Getting Polluters to Tell the Truth*

Marcelo Caffera†
Universidad de Montevideo

Juan Dubra‡
Universidad de Montevideo

March, 2005

Abstract

We study the problem of a regulator who must control the emissions of a given pollutant from a series of industries when the firms' abatement costs are unknown. We develop a mechanism in which the regulator asks firms to report their abatement costs and implements the most stringent emissions standard consistent with the firms' declarations. He also inspects one of the firms in each industry which declared the cost structure consistent with the least stringent emissions standard and with an arbitrarily small probability, he discovers whether the report was true or not. The firm is punished with an arbitrarily small fine if and only if its report was false.

This mechanism is simple, is implementable in practice, its unique equilibrium is truth telling by firms, it implements the first best pollution standards and shares some features of the regulatory processes actually observed in reality.

Keywords: Efficient Emissions Standards, Command and Control, Truth Telling, Undominated Nash Implementation.

Journal of Economic Literature **Classification numbers:** D02, D78, D82, Q20, Q52, Q53.

1 Introduction

In this paper we study the problem of a regulator who must control the emissions of a given pollutant from a series of industries. He wants firms to produce the optimal amount of pollution, when both the firms' abatement costs and the costs of pollution to society are considered. Such a

*We thank Atila Abdulkadiroglu, Ezequiel Aguirre, Anil Arya, Jean-Pierre Benoît, Carlos Chavez, Federico Echenique, Jeff Ely, Jonathan Glover, Ana María Ibáñez, Matt Jackson, Larry Kotlikoff, Carlos Lacurcia and Stephen Morris.

†Email: marcaffera@um.edu.uy

‡*Correspondence address*: Departamento de Economía, Universidad de Montevideo, Prudencio de Pena 2440, Montevideo 11600, Uruguay. Email: dubraj@um.edu.uy

regulator faces a fundamental problem faced by every regulator worldwide: that he rarely knows the exact nature of the pollution abatement technology of firms, which of course influences the optimal pollution level to be chosen. The regulator must therefore rely on whatever he can learn about firms' costs from the information they are willing to provide. Given the importance of the problem of regulating polluters, the issue of how to truthfully extract information about their costs has been at the heart of both academic and policymaking discussions for almost three decades.

We posit a model in which the regulator asks firms to declare what their cost functions are and uses these announcements to set an emissions standard for each industry: a maximum allowable level of emissions for every firm in that industry. After receiving the reports, the regulator implements in each industry the most stringent standard consistent with the declarations of the firms in that industry. He also inspects one of the firms in each industry which declared the cost structure consistent with the least stringent emissions standard (the firms most likely to be lying). With an arbitrarily small probability, he discovers whether the report was true or not. A firm which was sampled is punished with an arbitrarily small fine if and only if its report was false.

This mechanism has several important features. First, it is very simple, and therefore applicable in practice. In fact, as we will discuss later in more depth, it is very similar to the mechanism actually used in several countries, including the US National Pollutant Discharge Elimination System. Second, it fully implements truth telling by the firms, and results in the regulator setting the efficient standard in each industry. That is, since the unique equilibrium of this game is for firms to tell the truth, the informational asymmetry disappears, and the total welfare of society is maximized. Finally, a third advantage of the mechanism is that it is budget balanced: it implies no costs for the regulator.

There are other studies that have proposed mechanisms that both implement truth telling by the firms and result in an efficient level of pollution. The three most relevant works in this area are Kwerel (1977) and Dasgupta, Hammond and Maskin (1980) and Spulber (1988). The problem with these prior studies is that one does not observe the proposed mechanisms in practice. We believe that there are two main reasons why those mechanisms are not observed in reality. Moreover, our mechanism is free of those problems. The first reason why we don't observe those mechanisms in reality is that they are complicated. This has been a standard criticism about the literature of optimal mechanism design. We believe that another reason why previously proposed mechanisms are not observed is that they are based on taxes, subsidies, or tradeable permits and these types of instruments have several implementation problems as compared to classic "command and control" instruments. Although these types of instruments have been used recently, they have applied only in very specific contexts, and their implementation has been slow for several reasons. For example, regulators are not educated in environmental economics and do not see the advantages of these instruments in terms of cost-effectiveness and efficiency; they see "command-and-control" instruments as stronger statements of support for environmental protection. Moreover, regulators

usually think that it is immoral to let firms pollute just because they paid some taxes, or because they purchased pollution permits. Policymakers may also be reluctant to impose further costs on firms because of the impact on employment. Also, incentive-based instruments shift control decisions from regulatory staff to polluting firms, possibly affecting their prestige and job security.[1]

In Section 6 we will discuss the relationship of our mechanism with the literature on implementation, but it suffices here to stress two points. First, the implementability of the regulator's rule in our setting does not follow from any of the existing theorems in the literature. Second, and most important, our focus is not on the novelty of the theoretical arguments in the implementation of the regulator's rule, but on the possiblity of actually implementing it in real contexts.

We have argued that our mechanism is simple, shares some features of some regulatory practices around the world, implements truth telling and the efficient level of pollution, and is budget balanced. Also, we have argued that one of the reasons why one does not observe in practice alternative mechanisms that have been proposed in the literature is because they were complicated and relied on taxes and subsidies, which may be too difficult to implement for regulators. We now turn to the discussion of our assumptions.

2 Discussion of Assumptions

Our model is very similar to that in Kwerel (1977) and Dasgupta et al. (1980). In some dimensions our model is more general, but we make two additional assumptions.[2] First, we assume that if the regulator samples one firm, it can find out, with probability ε, for ε arbitrarily small, whether the report of abatement costs was true or not. Second, we assume that in each of m industries there are at least two firms with the same cost functions.

With the first assumption the asymmetry of information between the regulator and the firms ceases to be absolute. The assumption is quite weak for at least three reasons. First, we assume that the regulator inspects and samples just one firm out of a potentially large pool. Second, we assume that in case the inspection is successful and it provides some information, the regulator only learns whether the report was true or not, but in case of a false report, he does not get to know the true cost function. Third, and most important, the regulator only finds out whether the report is true or not with an arbitrarily small chance. That is, we fix any $\varepsilon > 0$, and the regulator only learns whether the report is true with probability ε.

Our assumption that the asymmetry of information is not absolute is also a reasonable one in the context we study. First, regulators worldwide engage in controlling or monitoring the statements of

[1] These and other arguments are well documented in the literature. See for example Bohm and Russell, 1985, Russell and Powell, 1996, Lewis, 1996, Keohane, Revesz and Stavins (1998).

[2] Like both these works, our model can be applied more generally, and not just to the problem of a regulator trying to fix the right level of pollution.

polluters about the abatement technology to be used, so our assumption reflects a common practice. In the US for example, before starting their operations firms are required to present an exhaustive description of their production processes, abatement technology and costs in order to obtain a pollution discharge permit.[3] Second, this common practice is well founded, since the regulators can check each piece of information provided by the firm, and assess its validity, or even in some cases be more proactive by pointing out to firms how other businesses have coped with the same abatement problems. Engineers from the EPA study the different abatement technologies available to a particular type of industrial activity and then establish effluent standards for each category of polluter and place of discharge (See Field, 1997). Since the regulation, and the standard-setting, occur at a basic "process level" and not at a more complicated "plant-level," the processes involved are standard across industries, and the regulator has a deep knowledge about costs as illustrated, for example, in the following quotation from the EPA (1992).

> "The document provides a generic process-by-process assessment of pollution prevention opportunities for the Kraft segment of the pulp and paper industry. The process areas covered are: wood yard operations, pulping and chemical recovery, pulp bleaching, pulp drying and papermaking, and wastewater treatment. These process areas are further broken down by specific process (e.g., oxygen delignification as one specific process under the pulping and chemical recovery area). For each specific process there is a description, a cost estimate, a discussion of applicability, and estimate of environmental benefits."[4]

Both the way the regulatory process takes place, and the depth of the knowledge of the regulator about each individual process suggest that the asymmetry of information between firms and the regulators is not absolute, so that our assumption seems appropriate.

Our second assumption, that there are at least two firms in each industry which have identical cost functions also follows from the way the regulatory process works (i.e. setting emissions standards on a process by process basis). If a firm buys cows and delivers leather shoes, it won't have the same abatement costs as a firm that buys cows and delivers leather seats for cars. But both firms will first produce raw hides and then tan the leather. Since both firms need to abate its pollution levels at each individual task, each of which is also undertaken in other firms producing different goods, our assumption reflects the fact that even very complicated production processes are based on some elemental processes that are repeated in several firms even across industries.

[3] Several countries have copied extensively the US NPDES, including our own Uruguay. Most of such systems share the "inspection" features of the US system that we are interested in.

[4] Similar quotations can be found for other industries. See for example EPA (2002) for the iron and steel industries and their process by process regulation.

Another reason why the assumption of at least two firms per industry is not so restrictive is that our exact same model would apply if it was common knowledge that costs in the same industry are just "vertical" translations of each other. That is, if firm 1 has a cost function of c, and firm 2 a cost function of $c + k$, those cost functions are "identical" as far as our mechanism is concerned. Therefore, if a firm in California and a firm in NY buy their abatement technology from a firm in NY, and the price in California is just the price in NY plus shipping, those two firms can be modeled as having identical costs. Finally, as we will argue in Section 5, even if there are some firms that have cost functions that no other firm in the whole economy share, our mechanism can still be used. Suppose that the regulator can estimate the cost functions of these firms and produce estimates which are "close" to the truth. Then, the unique equilibrium of our mechanism (when it is applied among the firms in industries with at least two firms) is still truth telling, and the standards set for each industry are "close" to the first-best, complete information, ones.

Another, less disputable, assumption that we make is that the regulator can fine the firms for lying. This is consistent with the practice of pollution regulators worldwide. In Uruguay, for example, as a consequence of "forgery" in the cost declaration, the person in charge of filling the reports about the abatement technology can be imprisoned. Another potential punishment is the temporary closing of the plant. Similar practices are common elsewhere. It is worth emphasizing that for our mechanism to work, the fine can be arbitrarily small. If fines were large, even a small probability of a false report being uncovered would suffice to make truth telling a dominant strategy. In our mechanism the fine is used exclusively for breaking ties.

We also assume that total damages to society are known or can be estimated. Although this has been the standard assumption in this branch of the literature (see Kwerel (1977) and Dasgupta et al., 1980) it is quite strong. As we will argue later, however, our mechanism is robust to whether the regulator knows total damages exactly, or approximately, or just wants to set a total level of emissions for the whole economy. The first extension is relevant if one is able to estimate total damages to society approximately, and is concerned that the emissions standards will be approximately correct. We show that that is indeed the case: our mechanism still fully implements truth telling, and if the regulator's estimate of total damages are close to the true damages, then the emissions standards that result from our mechanism are close to the ones that would be implemented if the regulator knew exactly the damages to society and abatement costs. In a second relaxation of the assumption that the regulator knows damages, we investigate how our mechanism fares when the regulator does not know, or is not interested in damages to society, but rather on achieving a certain level of emissions for the whole economy. This extension is important because, for example, the adoption of the Kyoto Protocol implies that the regulatory agencies must find the most efficient way to achieve a certain level of emissions for the economy as a whole. We show that our mechanism can be used to determine the standards which minimize the total cost to society of complying with, say, the Kyoto standards.

In this note we are only concerned with the problem of setting the right emissions standards. The enforcement of those standards is a different issue, and we therefore omit its study. Our mechanism does not assume that there is perfect enforcement, only that higher emissions standards are better for firms. If there is perfect enforcement, then our mechanism maximizes total welfare to society. If there isn't, the emissions standards are the correct ones, but if firms violate the standards, welfare is not maximized, and the regulator must try to maximize compliance subject to its enforcement budget (see footnote 5 for more on this issue).

In Section 5 we discuss three variations of our assumptions with which the mechanism still fully implements truth telling.

3 The Model

There are m industries and n_i, for $i = 1, ..., m$, firms in each industry. Firms in $I_1 = \{1, ..., n_1\}$ are those in industry 1, firms in $I_2 = \{n_1 + 1, ..., n_1 + n_2\}$ are those in industry 2 and so on. Each industry has at least 2 firms.

Let $\mathcal{C}$ be the set of continuous and weakly decreasing functions $c : \mathbf{R}_+ \to \mathbf{R}_+$. Each firm in industry i can abate its pollution level using an abatement technology which has a cost of $c_i(\cdot) \in \mathcal{C}$. That is, $c_i(x_j)$ for firm j polluting a level x_j in industry i is the difference in profits from (a) not engaging in abatement, and (b) abating its potential pollution to level x_j.[5] If the potential pollution level in industry i is $\overline{X}_i$, this is modeled as $c_i(x) = 0$ for all $x \geq \overline{X}_i$. Note that all firms in each industry have the same cost function.

The cost function c_i is unknown to the regulator. He only knows that $c_i \in \mathcal{C}$ for $i = 1, ..., m$. In the mechanism of this paper, the regulator asks firms to report their cost functions. In spite of the informational asymmetry, the regulator can inspect one firm. With probability $\varepsilon > 0$ he finds out whether the report was truthful or not, with probability $1 - \varepsilon$ the inspection is inconclusive. In case the regulator discovers that the report was not true, he does not find out the true c_i, but only that the report was false.

The total damages to society coming from pollution are a continuous and increasing function $D : \mathbf{R}_+ \to \mathbf{R}_+$ where total damages are given by $D(X)$ and X is the total pollution from every firm in every industry:

$$X_i = \sum_{j \in I_i} x_j, i = 1, ..., m \quad \text{and} \quad X = \sum_{1}^{m} X_i$$

We assume that the regulator knows or is able to estimate $D(X)$, but we relax this assumption

[5] If $c_i(x_j)$ is interpreted as the cost of abating pollution to x_j, one is implicitly assuming that there is perfect enforcement, and therefore our mechanism will maximize total welfare. If $c_i(x_j)$ is interpreted as the cost of having a standard of x_j, one is not assuming perfect enforcement, only that higher standards are better. In that case, our mechanism sets the right standard, but eschews the issue of whether they will be enforced.

in Section 5.2. Also, this definition of damages also assumes that what matters is the total level of pollution, and not its geographic distribution. Although this assumption is not essential for our mechanism to work, it can be justified on the grounds that the pollutant to be regulated is "uniformly mixed" in the sense that only the amounts emitted are relevant, and not their place of generation.

In this context, a social choice function is a function $f : \mathcal{C}^m \to \mathbf{R}_+^m$ that specifies for each possible profile of cost functions (one for each industry) the pollution level that each firm must produce. The regulator wishes to implement the social choice function that minimizes the net cost of pollution. Technically, f is a selection from the correspondence $F : \mathcal{C}^m \rightrightarrows \mathbf{R}_+^m$ defined by

$$F(c) = \arg \min_{(x_1,...,x_m)} \left[D\left(\sum n_i x_i \right) + \sum n_i c_i(x_i) \right], \tag{1}$$

for all $c = (c_1, ..., c_m) \in \mathcal{C}^m$. In order for F to be well defined, we assume that D is unbounded so that the set of minimizers is nonempty for all $c \in \mathcal{C}^m$.

4 The Mechanism and the Theorem

We now present our mechanism, and then show that it fully implements f. That is, we will show that in the unique equilibrium in undominated strategies of the game designed by the regulator, firms truthfully disclose their cost functions.

For our direct revelation mechanism, the strategy space for each firm is $\mathcal{C}$. Firms must announce their cost functions, and thereby, the cost function of the industry. For each profile of announcements $C = (C_1, ..., C_m)$, C_i will represent the profile of announcements of firms in industry i, so that

$$C = (C_1, ..., C_m) = \left(\underbrace{c_1, ..., c_{n_1}}_{\text{industry 1}}, \underbrace{c_{n_1+1}, ..., c_{n_1+n_2}}_{\text{industry 2}}, c_{n_1+n_2+1}, ..., c_{n_1+n_2+...+n_m} \right). \tag{2}$$

For each profile C let

$$x_{1j} = \min \{ f_1(c_j, c_{p_2}, ..., c_{p_m}) : p_i \in I_i, i = 2, ..., m \}.$$

The number x_{1i} is the emissions standard that would result for industry 1 if the regulator believed the announcement of firm j in this industry, and chose the announcement of each firm in each of the remaining industries which would result in the most stringent of standards for industry 1. A firm with a low x_{1j} is most likely telling the truth, since it is announcing a cost function that could result in a harsh environmental policy. Similarly, define x_{ij} for $i = 2, ..., m$ and $j \in I_i$ to be the standard that would be implemented for industry i if the regulator believed the announcement of firm j in that industry. Also, define

$$\underline{x}_i = \min_{j \in I_i} x_{ij} \qquad \text{and} \qquad \overline{x}_i = \max_{j \in I_i} x_{ij} \tag{3}$$

to be, respectively, the most (least) stringent standard announced consistent with the announcements of firms in industry i.

Our mechanism is as follows:

1. Firms announce their types

2. The regulator identifies in each industry i the firms which announced the cost functions which are consistent with $\overline{x}_i$, and samples one of these firms with equal probability. The idea is to monitor firms which are most likely lying. A firm is fined if and only if: it is sampled; its report is false; the inspection discovers (with probability ε) that the report was false. The size of the fine does not matter, it can be as small as one wants.

3. The emissions standards $(\underline{x}_1, ..., \underline{x}_m)$ are implemented.

Theorem 1. *Any efficient social choice function f defined by equation (1) is fully implementable in undominated strategies. That is, the unique equilibrium in undominated strategies of the direct revelation mechanism, is truth telling.*

Proof. We first show that truth telling is an equilibrium. Without loss of generality, consider the situation of firm 1 when all other firms in all industries are reporting the true costs $(c_1, c_2, ..., c_m)$. Notice that declaring the true c_1 leads to the implementation of the x_{1j} consistent with all the declarations of firms 2 through n_1. If firm 1 reports $\widehat{c}_1 \neq c_1$, and $x_{11} \geq x_{1j}$ for all $j = 2, ..., n_1$, there is a chance that the regulator will sample firm 1 and if (with probability ε) he finds out that the report was false, he will fine firm 1. Moreover, since declaring $\widehat{c}_1$ does not change the standard to be implemented, $x_{12} = ... = x_{1n_1}$, declaring the truth is better than declaring a $\widehat{c}_1 \neq c_1$, such that $x_{11} \geq x_{1j}$. Also, reporting the truth is better than declaring any $\widehat{c}_1 \neq c_1$ with $x_{11} < x_{1j}$, since such an x_{11} results in a smaller standard for the industry. Therefore, declaring the truth is better than declaring anything else, proving that truth telling is an equilibrium.

To show that there is no other equilibrium in undominated strategies, take any profile of announcements $C = (C_1, C_2, ..., C_m)$ (see equation (2)) in which not all firms are telling the truth and suppose it is an equilibrium. Without loss of generality, suppose that a firm in industry 1 is lying and let $\overline{x}_1$ be the largest standard for industry 1 consistent with C. First note that none of the reports consistent with $\overline{x}_1$ can be false. If there was only one firm declaring $\overline{x}_1$ and lying, it would be strictly better off declaring the same cost function as the firm declaring $\underline{x}_1$: it does not change the standard, and it reduces to at most 1/2 the probability of being sampled and fined. If there are at least two firms declaring $\overline{x}_1$, and at least one is lying, then a liar is strictly better off declaring a cost function that yields an x_{1j} only slightly smaller than $\overline{x}_1$, since it then avoids being sampled and if it reduces the standard to be implemented, it can do so by an arbitrarily small amount.

Second, notice that no firm j can declare a cost function yielding an $x_{1j} = \underline{x}_1 < \overline{x}_1$. Declaring such an x_{1j} is dominated by declaring the truth (which as we know from the previous paragraph yields $\overline{x}_1$). We conclude that all reports must yield $\overline{x}_1$, and since none of these reports can be false, the unique equilibrium in undominated strategies is truth telling. ∎

Remark. *It is worth emphasizing that our mechanism has a unique equilibrium in undominated strategies.*[6] *As has been argued in the literature on mechanism design the issue of multiplicity is very relevant, especially if the equilibria arising from the game can be Pareto ranked, and it becomes focal to lie. See Moore (1992) p. 186., fn. 5 and the references therein.*

5 Different Assumptions

In this Section we briefly discuss three variants of our assumptions and of the mechanism that still fully implement truth telling. This is relevant since the institutional settings may vary from country to country, making some versions impossible to implement, while rendering others feasible.

Before turning to the variations of the model, we note that our two main assumptions are necessary for the mechanism to fully implement truthtelling. If the regulator had no way of finding out whether the firms are lying, the following would be an equilibrium.[7] Given the potential pollution level in each industry $\overline{X}_i$, firms in industry i report a cost function c_i such that

$$D'\left(\sum n_i \overline{X}_i\right) = -c_i'\left(\overline{X}_i\right)$$

and the regulator then sets the non binding standard $\overline{X}_i$ in industry i. Also, note that even if the regulator could find out whether a report was true with probability ε, as in our model, if a firm were alone in the industry, it would maximize profits by declaring a cost function that yields $\overline{X}_i$ as its standard, provided ε and the fine are sufficiently small.

There are several variants of the mechanism that also yield truthtelling as the unique equilibrium. Here we analyze two. The first variant is concerned with our main assumption: that there are at least two firms in each industry. The second analyzes the case where damages to society are unknown, or there is no interest in determining them.

5.1 Industries with one firm.

Suppose industries 1 through k have just one firm, $n_1 = n_2 = ... = n_k = 1$, and that industries $k+1$ through $k+m$ have at least two firms, as has been our assumption so far. As before, we

[6] Kwerel (1977) and Dasgupta et al. (1980) study the issue of whether truth telling is an equilibrium. The issue of uniqueness and whether truth telling is the only equilibrium is not analyzed.

[7] For simplicity we assume that D is differentiable.

let I_i be the set of indexes of firms in industry i, even for industries with 1 firm. When firm j in industry i, pollutes x_j, define total pollution as

$$X_i = \sum_{j \in I_i} x_j, i = 1, .., k, k+1, ..., k+m \quad \text{and} \quad X = \sum_{1}^{k+m} X_i$$

Again, the regulator wishes to implement the social choice function that minimizes the net cost of pollution. Technically, he wishes to implement a selection f from the correspondence $F : \mathcal{C}^{k+m} \rightrightarrows \mathbf{R}_+^{k+m}$ defined by

$$F(c) = \arg \min_{(x_1,...,x_{k+m})} \left[D\left(\sum n_i x_i \right) + \sum n_i c_i(x_i) \right], \tag{4}$$

for all $c = (c_1, ..., c_{k+m}) \in \mathcal{C}^{k+m}$. For simplicity, in this section and the next we assume that D is strictly convex and that the set $\mathcal{C}$ is the set of bounded, convex, weakly decreasing functions $c : \mathbf{R}_+ \to \mathbf{R}_+$, and we endow $\mathcal{C}$ with the sup norm. These assumptions are just to ensure that the set of minimizers is unique for each $c \in \mathcal{C}^{k+m}$.

Suppose that the regulator can estimate, not necessarily exactly, the cost functions of industries 1 through k and call $\widehat{c}_i$ those estimates. As before, the regulator will ask firms in industries $k+1$ through $k+m$ to report their cost structures. For each profile of announcements,

$$C = (C_{k+1}, ..., C_{k+m}) = \left(\underbrace{c_{k+1}, ..., c_{k+n_{k+1}}}_{\text{industry } k+1}, ..., c_{k+n_{k+1}+n_{k+2}+...+n_{k+m}} \right)$$

let

$$\widehat{x}_1 = \min \{ f_1(\widehat{c}_1, ..., \widehat{c}_k, c_{p_1}, ..., c_{p_m}) : p_i \in I_{k+i}, i = 1, ..., m \}$$

and similarly for industries $2, ..., k$. As before, define

$$x_{k+1j} = \min \{ f_{k+1}(\widehat{c}_1, ..., \widehat{c}_k, c_{p_2}, ..., c_{p_m}) : p_i \in I_{k+i}, i = 1, ..., m \}$$

and similarly for industries $k+2$ through $k+m$. The definitions of $\overline{x}_i$ and $\underline{x}_i$ are as before, from equation (3).

Consider the following mechanism:

1. The regulator estimates a cost function $\widehat{c}_i$ for firms in industries $i = 1, ..., k$.

2. Firms in industries $k+1$ through $k+m$ announce their types

3. The regulator identifies in each industry $i = k+1, ..., k+m$ the firms which announced the cost functions which are consistent with $\overline{x}_i$, and samples one of these firms with equal probability. A firm is fined if and only if: it is sampled; its report is false; the inspection discovers (with probability ε) that the report was false. The size of the fine does not matter, it can be as small as one wants.

4. The emissions standards $\left(\widehat{x}_1, ..., \widehat{x}_k, \underline{x}_{k+1}, ..., \underline{x}_{k+m}\right)$ are implemented.

Theorem 2. *For any estimates $(\widehat{c}_1, ..., \widehat{c}_k)$ the unique equilibrium in undominated strategies of the direct revelation mechanism, is truth telling. Moreover, the standards $\left(\widehat{x}_1, ..., \widehat{x}_k, \underline{x}_{k+1}, ..., \underline{x}_{k+m}\right)$ are continuous in $(\widehat{c}_1, ..., \widehat{c}_k)$ so that if the estimated $(\widehat{c}_1, ..., \widehat{c}_k)$ are close to the truth, the standards in all industries will be close to the first best standards.*

Proof. The proof that the unique equilibrium in undominated strategies is truth telling mirrors exactly the proof of Theorem 1, and is therefore omitted.

Continuity of the standards follows from applying Berge's Maximum Theorem (see Aliprantis and Border (1999), p. 539) to $F(c)$ in equation (4): when $(c_{k+1}, ..., c_{k+m})$ are fixed in their true levels,

$$D\left(\sum n_i x_i\right) + \sum_{1}^{k} \widehat{c}_i(x_i) + \sum_{k+1}^{k+m} n_i c_i(x_i) \tag{5}$$

is a function of $(\widehat{c}_1, ..., \widehat{c}_k)$ and $x = (x_1, ..., x_{k+m})$. Then, the set $x(\widehat{c})$ of minimizers of (5) is upper hemicontinuous, and therefore continuous, as was to be shown. ∎

5.2 Unknown Damages

In this section we consider two extensions to our basic model that address the question of whether our mechanism works when either D is unknown, or irrelevant.

Suppose first that the regulator is able to estimate D. Then, as in the previous section, we have that the mechanism works, and that if the estimate of D is accurate, the emissions standards will be close to the complete information ones.

Theorem 3. *For any estimate $\widehat{D}$ the unique equilibrium in undominated strategies of the direct revelation mechanism of Section 4, is truth telling. Moreover, the standards are continuous in $\widehat{D}$ so that if the estimated $\widehat{D}$ is close to the truth, the standards in all industries will be close to the first best standards.*

The proof of Theorem 3 is similar to that of Theorem 2, and is therefore omitted.

Another extension of the model that is relevant is one in which total damages to society are irrelevant. Consider the case of a country that wants to achieve a certain level of pollution $\overline{X}$ in the most efficient way. This could be the case, for example, of countries that adopted the Kyoto Protocol: they have committed to achieving by 2012 a certain level of emissions. Europe, for instance, must abate its 1990 levels of green house gases by 8%. The problem of the regulator is therefore to find the standards for each industry that minimize the total costs of abatement, and

that achieve the desired level of emissions. Formally, suppose that the Kyoto standard is $\overline{X}$, and let

$$\Gamma\left(\overline{X}\right) = \left\{(x_1, ..., x_m) : \sum n_i x_i \leq \overline{X}\right\}.$$

Then, the regulator wants to implement a selection f from

$$F(c) = \arg \min_{(x_1,...,x_m)\in\Gamma\left(\overline{X}\right)} \sum n_i c_i(x_i).$$

We have that our mechanism still implements truth telling, and this results in the complete information standards for this problem.

Theorem 4. *For any $\overline{X}$ the unique equilibrium in undominated strategies of the direct revelation mechanism of Section 4, is truth telling.*

The proof is identical to that of Theorem 1, and is therefore omitted.

6 On the Novelty of Our Theorems

We believe that the main merit of our results is their applicability given the simplicity of the mechanism and of the proof, which makes it "likely" that players will understand their incentives.[8] In particular, we do not use some of the standard techniques, like cross reporting, used in the literature on implementation with complete information. Nevertheless, in this section we argue that our results are new, and discuss the relationship with the literature on mechanism design.

First, our results do not follow from any of the existing theorems in the literature. That is, there is no theorem that ensures that the social choice correspondence defined by equation 1, or any selection from it, is fully implementable in undominated Nash equilibrium. The results in Jackson, Palfrey and Srivastava (1994) do not apply either to our mechanism, or to the simpler version in which there is only one industry and two firms. Their Theorem 1 is for three or more firms, and their Theorem 3 requires the existence of a "worse outcome" that is not present in our setup.[9] It is also worth noting that our mechanism is bounded.[10]

[8] We thank Matt Jackson for many of the references in this Section, and for his comments regarding the importance of the simplicity of the mechanism and the proof.

[9] A worse outcome in that setting would be a standard of 0 and for each firm a lottery which yields the fine with probabilty ε. We do not need to include such an outcome in our space of allocations for our mechanism to work. Our mechanism inspects only one firm.

[10] A mechanism is bounded if any dominated strategy is dominated by an undominated strategy. In our setup, strategies that yield standards larger than the one resulting from truth telling are not dominated by any strategy. Suppose c is one such strategy, and consider another c'. If c' yields a larger standard, it can be fined when c is not, and if it yields a lower standard, it may result in a more stringent standard than c.

Only strategies that yield a standard lower than the standard consistent with the truth are dominated, and they are dominated by truth telling, which is undominated.

Second, slight variations of our mechanism fully Nash implement the social choice correspondence in equation 1 (we have removed the qualifier "in undominated strategies"). These variations of the model are relevant because it has been argued that almost any social choice function can be implemented in complete information environments in undominated strategies (see Palfrey and Srivastava (1991), Jackson (1992), Jackson, Palfrey and Srivastava (1994) and Jackson and Srivastava, 1996). In one of these variations, for example, the regulator inspects all firms unless there are declarations yielding at least two different standards and only one firm is making a declaration consistent with the least standard. In that case, he inspects all firms except the one declaring the least standard. This mechanism is also bounded and fully Nash implements any selection from F.

Third, although inspections and fines have been used in the past and it is "known" that they help in the implementation problem, our assumptions are weaker and different than the ones that have been used before. For example, the important works of Mookherjee and P'ng (1989) and Ortuño-Ortin and Roemer (1993) used costly but perfectly informative inspections and sizeable fines. Our inspections can be as uninformative as one wants, and the fines can be arbitrarily small. Arya and Glover (2005) use a public signal that may be only slightly correlated with the player's reports to implement truth telling (to the owner of a firm) by a manager and his auditor. In their model, however, fines for lying can be large.

Finally, our results are not subject to the criticisms to full implementation in complete information that have been raised by Chung and Ely (2003), since our setup is, in their terminology, one of "private values".

7 Summary

We have presented a mechanism that may help in solving the important problem of how to get polluters to tell the truth about their abatement costs. Our solution is simple, shares some features of how the actual regulatory process works in the US and other places, it implements truth telling by firms and the efficient level of pollution. Also, we have argued that one of the reasons why one does not observe in practice alternative mechanisms that have been proposed in the literature is because they were complicated and relied on taxes and subsidies, which may be too difficult to implement for regulators.

Our main assumption is that there are at least two firms in each industry. We have argued that this is a reasonable assumption, and we have shown how our mechanism can still be used even when that assumption is not satisfied.

References

[1] Aliprantis, C. and K. Border, (1999), *Infinite Dimensional Analysis*, Springer: Heidelberg.

[2] Arya, A. and J. Glover, (2005), "Maintaining Auditor Independence," mimeo.

[3] Böhm, P. and C. S. Russell, (1985), "Comparative analysis of alternative policy instruments," in *Handbook of Natural Resource and Energy Economics*, Vol. I, A.V. Kneese and J.L. Sweeney. (Eds.), Elsevier.

[4] Chung, K. and J. Ely, (2003), "Implementation with Near Complete Information," *Econometrica*, 71, 3, 857-871.

[5] Dasgupta, P., P. Hammond and E. Maskin, (1980), "On Imperfect Information and Optimal Pollution Control," *Review of Economic Studies*, **47** (5), 857-60.

[6] EPA (1992), "Model Pollution Prevention Plan for the Kraft Segment of the Pulp and Paper Industry," NTIS # PB93-145605.

[7] EPA (2002), "Development Document for Final Effluent Limitations Guidelines and Standards for the Iron and Steel Manufacturing Point Source Category," EPA-821-R-02-004, at http://www.epa.gov/ost/ironsteel/tdd.htm

[8] Field, B. C., (1997), *Environmental Economics, an Introduction,* Irwin/McGraw-Hill: Boston.

[9] Jackson, M. O. (1992): "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," Review of Economic Studies, 59, 757–775.

[10] Jackson, M. O., T. R. Palfrey, and S. Srivastava (1994): "Undominated Nash Implementation in Bounded Mechanisms," Games and Economic Behavior, 6, 474–501.

[11] Jackson, M. O., and S. Srivastava (1996): "Characterizations of Game Theoretic Solution Concepts which Lead to Impossibility Theorems," Review of Economic Studies, 63(1), 23–38.

[12] Keohane, N. O., R. L. Revesz and R. N. Stavins, (1998), "The Choice of Regulatory Instruments in Environmental Policy," *Harvard Environmental Law Review*, 22 (2), 313-367.

[13] Kwerel, E. (1977), "To Tell the Truth: Imperfect Information and Optimal Pollution Control," *Review of Economic Studies*, **44** (3), 595-601.

[14] Lewis, T. (1996), "Protecting the Environment when Costs and Benefits are Privately Known," *RAND Journal of Economics*, **27**, 819-47.

[15] Mookherjee, D. and I. Png, (1989), "Optimal Auditing, Insurance, and Redistribution," *the Quarterly Journal of Economics,* Vol. 104, 2, pp. 399-415.

[16] Moore, J. (1992). Implementation, contracts, and renegotiation in environments with complete information. In *Advances in Economic Theory: Sixth World Congress*, vol. I, edited by J.-J. Laffont. Cambridge, UK: Cambridge University Press.

[17] Ortuño-Ortin, I. and J. Roemer (1993), "Politically realistic implementation with inspection: The equity-honesty-welfare trade-off," *Economics and Politics* 5, 255-270.

[18] Palfrey, T. R., and S. Srivastava (1991): "Nash Implementation Using Undominated Strategies," *Econometirca*, 59, 479–501.

[19] Russell, C. S. and P. T. Powell, (1996), "Choosing Environmental Policy Tools, Theoretical Cautions and Practical Considerations," IADB, Washington D.C., - No. ENV-102.