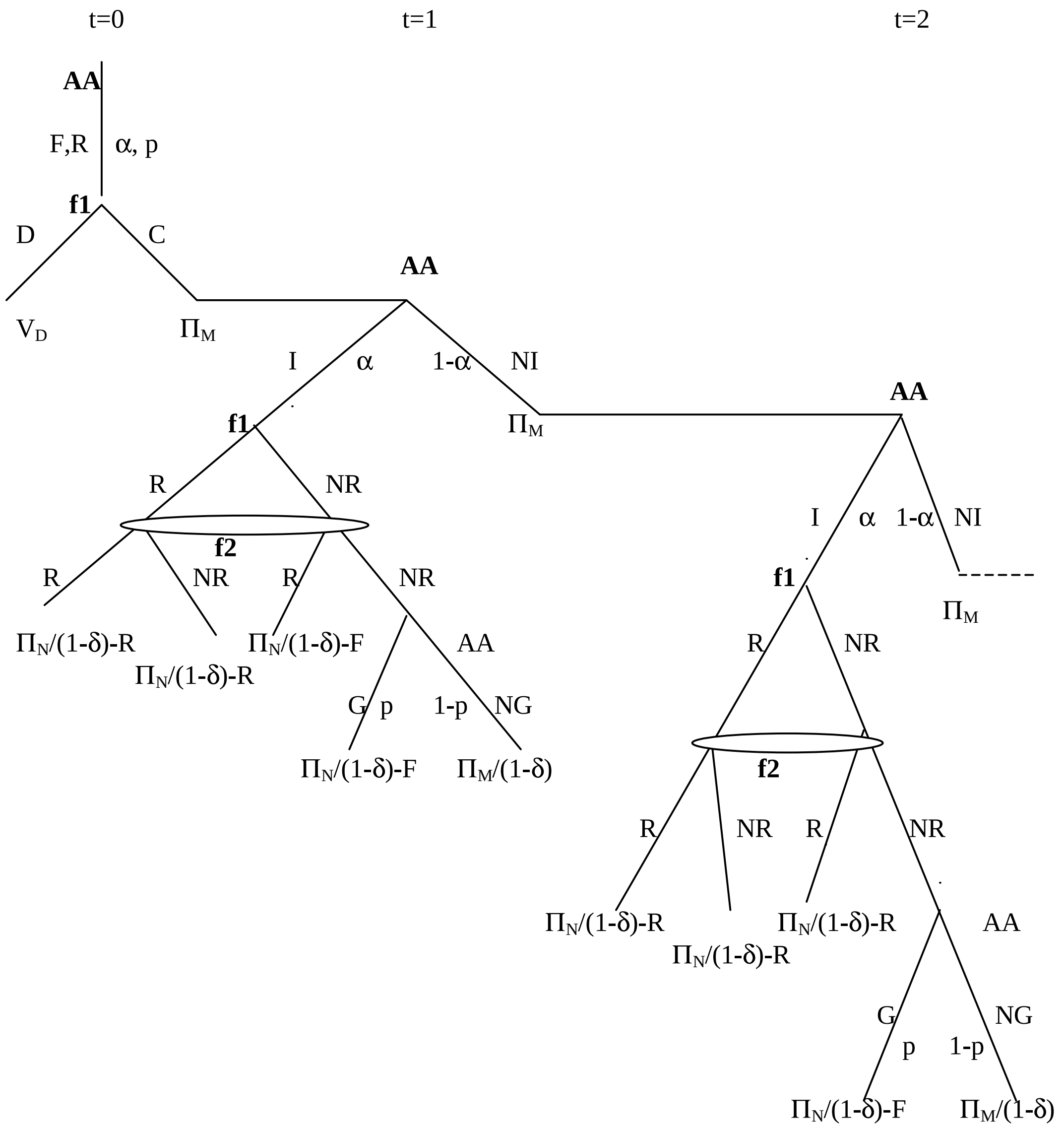

Figure 1: The game tree (D: deviation; C: collusion; I: investigation; NI: no investigation; R: reveal; NR: not reveal; G: guilty; NG: not guilty)

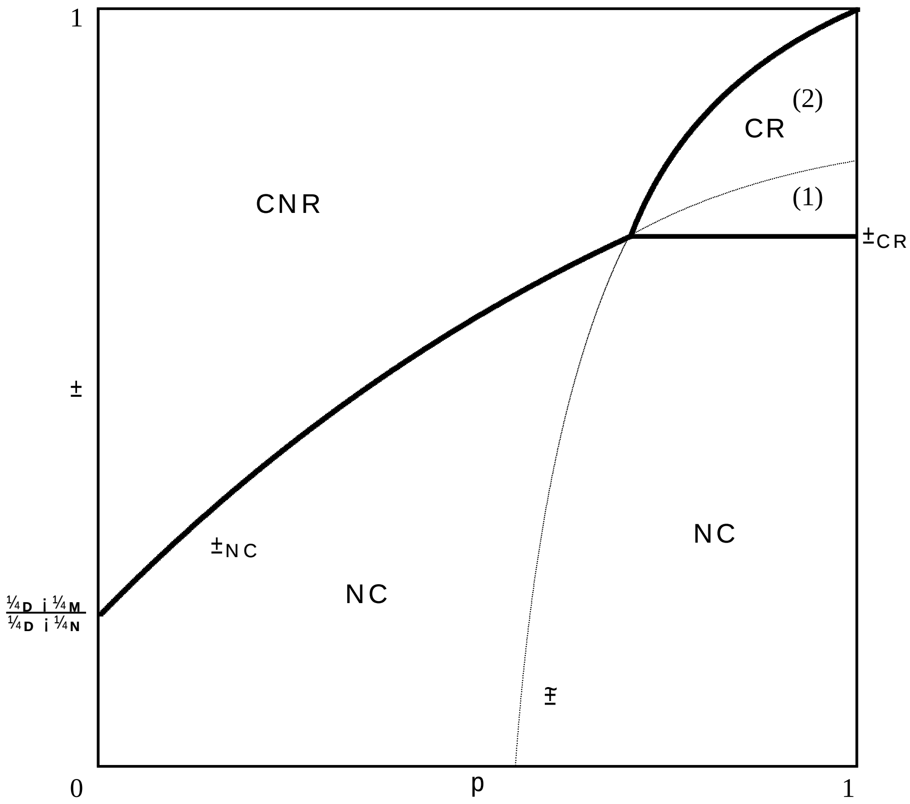

Figure 2.a - Implementable allocations

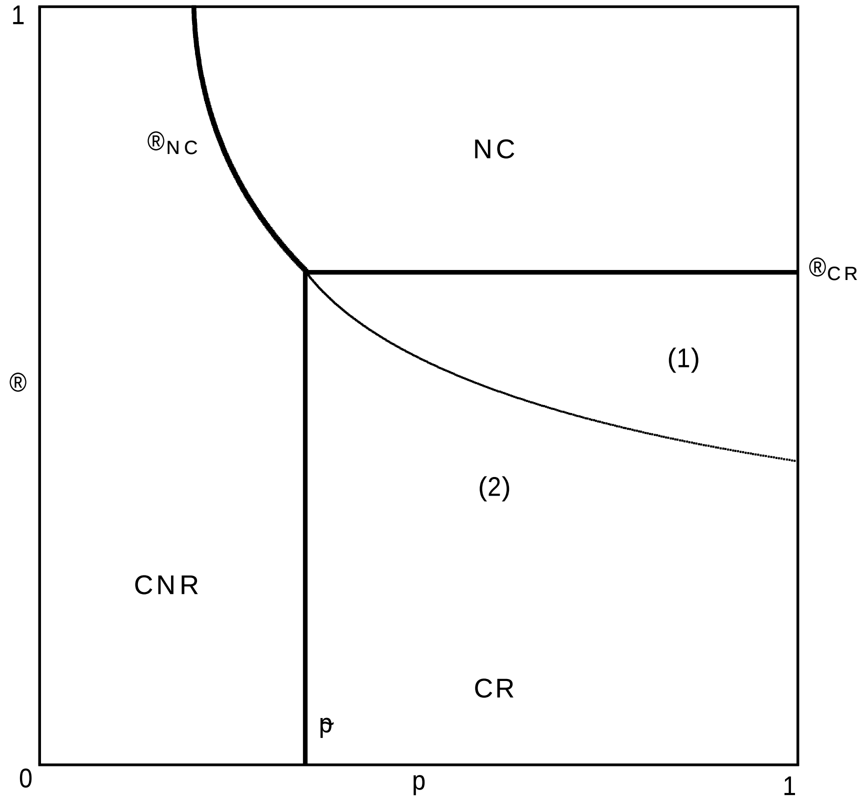

Figure 2.b - Implementable allocations

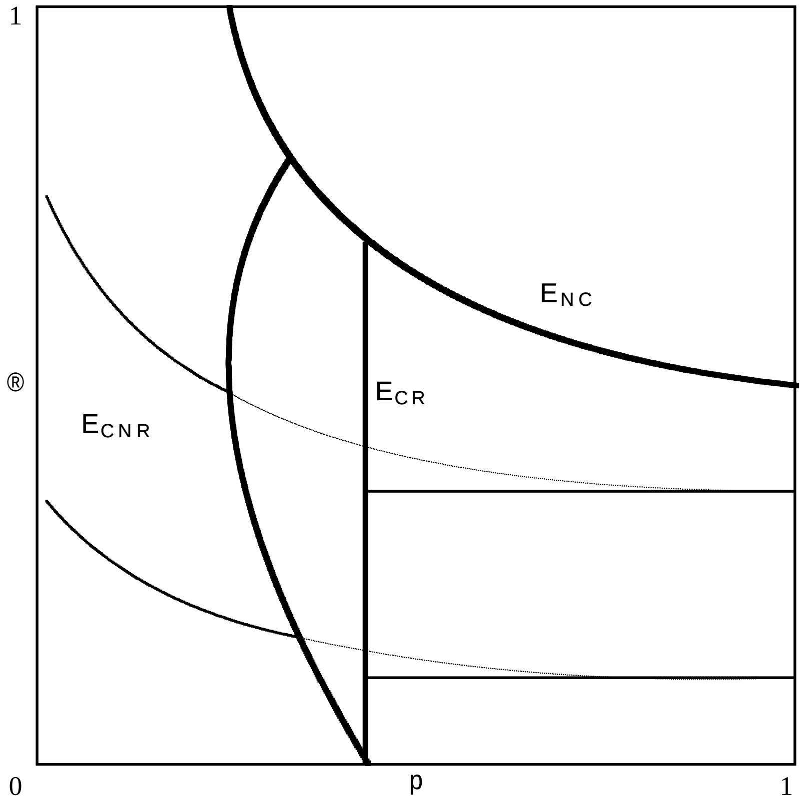

Figure 3 - Equilibrium policy combinations

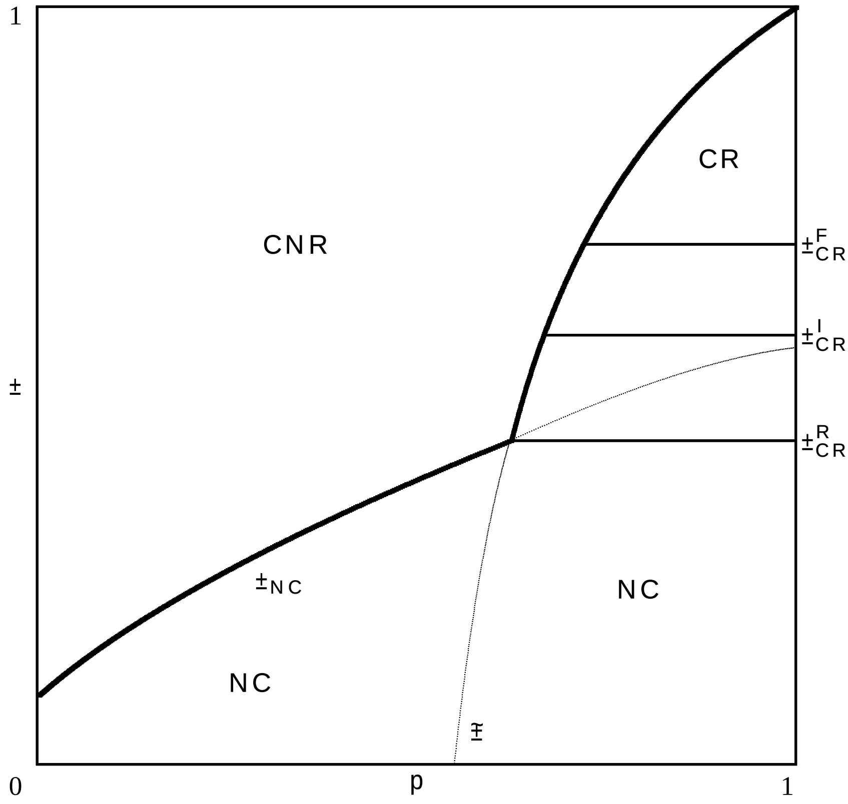

Figure 4 - Alternative Leniency rules and implementable allocations

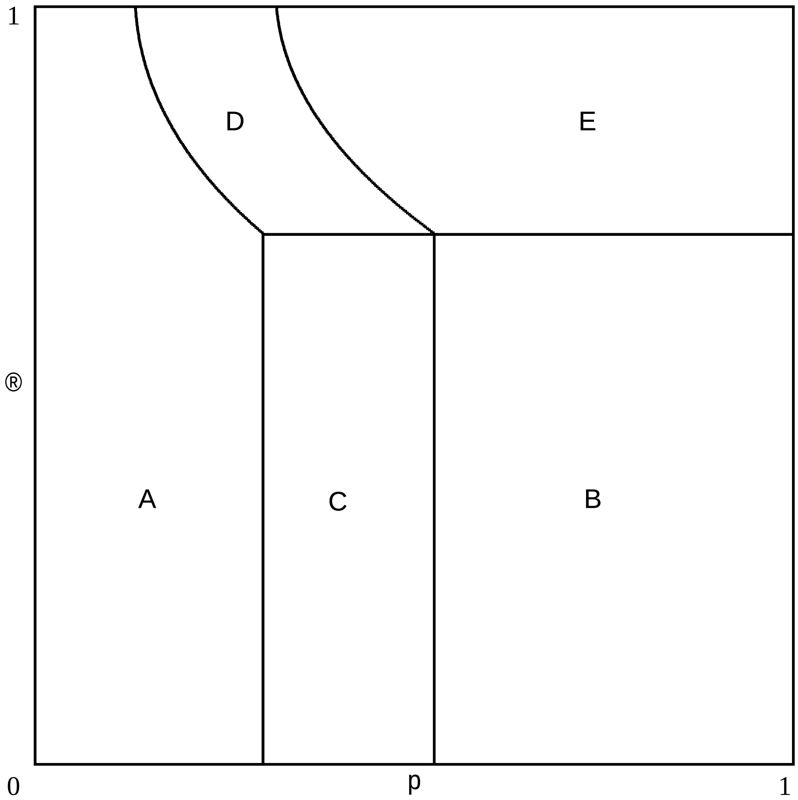

Figure 5 - Multiple types and implementable allocations

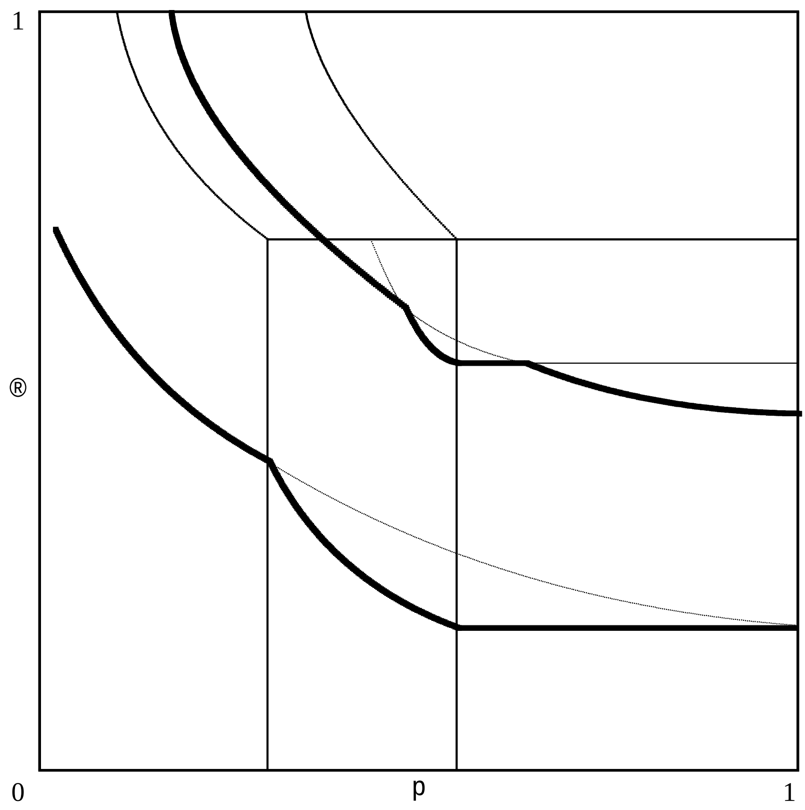

Figure 6.a - Iso-welfare curves with multiple types

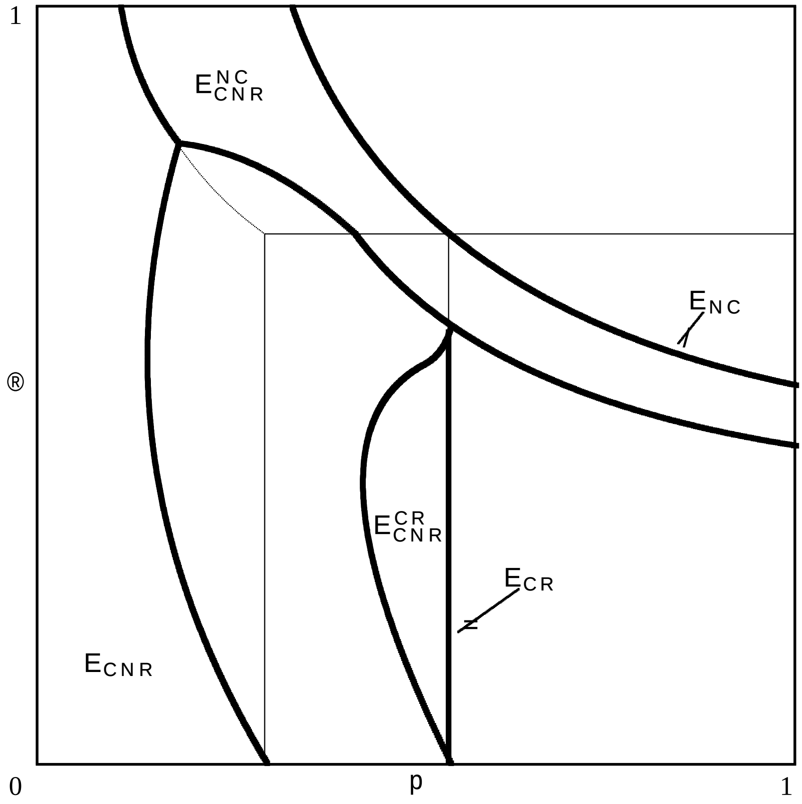

Figure 6.b - Equilibrium policy combinations with multiple types