Collective penalties and inducement of self-reporting

Katrin Millock, David Zilberman

Centre d'Economie de la Sorbonne
UMR 8174

Cahiers de la MSE

Collective Penalties and Inducement of Self-Reporting

Katrin MILLOCK
David ZILBERMAN

2006.48

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://mse.univ-paris1.fr/Publicat.htm
ISSN : 1624-0340

Collective Penalties and Inducement of Self-Reporting

Katrin Millock[1] and David Zilberman[2]

[1]Centre d'Economie de la Sorbonne, CNRS
Maison des Sciences Economiques
106/112 Bd. de l'Hôpital
75 647 Paris Cedex 13
France

Email: millock@univ-paris1.fr

[2] Department of Agricultural and Resource Economics
University of California, Berkeley
207 Giannini Hall
Berkeley CA 94 720-3310
USA

Email: zilber@are.berkeley.edu

COLLECTIVE PENALTIES AND INDUCEMENT OF SELF-REPORTING

Abstract

Random accidents can be contained by collective penalties. These penalties are not likely to be enforced but rather induce self-reporting that enhances welfare due to early containment. Self-reporting under collective penalties increases overall welfare, but may increase expected environmental cost. Even when regulation is constrained by an upper limit on the acceptable collective penalty, the threat of collective penalties can induce an incentive-compatible mutual insurance scheme under which a side-payment is made to the agent that self-reports an accident. This self-reporting mechanism is welfare-improving, but first-best outcomes can only be obtained when the collective penalty is unconstrained, or when an honor system applies. In cases when there is a new externality that requires fast response (avian flu), collective penalties can compliment or substitute for monitoring.

Résumé : Pénalités collectives – incitations à l'auto-déclaration

Nous étudions le rôle des pénalités collectives pour donner des incitations à déclarer un accident environnemental, dans un modèle d'agents hétérogènes et dans lequel les inspections sont irréalisables. Une déclaration immédiate d'un accident permet au régulateur de prendre des mesures pour minimiser le coût social de l'accident (en effet, sans une auto-déclaration, des problèmes d'information créent des délais dans la réparation du dommage). Nous démontrons l'impact distributif d'une politique d'auto-déclaration des dommages, ce qui pourrait expliquer la réticence des associations environnementales envers une telle politique. Ensuite nous analysons le cas où il existe une limite supérieure sur le montant de la pénalité collective acceptable et nous montrons comment une politique de pénalités collectives peut inciter à une action collective pour récompenser un agent qui fait une auto-déclaration de son accident. Le mécanisme que nous proposons fonctionne comme une assurance mutuelle afin de récompenser un agent qui auto-déclare un accident. Nous évaluons le bien-être social sous l'action collective par rapport à un statu quo avec pénalités individuelles, et nous proposons des conditions qui entraînent une amélioration de bien-être. Néanmoins, les choix optimaux d'effort de prévention nécessitent soit une pénalité collective sans contrainte, soit un code d'honneur.

Keywords: ambient tax, collective penalties, enforcement, self-reporting

JEL: H23, K42, Q28

COLLECTIVE PENALTIES AND INDUCEMENT OF SELF-REPORTING

1. Introduction

Many negative externalities pose a regulatory challenge because the regulator cannot identify the individual source of the damage. The agents are often members of a group, a particular industry, or a regional cluster. This is the essence of the nonpoint source pollution problem, which is basically a moral hazard problem since the regulator cannot monitor individual actions. Second-best regulations have been implemented to deal with the problem, such as input taxes or regulation based on observable technology or some other proxy variable related to the creation of pollution. Based on Holmstrom's (1982) work on optimal incentives for teams, Meran and Schwalbe (1987) and Segerson (1988) suggested that collective penalties can be used as a first-best solution. The collective penalty has been termed the ambient tax when based upon overall ambient quality. It works by imposing the full social marginal damage cost on each polluter, who thus faces incentives to cut back pollution to the socially optimal level.

The particular moral hazard problem that we study here is the problem of stochastic accidents when the regulator cannot trace the damage to its source among a large population of potential polluters. Examples include spills of hazardous chemicals, leakage from oil fields into groundwater, the release of a microbe or the propagation of a virus that spread a disease, or forest fires. These examples are all cases of externalities that are difficult to trace to their origin, but where quick detection will reduce the social damage. In some instances, public monitoring systems can be built up to contain the damage. The U. S. Customs' control of imported fruit and vegetables in order to prevent exotic pests from infesting agriculture is one example. But monitoring systems take time to establish, and may be very costly, or simply not feasible. In such cases, self-reporting of accidents would enable early containment and thus a lower social damage than if the regulator at a later time were to discover the accident. The problem is how to induce self-reporting when it is optimal to do so. Here we study the role that collective penalties can play in a policy to induce self-reporting.

Collective punishments exist in many different contexts. Markets sometimes exercise collective punishments and rewards (Winfree and McCluskey, 2005). When consumers cannot distinguish between many different producers, they have to pay a regional price premium for French wine, Italian design, and punish Romanian wines. If a producer wants to distinguish herself, she has to invest in building a brand. Externality control, which is what we address in this paper, is about building market-like solutions: collective punishment and rewards exist in the market to overcome information deficiencies. Collective punishment is used to slow the spread of diseases. In May 2003, the discovery of a single reported case of Bovine Spongiform Encephalopathy (BSE) led to an immediate ban on exports of beef from the United States to Japan and Canada. The European Union imposed a worldwide ban on beef exports from the United Kingdom in order to limit the effects of its outbreak of BSE. Similar export bans were imposed as a means to limit the propagation of foot-and-mouth disease. Japan and China will not import produce from California if an outbreak of the Mediterranean fruitfly has been reported in the state.

Collective punishment has furthermore been used to end bad practices causing environmental or behavioral externalities. The US Food and Drug Administration banned

Chilean fruit after cyanide was found in two grapes in 1989. The most famous case is probably the US 1990-91 boycott of Mexican tuna to stop the use of net threatening dolphins. Collective environmental punishments may also be used as a trade barrier. Indeed, the GATT dispute settlement panel eventually ruled in favor of Mexico. There are several other noted examples of use of collective penalties to achieve environmental ends. In the Kesterson debacle, the U. S. government threatened to cut federal water project supply to farmers in the California San Joaquin Valley if the ramifications of the drainage problem had not been addressed. Under the Everglades Forever Act in Florida, a land tax is automatically increased every year for all farmers in the adjacent area if the aggregate phosphorus reduction goal is not met for the wetlands (Ribaudo and Caswell, 1999). There are various incidents of close-down of industrial activities when smog levels exceed a certain level. The ban on South African exports to protest apartheid can also be seen as a collective punishment to counteract a behavioral externality.

All these examples of collective penalties demonstrate that they can act as a credible threat. In this paper we will argue that the main advantage of a policy based on collective penalties, such as the ambient tax, is to induce collective action for early containment and reduction of the damage imposed on society. Damage from pollution and health risks can be assessed by using a risk-generating function which describes the stages of contamination, transfer and fate, and exposure and gives a dose-response function (Lichtenberg and Zilberman, 1988). The key to minimizing damage is early intervention. Since risk is a dynamic process (a plume of chemicals moves, a virus propagates), there are time lags, and the earlier the regulator can intervene in the transfer-and-fate process, the smaller will be the damage. We do not develop a dynamic model, but the essence of the problem is that early action brings about a lower social damage cost. In cases when there is a new externality or a source of damage that requires fast response and the regulator has limited capacity to monitor individual agents, collective punishment may be used or introduced as a threat. The avian flu illustrates the importance of early warning and self-reporting in order to reduce the costs of a human flu pandemic (Fouchier et al., 2005). A regional outbreak may not be reported, and even if the regulator knows that there has been an outbreak in a certain region, she does not always know the source. It is important to require poultry farmers to report an outbreak early and to refrain them from selling the infected chicken.

In suggesting that a practical application of collective penalties can be to induce self-reporting by polluters, the current article is closely related to the literature on self-reporting and the work on law enforcement spawned by the seminal articles by Becker (1968) and Stigler (1970). Existing studies of self-reporting (Malik, 1993; Kaplow and Shavell, 1994; Innes, 1999a, 1999b, 2000) rely on the background threat of a public monitoring system that imposes a penalty on the agent with some probability.[1] If the penalty upon self-reporting is set at a lower level than the standard penalty, self-reporting can be induced. However, random monitoring systems are not always feasible or can be prohibitively costly to establish. When the outbreak of a disease occurs, its origin is not easy to establish. Random monitoring may also be impractical when dealing with infrequent accidents, for example, when toxic releases lead to water contamination, where the pollution dissipates quickly and it may be very costly to obtain information that enables the regulator to trace the pollution back to its source.

---

[1] Some other articles on environmental law enforcement also discuss self-reporting but as part of an exogenous government policy (Harford, 1987; Livernois and McKenna, 1999).

In this article we investigate self-reporting in the case when the regulator cannot invest in random monitoring. As in Innes (2000), the model allows for heterogeneous agents. We show how a collective penalty can work as a threat and induce agents to self-report, first in a simple model with no constraint on the collective penalty, then in a model encompassing an upper limit on the feasible collective fine. In this situation, we analyze an incentive-compatible mutual insurance scheme under which a side-payment is made to the agent that self-reports an accident, and identify the conditions under which it is welfare-improving.

Levinson's (2003) discussion of legal cases contends that one of the merits of collective penalties is "creating incentives for the wrongdoer himself to confess." He also argues that regulated groups (e.g., firms) have better information than regulators. Avoiding collective sanctions can be a cause of group solidarity to overcome transaction costs and develop mechanisms to reduce the burden. The mutual insurance scheme presented here is such a mechanism.[2]

The outline of the paper is the following. In section 2, we present the model and derive the first-best benchmark. In section 3, we first show how collective penalties can be used to induce self-reporting, and we study the distributional impact of self-reporting on private surplus and on the environment. Next, we introduce limits on the level of collective penalties and determine the incentives for self-reporting in this constrained situation. Section 4 then shows how, in this constrained situation, collective penalties

[2] Economists have also recognized the leverage opportunities from group delegation and have studied the use of peer monitoring to alleviate the regulator's incomplete information, in particular with respect to credit networks (see, for example, Varian, 1990; Arnott and Stiglitz, 1991; Armendariz, 1999; Ghatak 1999), and work teams (Holmstrom and Milgrom, 1990). The idea of solidary networks of group members is not new; already the Medieval Guilds of merchants introduced informal rules of conduct that relied upon group sanctions (Greif, Milgrom, and Weingast, 1994).

nevertheless can induce the group of agents that is regulated to set up a collective compensation system for self-reporting. The welfare effects of such collective action are analyzed. We conclude with a discussion of the role of collective penalties and honor systems.

2. The Model

We model a large population of *N* risk neutral agents, for which the regulator cannot trace any accidents that may happen. The accident can take the form of a release of a pernicious microbe, or a hazardous waste spill.[3] The probability that agent *i* causes an accident is g_i. The probabilities are independent and it is furthermore assumed that the probability of an accident is small so that only one accident occurs at a given time. The probability of an accident is a decreasing function of the unobserved level of care agent *i* uses, X_i, with $\frac{\partial g_i}{\partial X_i} < 0, \frac{\partial^2 g_i}{\partial X_i^2} > 0$. The regulator cannot infer the level of care chosen by the firm from firm characteristics such as technology. The chosen level of care, *X*, is assumed to be contained in $\left[0, \overline{X}\right]$. By definition, g_i is in $\left[0,1\right]$ and, depending on the hazard problem, g_i is bounded above by some small number $\overline{g}$ where $g_i\left(X=0\right)=\overline{g}$, with $\frac{\partial g_i(0)}{\partial X_i}$ arbitrarily large and $\frac{\partial g_i\left(\overline{X}\right)}{\partial X_i}$ arbitrarily small.

---

[3] Strand (1994) also studies regulation of stochastic accidents where the damage from an accident is fixed, the firm's output is exogenous and the regulator can observe the fixed initial investment in technology. The solution proposed by Strand involves the regulator subsidizing the firm to take the first-best level of care, also when technology is variable. As noted by Strand (1994), practical implementation of such a policy is difficult when regulators are budget-constrained.

Agent i's profit before any regulatory penalty is denoted $\pi_i(X_i)$. We assume that devoting time and resources to accident prevention has an opportunity cost on private profits, such that:

$$\frac{\partial \pi_i}{\partial X_i} < 0, \frac{\partial^2 \pi_i}{\partial X_i^2} \leq 0,$$

by standard assumptions of a convex cost function. Furthermore, $\pi_i(X=0)=\bar{\pi}$, and $\pi_i(\bar{X})=0$.

Kaplow and Shavell (1994) analyzed two fundamental advantages of self-reporting: monitoring cost savings for the regulator and a reduction in the cost of risk-bearing. An additional advantage of self-reporting (noted in Kaplow and Shavell, 1994, and modeled formally in Innes, 1999a) is the reduction in the damage cost when agents can undertake remediation activities. This is the essential advantage of self-reporting in our model.

The social cost of an accident that is not reported is C_H. However, if an agent takes early action, or reports the accident promptly so that containment measures can be implemented by the regulator as early as possible before damage spreads, the resulting damage would be $C_L < C_H$. We call this early containment since we wish to distinguish this case from when it is the regulator that remediates the damage upon gaining knowledge of the accident. An agent's containment actions are assumed to be site-specific and visible (as in Hutchinson and van't Veld, 2005). Examples include containing the diffusion of a plume of toxic chemicals or the movement of an infectious disease such as foot and mouth disease in cattle, or limiting the spread of a forest fire by action while it is in the agent's territory. Let m_i represent the cost to agent i of early containment and self-reporting. We thus link the two decisions of early containment and self-reporting as one. If an agent has an accident and takes some action to contain it, he will also self-report the accident, since he can show that action has been taken to minimize damage. On the other hand, an agent that self-reports always has to make some extra effort to contain the problem once deciding to self-report.[4] The larger part of the agent's cost is related to early containment, whereas the actual self-reporting cost is likely to be quite low (the equivalent of a phone call in some cases). When $C_L + m_i < C_H$, early containment and self-reporting improves social welfare. We assume m_i is contained in an interval defined as $\underline{m} < m_i < \overline{m}$. While we know that $C_L + \underline{m} < C_H$, we cannot exclude the case that $C_L + \overline{m} > C_H$, and the model will allow for both possible cases.

2.1 The first-best benchmark

We can now define the socially optimal level of care under complete information. First, define the optimal level of care with or without early containment:

$$X_i^C = \arg\max \; \pi_i(X_i) - \left(C_L + m_i\right) g_i(X_i) \tag{1}$$

$$X_i^N = \arg\max \; \pi_i(X_i) - C_H \; g_i(X_i) \tag{2}$$

[4] Tort law requires some effort. The most intuitive example being a hit-and-run car accident; someone who self-reports an accident but takes no effort to mitigate the impact, seriously prejudices his position with the regulator.

where X_i^C is the level of care taken when early containment and self-reporting costs are accounted for, and X_i^N is the level of care taken with no containment. Under either case each agent should balance its marginal cost of care with the marginal social benefit resulting from accident prevention.

Depending on the social cost of an accident, we can have a corner solution or an interior solution. When both $\pi_i(X_i^N) - C_H\, g_i(X_i^N) < 0$ and $\pi_i(X_i^C) - \left(C_L + m_i\right) g_i(X_i^C) < 0$, a corner solution is optimal, i.e. close-down of the firm. We will assume that an interior solution always exists. For such an interior solution, in the unlikely case of two possible equilibria, the profit-maximizing level of care with early containment is where $\frac{\partial^2 \pi_i(X_i)}{\partial X_i^2} - \left(C_L + m_i\right)\frac{\partial^2 g_i(X_i)}{\partial X_i^2} \le 0$, and the marginal profit function intersects the marginal expected cost function from above. A similar condition holds for the profit-maximizing level of care without early containment.

The social optimum can now be defined as the solution to:

$$\underset{\gamma_i}{Max} \sum_{i=1}^{N} \gamma_i \left\{ \pi_i(X_i^C) - \left(C_L + m_i\right) g_i(X_i^C) \right\} + \sum_{i=1}^{N} \left(1 - \gamma_i\right) \left\{ \pi_i(X_i^N) - C_H\, g_i(X_i^N) \right\} \qquad (3)$$

where γ_i defines whether it is socially optimal to contain and self-report environmental damage from source *i*.

The optimal value of γ_i is:

$$\gamma_i = 1 \text{ when } \quad \pi_i(X_i^C) - \left(C_L + m_i\right) g_i(X_i^C) \ge \pi_i(X_i^N) - C_H\, g_i(X_i^N) \qquad (4a)$$

$$\gamma_i = 0 \text{ when } \pi_i(X_i^C) - \left(C_L + m_i\right) g_i(X_i^C) < \pi_i(X_i^N) - C_H\, g_i(X_i^N).$$

For the subsequent analysis we will define the index $\hat{i}$ for which an agent would be indifferent between self-reporting and not. If agents are ordered according to the cost of containment, m_i, where agent i=1 has the lowest cost and agent i=N has the highest cost of containment, there will be at most one agent, defined by $\hat{i}$, for which the following condition holds:

$$\pi_i(X_i^C) - \left(C_L + m_i\right) g_i(X_i^C) = \pi_i(X_i^N) - C_H\, g_i(X_i^N) \tag{4b}$$

For agents with $i \leq \hat{i}$, it will be socially optimal to contain environmental damage, whereas for agents with $i > \hat{i}$, it will not be optimal to take early action.

3. Self-Reporting and Early Damage Detection

The occurrence of an accident, however, is private information before the regulator detects it. All the regulator can do is to impose a fixed penalty on each agent in the population when the regulator detects the damage. In this Section, we will show that when there are no constraints on the level of the collective penalty, it can be used to induce self-reporting and lead to the socially optimal outcome.

Given this policy, the decision choice of agent i, facing a collective penalty P is:

$$\underset{X_i}{Max} \qquad \pi_i(X_i) - P \sum_{i=1}^{N} g_i(X_i).$$

The agent will choose a level of care characterized by the following first-order condition:

$$\frac{\partial \pi_i(X_i)}{\partial X_i} - P \frac{\partial g_i(X_i)}{\partial X_i} = 0 \qquad \forall i. \tag{5}$$

As usual, the level of care increases with the level of the penalty:

$$\frac{dX_i}{dP} = \frac{\frac{\partial g_i(X_i)}{\partial X_i}}{\left[\frac{\partial^2 \pi_i(X_i)}{\partial X_i^2} - P \frac{\partial^2 g_i(X_i)}{\partial X_i^2}\right]} > 0$$

If the regulator does not consider self-reporting as an option, he will set the collective penalty $P = C_H$ in order to internalize the damage, given that the social cost of an accident that is not reported is C_H. In the extreme case with a very high level of damage cost without early containment this can lead to a situation where the policy forces all firms to close down. If firms still operate, welfare is defined as

$$W^0 = \sum_{i=1}^{N} \pi_i\left(X_i^N\right) - \sum_{i=1}^{N} C_H g_i\left(X_i^N\right).$$ [5]

[5] We assume a small industry with perfectly elastic demand, so the model excludes any effects on consumer surplus. In order to keep a simple model, we also do not include a marginal cost of public funds and any increase in government revenue is here a pure transfer.

Now, allow self-reporting and early clean-up. Detecting and containing an accident early due to self-reporting reduces social damages from C_H to C_L. On the other hand, self-reporting is costly to the agent since the nature of a dose-response function requires a self-reporting agent to incur some extra effort to contain the damage. This cost is represented by m_i in our model. A self-reporting policy would thus offer a reduced penalty upon self-reporting, defined as $F_S < P$. Given that an accident has occurred, an agent *i* will self-report iff

$$\left(F_S + m_i\right) < P \tag{6}$$

If the condition in Equation (6) holds for all agents, then complete self-reporting will occur. Depending on the cost of containment and self-reporting, the condition may not hold for some agents and those agents will not self-report an accident.

Introducing self-reporting now changes the care levels of those agents that choose to self-report in case of an accident, since the level of care is increasing with the penalty. From the definition in Equation (2), the level of care taken under the collective penalty P=C_H equals X_i^N. The level of care taken under self-reporting is defined as $X_i^S = \arg\max \pi_i(X_i) - \left(F_S + m_i\right) g_i(X_i)$.

Agents that choose the strategy to self-report will take a lower level of care than under the collective penalty ($X_i^N - X_i^S$> 0), since $F_S + m_i < P$. Consequently, the risk of having an accident increases for agents that self-report: $g_i\left(X_i^S\right) > g_i\left(X_i^N\right)$. However, because there will be early containment, the social cost of damages will be smaller if $C_L + m_i < C_H$.

The regulator's problem is to choose P and F_S to induce the optimal level of self-reporting:

$$\underset{F_S, P, \delta_i}{Max} \sum_{i=1}^{N} \delta_i^S \left\{ \pi_i(X_i(F_S)) - \left(C_L + m_i\right) g_i(X_i(F_S)) \right\} + \sum_{i=1}^{N} \left(1 - \delta_i^S\right) \left\{ \pi_i(X_i(P)) - C_H g_i(X_i(P)) \right\} \tag{7}$$

where δ_i^S is an indicator variable; δ_i^S =1 if agent *i* self-reports an accident, δ_i^S =0 if not.

The solution to this problem leads to the following Proposition:

Proposition 1: *A collective penalty $P = C_H$ with a reduced fine for self-reporting $F_S = C_L$ will induce self-reporting of stochastic accidents. Under this policy the socially optimal level of care can be induced and a welfare gain is obtained compared to the benchmark policy of collective penalties only.*

Proof: In Appendix.

The proof in the Appendix shows that the levels of care induced by this policy correspond to the socially optimal levels of care: X_i^N and $X_i^S = X_i^C$. With information about C_H and C_L, the regulator can thus induce all agents below the previously defined critical level $\hat{i}$ to self-report, and agents with $i > \hat{i}$ not to self-report. Depending on the costs of self-reporting, different patterns of self-reporting can occur. If $\hat{i} \geq N$, all agents should self-report. If $\hat{i} < 1$, then no self-reporting is optimal. Whenever the critical agent

$\hat{i}$ is in the interval $[1, N]$, some intermediate level of self-reporting is optimal. An alternative expression for welfare under the optimal policy is thus:

$$W^* = \sum_{i=1}^{\hat{i}} \left\{ \pi_i(X_i^S) - \left(C_L + m_i\right) g_i(X_i^S) \right\} + \sum_{i=\hat{i}}^{N} \left\{ \pi_i(X_i^N) - C_H\, g_i(X_i^N) \right\}. \qquad (8)$$

To sum up, the literature has previously shown that self-reporting can be induced with a policy of random penalties (Kaplow and Shavell, 1994; Innes, 1999a, 1999b, 2000). We study a different benchmark policy here: the case when the regulator cannot invest in a monitoring and detection system. In this case, we have shown that self-reporting can be induced using fixed certain penalties. The self-reporting policy works because of the threat of the collective fine imposed on all firms when an accident is discovered. The role of the collective penalty is thus double; on the one hand, it induces higher levels of care by itself. On the other hand, it also serves to increase the incentives for agents to self-report an accident.

3.1. The distributional effects of self-reporting

The inducement of self-reporting by a collective penalty has a distributional impact. Since there is no change for agents that do not self-report, it is sufficient to study the change in net welfare resulting from the agents that choose to self-report, i.e., the first term in Equation (8). The introduction of incentives for self-reporting will result in a social welfare improvement as well as an increase in private surplus. Since we consider two policy options where the externalities are fully internalized, the social welfare gain from allowing self-reporting is identical to the improvement in the net private surplus after penalties. The net social gain from self-reporting when there is a collective penalty is thus:

$$\Delta W = W^* - W^0 = \sum_{i=1}^{\hat{i}} \left\{ \underbrace{\left[C_H\, g_i(X_i^N) - \left(C_L + m_i\right) g_i(X_i^S) \right]}_{\text{environmental impact}} + \underbrace{\left[\pi_i(X_i^S) - \pi_i(X_i^N) \right]}_{\text{market impact}} \right\} \qquad (9)$$

The net social gain from a self-reporting policy is made up of the environmental impact and the increase in the private surplus in the final good market – the market impact. The environmental impact of self-reporting can be decomposed further into two terms:

$$\Delta W = \sum_{i=1}^{\hat{i}} \left\{ g_i(X_i^N)\left(C_H - C_L - m_i\right) - \left[g_i(X_i^S) - g_i(X_i^N) \right]\left(C_L + m_i\right) + \left[\pi_i(X_i^S) - \pi_i(X_i^N) \right] \right\} (10)$$

The first term is the gain from early action. However, self-reporting also induces a lower level of care than under a policy of fixed penalties: $X_i^N - X_i^S > 0$. The impact of this is two-fold; on the one hand, it increases the private surplus from production (the third term), a benefit that can be added to the gain from lower damage due to early action. On the other hand, it simultaneously increases the risk of an accident (the second term), and this loss has to be deducted from the welfare gain. The optimal level of care obtained under self-reporting thus implies a trade-off between the gain from lower damage when an accident occurs and the impact from two effects resulting from a lower level of care than under a fixed penalty. By a revealed preference argument the private surplus is always larger under self-reporting if it is voluntary. As seen above, the environmental impact of self-reporting is ambiguous, however. It is defined as the sum of the environmental cost savings following early containment (net of containment and self-reporting costs) when an accident occurs and the increase in the expected cost of an accident due to the increase in the probability of having an accident. By using the elasticity of the accident probability function with respect to the level of care, ε_{gx} = $-\left(\frac{\partial g_i(X_i^S)}{\partial X_i^S}\frac{X_i^S}{g_i(X_i^S)}\right)$, we can derive a condition that determines when the environmental impact is positive.

Proposition 2a: *The private surplus always increases under self-reporting, but the introduction of self-reporting may increase the expected environmental damage.*

Proposition 2b:

If $\frac{C_H}{C_L + m_i} > \frac{g_i(X_i^S)}{g_i(X_i^N)}$*, or* $\varepsilon_{gx} < \left(\frac{C_H - (C_L + m_i)}{C_H}\right)\left(\frac{X_i^S}{X_i^N - X_i^S}\right) \quad \forall i$ *, the environmental impact from self-reporting is positive.*

Proof: In Appendix.

Although the net social gain from self-reporting is positive, Proposition 2 shows that such a policy can encounter resistance from environmental groups. As shown above, the gain from self-reporting comes from the increase in private surplus following the lower level of care under a self-reporting policy. The necessary condition for a net reduction in damages requires that the elasticity of the accident probability function is smaller than an amount determined largely by the percentage reduction in environmental damage costs from self-reporting if the accident probability were fixed. If the accident probability function is relatively inelastic with respect to the level of care, then the direct effect on damage reduction will outweigh the increase in the probability of an accident. This may not always be the case, though. If some groups consider that environmental damage reduction should carry more weight in social decision-making than private surplus, they are likely to focus on the increased environmental risk induced by a self-reporting policy. The above analysis thus suggests why environmental groups may oppose a self-reporting policy.

Up to now, we assumed that there were no upper limits on the penalties imposed by industry participation constraints. In the following, we will introduce the possibility of an upper bound on the fixed penalty (Section 3.2).

3.2 Restricted collective penalties

Now suppose that, because of political constraints, there exists an upper limit K to the fixed penalty that the regulator can implement. Environmental policy in particular is constrained by industrial profitability considerations. For example, effluent standards as defined in the U.S. Clean Water Act should be developed on the basis of economic achievability.[6] Also in Europe, permits for industrial facilities are often based on an analysis of best available technology and its costs (OECD, 1999). Hence, the parameter K

reflects political considerations and commitment to achieving environmental goals. The direct consequence of a low K is to restrain the regulator's choice of the collective penalty, and preclude the optimal solution from being reached.

There are three possible cases for the interval of K: a) $K>C_H$, b) $C_L+m_i<K<C_H, \forall$ i , and c) $K < F_S = C_L$. If a) holds, the constraint is irrelevant. If b) holds, the regulator can still induce self-reporting and early damage containment by setting the fixed penalty equal to K and the fine under self-reporting equal to C_L, and the optimal level of self-reporting will be obtained. Given the condition on K, each individual agent will still have incentives to choose self-reporting. The interesting case that we will focus on here is case c) $K < F_S = C_L$.

Under the constrained maximum penalty, the firm's level of care will be:

$$X_i^K = \arg max\ \pi_i\left(X_i\right) - K\, g_i\left(X_i\right) \tag{11}$$

Given that $K<F_S$, it is clear that the level of care X_i^K will be lower than the level of care under the optimal self-reporting fine (F_S), X_i^S . In fact, it is sub-optimal from society's perspective. Welfare under a collective penalty set at K equals:

$$W^K = \sum_{i=1}^{N} \pi_i(X_i^K) - \sum_{i=1}^{N} C_H g_i(X_i^K).$$

Would a policy of collective penalties, although restricted, still give firms incentives to self-report accidents? With a constrained penalty, agents can be induced to self-report as

[6] 33 U.S. Code 1311 Clean Water Act Sec 301 Effluent limitations.

long as the limit on the collective penalty exceeds the environmental damage cost when there is early action plus the individual cost of self-reporting. However, if the constraint is binding ($K < C_L + m_i$), $\forall\, i$, agents will have no individual incentives to self-report an accident unless the penalty upon doing so is set lower than or equal to $K\text{-}m_i$.

Since each agent pays for the occurrence of an accident, the group of agents could gain $NK – (C_L + m_i)$ in total if the agent causing an accident voluntarily reported it.[7] In Section 4 below, we will study how collective action could occur to organize self-reporting in the case when the level of the fixed collective penalty is not enough to induce the individual agent to report voluntarily.[8]

4. Collective Action and Self-Reporting Under a Constrained Collective Penalty

When $C_L + m_i > K$, no agent will report an accident out of own self-interest. The only situation under which self-reporting would occur would be if the agent were to receive compensation for this from the other agents. In particularly homogenous communities, solidarity develops over time and is sustained by social norms. Community enforcement has been shown to be able to sustain cooperation in a repeated game context (Kandori, 1992) or in a setting where there is strong peer pressure to deter free-riders (Kandel and

---

[7] In this model, it is not possible for the agents to exploit the asymmetric information between them and the regulator and maximize the gain from collective self-reporting by having the agent with the minimum cost report the accident. The reason is that, once an accident is reported, measures to contain it necessarily need to be site-specific and visible to the regulator.

[8] Under the constraint on the maximum collective penalty, the regulator might be tempted to introduce incentives for self-reporting of the following sort: if an agent self-reports, he would only pay a penalty $K\text{-}m_i$. Under such a policy, the social damage costs would amount to $C_L + m_i$, and the other agents would be exempted from paying the collective penalty K. Nevertheless, a firm that chooses the strategy to self-report will still choose a sub-optimal level of care, since its effective penalty is $K\text{-}m_i$. It is easy to show that the decentralized self-reporting scheme proposed in Section 4 dominates a self-reporting policy of this kind that is organized by the regulator under the constrained penalty K.

Lazear, 1992). In close-knit communities, agents can have explicit or implicit agreements to reimburse an agent that self-reports and pays the damage costs of an accident.[9] The objective of this Section is to investigate whether such a collective side-payment system is feasible and examine its effects on social welfare.

4.1. A mutual insurance scheme

We will analyze a mutual insurance scheme under which agents agree on a symmetric side-payment S_i to the agent who has an accident and reports it. We assume that the risk functions of all agents are known and that there are no transaction costs connected with the side-payments. There are several examples where this can be the case, in particular, it applies to industrial clusters or industry sector associations, where firms already interact and have means of communication. In the context of risk neutral agents that we analyze here mutual insurance does not arise because of risk aversion, but rather because the agent wishes to insure herself against the reduction in profits resulting from the imposition of the collective penalty (which as we have seen can be as high as to close down operations).

Under the mutual insurance scheme considered here, in the case where agent i has an accident, he would receive a payment S_i by the other agents, and whenever another agent has an accident, agent i has to be willing to pay his share $(1/N-1)$ of the

[9] It is natural to assume that the regulator can impose the full damage cost on the responsible agent once an agent self-reports an accident. When no responsibility can be assigned, the collective penalty is constrained by a maximum limit K.

compensation for self-reporting. The following incentive-compatibility constraint would have to hold for an agent to voluntarily participate in and abide by such a scheme:

$$\pi_i(X_i^{M_i}) - g_i(X_i^{M_i})\left(C_L + m_i - S_i\right) - \sum_{j\neq i} g_j(X_j^{M_j})\frac{S_j}{N-1} \geq \pi_i(X_i^K) - \sum_{i=1}^{N} g_i(X_i^K)K \qquad \forall\, i \tag{12}$$

where X_i^M and X_i^K are defined by:

$$X_i^M = \arg\max\left\{\pi_i(X_i) - (C_L + m_i - S_i)g_i(X_i)\right\}$$
$$X_i^K = \arg\max\left\{\pi_i(X_i) - K\, g_i(X_i)\right\}$$

A simple rule for the compensation payments by the other agents is that they will reimburse the self-reporting agent for his additional costs above the collective penalty: C_L+ m_i-*K*. The subsidy is thus defined as $S_i = C_L + m_i - K$. With a side-payment $S_i = C_L + m_i - K$, it follows directly from Equation (12) that the sufficient and necessary condition for full participation is:

$$\sum_{j\neq i} g_j(X_j^K)\left[K - \frac{C_L + m_j - K}{N-1}\right] \geq 0 \qquad \forall\, i \quad . \tag{13}$$

The expected payment $\sum_{j\neq i} g_j(X_j^K)\frac{C_L + m_j - K}{N-1}$ can be interpreted as a form of insurance premium that is paid in order to receive compensation by other agents in case of an accident. When all agents have identical costs of self-reporting, *m,* condition (13) simplifies to $K \geq \frac{C_L + m}{N}$. Hence, if the collective fine is sufficiently high, or the self-reporting and containment costs are sufficiently low, agents can organize a collective compensation system for self-reporting. In the case where agents have different costs of self-reporting, there will exist a minimum subsidy such that all agents are induced to participate, this minimum subsidy being defined by the agent with the highest costs of self-reporting:

$$\max_{1 \leq i \leq N} m_i = NK - C_L .$$

Proposition 3a: *When the liability limit of the collective penalty is sufficiently high, or the costs of self-reporting are sufficiently low, a collective reporting system can be agreed upon among all agents if* $NK - C_L \geq \max_{1 \leq i \leq N} m_i$.

The preceding results did not depend on the self-reporting costs being identical for every agent. Some homogeneity in self-reporting costs is necessary on the other hand. If there exists an agent *i* with self-reporting and containment costs that exceed those of the other agents by a large amount, $m_i >> m_j$, then $S_i = 0$ but $S_j = m_j \quad \forall\, j \neq i$. The side-payment scheme would no longer be symmetric, and the agent with high containment costs will not self-report, although he will pay his share of the compensation to other self-reporting agents.

Proposition 3b: *If an agent i has significantly higher self-reporting costs than the other* *agents, such that* $g_i(X_i^M)\left(C_L + m_i - K\right) \geq \frac{1}{N-2}\sum_{k \neq j \neq i} g_k(X_k^{M_k})\left(C_L + m_k - K\right)$, *he will be excluded from the scheme, and the side-payment will equal* $S_i = 0, \quad S_j = m_j \quad \forall\ j \neq i$.

Proof: In Appendix.

The mutual insurance scheme suggested in this Section is designed so that agents have no incentives to renege on the mutual agreement and not pay their share of the costs once an accident has been reported. The scheme is quite information-demanding, however, since it requires agents to know not only the self-reporting costs of every single agent but also the accident probabilities of each other.

One case where agents have an informational advantage over the regulator and also may know reasonably well each others' costs and accident probabilities is the case of industrial zones. In India, much of the chemical industry is clustered in industrial estates. Although aggregate pollution is noticeable, it is very difficult for the state regulatory agency to monitor individual plants. Firms that are members of the industrial association of the estate have a significant advantage in monitoring efficiency and the provision of incentives for mutual monitoring and information exchange. Kathuria and Sterner (2002) analyze efforts to regulate chemical plants in the Ankleshwar Industrial Estate in the Indian state of Gujarat and argue that the local industrial association is much better informed than the state environmental agency, which has difficulties controlling the many small firms within the estate. Firms in the industrial association have a common interest in avoiding accidents and pollution incidents in order to maintain or enhance the reputation of the industrial estate. Existing interactions on the product market or similarities in the production process among the firms in the estate make it likely that firms have better information on each other than the regulator.

The capacity of collective penalties to induce self-reporting and an associated compensation system relies on the existing cooperation among agents. In the industrial sector, membership of professional associations serves the purpose of creating a collective. In the area of agriculture and the environment, water and drainage districts provide a forum for agents to create collective responsibility. In the United States there is an established legal tradition for self-taxing organizations to manage environmental problems. Drainage districts can be established with a qualified majority vote of property owners for the construction of canals, drains, and levees. Such drainage districts have the authority to inspect all land and levy tax payments on agents within their boundaries, and could work as an institution for decentralized regulation under collective penalties. Formal institutions such as drainage and water districts or informal social interaction that provides clear sanction and punishment rules and a sense of fairness among agents can serve as a foundation for effective collective action (Baland and Platteau, 1993, 1996). Informal cooperation plays a particularly important role in developing countries and in rural communities (Bardhan, 1993).

4.2 The welfare effects of collective action for early containment and self-reporting

Thus far we have not drawn any conclusions on the welfare effect of a full coalition between agents, but only studied the necessary and sufficient condition for agents to participate. The next step is to ask whether such collective action would lead to the socially optimal level of care. Welfare under this arrangement can be written as follows (with δ_i^M =1 when agent *i* self-reports, and δ_i^M =0 if not):

$$W^M = \sum_{i=1}^{N} \delta_i^M \left[\pi_i(X_i^M) - \left(C_L + m_i\right) g_i(X_i^M) \right] + \sum_{i=1}^{N} \left(1 - \delta_i^M\right) \left[\pi_i(X_i^K) - C_H \, g_i(X_i^K) \right]$$

$$\text{s.t.} \quad \begin{aligned} X_i^M &= \arg\max \left\{ \pi_i(X_i) - (C_L + m_i - S_i) g_i(X_i) \right\} \\ X_i^K &= \arg\max \left\{ \pi_i(X_i) - K \, g_i(X_i) \right\} \end{aligned}$$

When Equation (13) holds for every agent, there is full participation in the mutual insurance scheme ($\delta_i^M = 1 \quad \forall i$), and a comparison of welfare under the constrained collective penalty without self-reporting implies that welfare improves iff:

$$\Delta W = W^M - W^K = \sum_{i=1}^{N} \left[\pi_i(X_i^M) - \pi(X_i^K) - \left(C_L + m_i\right) g_i(X_i^M) + C_H \, g_i(X_i^K) \right] > 0 \qquad (14)$$

With a subsidy to the self-reporting agent equal to $S_i = C_L + m_i - K$, each agent will choose a level of care identical to the level under the constrained collective penalty (X_i^K), and the above expression is positive iff

$$\sum_{i=1}^{N} \left\{ g_i(X_i^K) \left(C_H - C_L - m_i\right) \right\} > 0 \,. \qquad (15)$$

The term $(C_H - C_L)g_i(X_i^K)$ represents the expected gain from early containment under self-reporting, whereas the expected cost of self-reporting is equal to $m_i g_i(X_i^K)$. As known, a collective reporting system that induces self-reporting will increase welfare if the gain from early containment and self-reporting exceeds its cost: $C_H - C_L > m_i \quad \forall i$. With heterogeneous agents, a sufficient condition for a welfare improvement is that $C_H - C_L - \max_{1 \le i \le N} m_i > 0$.

A comparison of the condition for full participation (Equation 13) with the condition for a welfare improvement (Equation 15) indicates when collective action for self-reporting improves welfare compared to the benchmark collective penalty.

Proposition 4:

(i) *When* $\frac{C_L + m_i}{N} < K \le \frac{C_H}{N} \quad \forall i$, *mutual insurance organized under the threat of a collective penalty K will occur and will improve welfare (* $\Delta W = W^M - W^K \ge 0$ *). Within this range, increases in K improve welfare.*

(ii) *If* $K > \frac{C_H}{N}$ *an inefficiently high level of self-reporting may occur under mutual insurance.*

(iii) *If* $K < \frac{C_L + m_i}{N} \quad \forall i$, *no self-reporting will occur under the collective penalty.*

Proof: In Appendix.

The proposition leads to a counter-intuitive result that there may be situations when the collective penalty is too low to induce the first-best outcome and yet sufficiently high to induce excessive self-reporting under mutual insurance. Full participation in a mutual insurance scheme may lead to situations where individuals with high cost of early containment will self-report. Thus, at least theoretically there may be situations where an increase in K may lead to a reduction in welfare. This may occur if the benefit of increased safety is smaller than the cost of excessive self-reporting.

4.3. Discussion and perspectives

We have shown that in cases where the overall damage cost with self-reporting is not greater than without self-reporting, $C_H > C_L + \max_{1 \le i \le N} m_i$, and if the collective penalty is greater than $C_L + \max_{1 \le i \le N} m_i$, the collective penalty will provide incentives for self-reporting that will result in the first-best outcome and will prevent the payment of the draconian collective penalty. If political economy reasons lead to a collective penalty that is smaller than $C_L + \max_{1 \le i \le N} m_i$, the collective penalty will be paid if it is small and below $\min_{1 \le i \le N} \frac{C_L + m_i}{N}$, and producers will assume a relatively low level of care. If the constrained penalty is sufficiently high, above $\max_{1 \le i \le N} \frac{C_L + m_i}{N}$ (yet below $C_L + \max_{1 \le i \le N} m_i$), there is a potential for collective action that will lead to self-reporting and avoidance of the payment of the collective penalty. We introduced a mutual insurance scheme that will lead to such collective action. It assumes transaction costs are zero or at least so low as to not exhaust the gains from cooperation. The scheme also assumed that agents have full information on each others' containment costs and some degree of social capital that will allow collaboration. Several features of industrial organization and community interaction make us believe that the application of the schemes suggested here does have some interest as an alternative means of regulation whenever government monitoring is prohibitively costly. Industries tend to locate in clusters or at least interact in industrial associations which permit some information exchange which would make agents have better information about each other than the regulator. In the context of environmental policy, agricultural settings with several sources located in the same geographical neighborhood constitute another area of application of the mechanism suggested here.

The outcome of collective action when the collective penalty is constrained and $C_H > C_L + \max_{1 \le i \le N} m_i$, is more efficient than the outcome without cooperation when the collective penalty is paid. The collective action does not improve the level of care, but reduces the overall expected damage cost. Because of the low level of care, due to the constrained collective penalty, the outcome of self-reporting under collective action is inefficient, though. Increases in the collective penalty K will increase the overall efficiency of the collective action outcome as it will lead to firms taking a higher level of care. With a constraint on the collective penalty the mechanism of mutual insurance will thus lead to a welfare improvement but not to the first-best outcome. One mechanism that would lead to the first-best outcome is a well-functioning honor system. With an honor system individuals will assume responsibility for their action, self-report and pay $C_L + m_i$. That will lead to the optimal level of care defined in Equation (1): X_i^C. Honor systems are frequently associated with collective penalties. Economists may be too pessimistic about the occurrence of honor systems, as some evidence discussed by Dubner and Levitt (2004) show. Our analysis provides a case where society would benefit from the emergence of honor code systems since this is the only system that will result in a first-best outcome under a case of constrained collective penalties. For other cases, we have shown that mutual insurance schemes may have agents self-report accidents and, although second-best, this policy will improve welfare compared to the constrained initial situation.

The desired performance of self-reporting is not obtained in cases where some agents have a high cost of early containment and self-reporting, so that $C_H \leq C_L + m_i$. In particular, in the case when $C_H \leq C_L + m_i \leq NK$, self-reporting by these agents is suboptimal. This case is quite unlikely, though, since it would imply that the costs of self-reporting and early containment are so high that they eliminate the damage cost reduction from early action. It serves as a reminder, though, that collective action is no panacea; in some cases where the collective penalty is constrained, collective action may not be socially optimal and may lead to excessive self-reporting.

5. Conclusions

The paper has shown that random accidents can be contained by a policy of collective penalties. The penalties are not likely to be enforced but rather induce self-reporting that enhances welfare due to early containment. The success of this policy depends on political commitment to establish sufficiently high penalties. When the penalty is equal to or greater than the damage caused by an accident, the collective penalty will lead to the first-best outcome. Penalties below the damage cost may lead to the first-best outcome only under an honor system. Otherwise, systems of mutual insurance may lead to self-reporting, but with higher probability of accidents than socially optimal. The magnitude of the social benefits induced by self-reporting largely depends on the cost saving because of early containment. The self-reporting also reduces the harsh impact that implementation of collective penalties may have on producers. The harshness of collective penalties may prevent one from using it regularly, but in some situations it may be used for a transitional period until routine monitoring is introduced. Outbreaks of disease, such as the avian flu, illustrate the importance of early warning when timely reporting and response reduce damage costs. In cases of routine pollution, when detection of due care may be easy, monitoring remains the appropriate policy.

The comparison between collective penalties and investment in monitoring depends on the efficiency of the monitoring technology and the urgency with which a response must be obtained. In cases where there is a new externality or a source of damage that requires a fast response, or when routine monitoring of pollution carries a low probability of detection, and the regulator has limited capacity to monitor all agents, collective penalties may be used or introduced as a threat. Building monitoring systems takes time and can be extremely expensive, so collective penalties can complement or be substitutes for monitoring.

Future research could extend the model in several directions. The analysis considered the case when the damage of the accident is known with certainty, and the model could be extended to analyze the case when the damage is random. Even under random damages, imposing collective penalties equal to the realized damages will induce self-reporting if the expected damage does not increase with self-reporting. The analysis could also be expanded to situations where the identity of the agent that caused the accident may be known with some probability to the regulator or other members of the industry. Collective penalties seem to be effective in these cases, and the incentives to self-report may actually increase. It may also be useful to consider the performance of systems that combine random monitoring and self-reporting with collective penalties, especially in situations when the capacity to utilize these two approaches is constrained technically or politically.

References

Armendariz de Aghion, Beatriz. "On the Design of a Credit Agreement with Peer Monitoring." *Journal of Development Economics* 60 (October 1999), 79-104.

Arnott, Richard, and Stiglitz, Joseph E. "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out or Peer Monitoring?" *American Economic Review* 81 (March 1991), 179-190.

Baland, Jean-Marie, and Platteau, Jean-Philippe. "A Multi-Disciplinary Analysis of Local-Level Management of Environmental Resources." Rome, Italy: Food and Agricultural Organization of the United Nations, 1993.

Baland, Jean-Marie, and Platteau, Jean-Philippe. "Halting Degradation of Natural Resources. Is There a Role for Rural Communities?" Rome, Italy: Food and Agricultural Organization of the United Nations, 1996.

Bardhan, Pranab. "Analytics of the Institutions of Informal Cooperation in Rural Development." *World Development* 24 (1993), 633-639.

Becker, Gary S. "Crime and Punishment: An Economic Approach." *Journal of Political Economy 76* (March/April 1968*)*, 169-217.

Dubner, Stephen J., and Levitt, Steven D. "What the Bagel Man Saw." *The New York Times Magazine*, June 6, 2004.

Fouchier, Ron, Kuiken, Thijs, Rimmelzwaan, Guus, and Albert Osterhaus. "Global task force for influenza." Commentary, *Nature* 435 (26 May 2005), 419-420.

Ghatak, Maitreesh. "Group Lending, Local Information and Peer Selection." *Journal of Development Economics* 60 (October 1999), 27-50.

Greif, Avner, Milgrom, Paul, and Barry R. Weingast. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102 (August 1994), 745-776.

Harford, Jon D. "Self-reporting of Pollution and the Firm's Behavior under Imperfectly Enforceable Regulations." *Journal of Environmental Economics and Management* 14 (September 1987), 293-303.

Holmstrom, Bengt. "Moral Hazard in Teams." *Bell Journal of Economics* 13 (1982), 324-340.

Holmstrom, Bengt, and Milgrom, Paul. “Regulating Trade Among Agents.” *Journal of Institutional and Theoretical Economics* 146 (March 1990), 85-105.

Hutchinson, Emma and van’t Veld, Klaas. “Extended Liability for Environmental Accidents: What You See Is What You Get.” *Journal of Environmental Economics and Management* 49 (2005), 157-173.

Innes, Robert. “Remediation and Self-Reporting in Optimal Law Enforcement.” *Journal of Public Economics* 72 (June 1999a), 379-393.

Innes, Robert. “Self-Policing and Optimal Law Enforcement When Violator Remediation Is Valuable.” *Journal of Political Economy* 107 (1999b), 1305-1325.

Innes, Robert. “Self-reporting in Optimal Law Enforcement when Violators have Heterogeneous Probabilities of Apprehension.” *Journal of Legal Studies* 29 (January 2000), 287-300.

Kandel, Eugene, and Lazear, Edward P. “Peer Pressure and Partnership.” *Journal of Political Economy* 100 (August 1992), 801-817.

Kandori, Michihiro. “Social Norms and Community Enforcement.” *The Review of Economic Studies* 59 (January 1992), 63-80.

Kaplow, Louis, and Shavell, Steven. “Optimal Law Enforcement with Self-Reporting of Behavior.” *Journal of Political Economy* 102 (June 1994), 583-606.

Kathuria, Vinish, and Sterner, Thomas. “Monitoring and Enforcement: Is Two-Tier Regulation Robust?” Discussion Paper 02-17. Washington D.C.: Resources for the Future, 2002.

Levinson, Daryl J. “Collective Sanctions.” *Stanford Law Review* 56 (2003), 345-428.

Lichtenberg, Erik, and Zilberman, David. “Efficient Regulation of Environmental Health Risks.” *Quarterly Journal of Economics* 49 (February 1988), 167-178.

Livernois, John, and McKenna, C.J. “Truth or Consequences: Enforcing Pollution Standards with Self-Reporting.” *Journal of Public Economics* 71 (March 1999), 415-440.

Malik, Arun S. “Self-Reporting and the Design of Policies for Regulating Stochastic Pollution.” *Journal of Environmental Economics and Management* 24 (May 1993), 241-257.

Meran, Georg, and Schwalbe, Ulrich. “Pollution Control and Collective Penalties.” *Journal of Institutional and Theoretical Economics* 143 (December 1987), 616-629.

Organization for Economic Cooperation and Development. *Environmental requirements for industrial permitting. Volume 3. Regulatory approaches in OECD countries*. Paris: OECD, 1999.

Ribaudo, Marc, and Caswell, Margriet. “Environmental Regulation in Agriculture and the Adoption of Environmental Technology.” In *Flexible Incentives for the Adoption of Environmental Technologies in Agriculture*. Frank Casey, Andrew Schmitz, Scott Swinton, and David Zilberman, eds., pp. 7-25. Dordrecht: Kluwer Academic Publishers, 1999.

Segerson, Kathleen. “Uncertainty and Incentives for Non-Point Pollution Control.” *Journal of Environmental Economics and Management* 15 (1988), 87-98.

Stigler, George J. “The Optimum Enforcement of Laws.” *Journal of Political Economy* 78 (May/June1970), 526-536.

Strand, Jon. “Environmental Accidents under Moral Hazard and Limited Firm Liability.” *Environmental and Resource Economics* 4 (October 1994), 495-509.

Varian, Hal R. “Monitoring Agents with Other Agents.” *Journal of Institutional and Theoretical Economics* 146 (1990), 153-174.

Winfree, Jason A., and McCluskey, Jill J. “Collective Reputation and Quality.” *American Journal of Agricultural Economics* 87 (February 2005), 206-214.

APPENDIX

Proof of Proposition 1:

We know that $P = C_H$ yields the level of care X_i^N (defined in Equation 2).

The conditions that determine whether an agent self-reports or not are:

$$\delta_i^S = 1 \ \ if \ \ \pi_i(X_i^S) - \pi_i(X_i^N) - m_i g_i(X_i^S) + C_H\, g_i(X_i^N) - F_S\, g_i(X_i^S) \geq 0 \qquad \text{(A1)}$$

$$\delta_i^S = 0 \ if \ \pi_i(X_i^S) - \pi_i(X_i^N) - m_i g_i(X_i^S) + C_H\, g_i(X_i^N) - F_S\, g_i(X_i^S) < 0 \qquad \text{(A2)}$$

With $P = C_H$ and $F_S = C_L$, individual agents' choice of self-reporting will coincide with the socially optimal choice of self-reporting as defined in Equations (4a). Thus, when the regulator knows the damage levels C_H and C_L, $\delta_i^S = \gamma_i$, and the first-best choices are implemented: X_i^N, $X_i^S = X_i^C$. Q.E.D.

Proof of Proposition 2:

Private surplus:

If agent i self-reports then

$$Max \ \left[\pi_i(X_i) - C_H \sum_{i=1}^{N} g_i(X_i)\right] < Max \ \left[\pi_i(X_i) - (C_L + m_i) g_i(X_i) - \sum_{j \neq SR} C_H g_j(X_j)\right]$$

by a revealed preference argument.

It follows directly from $\pi_i(X_i)$ decreasing and concave in X_i, and $X_i^N - X_i^S > 0$, that $\pi_i(X_i^S) - \pi_i(X_i^N) > 0$.

The environmental impact:

Define the elasticity of the accident probability function (evaluated at X_i^S) as

$$\varepsilon_{gX} = -\frac{\partial g_i(X_i^S)}{\partial X_i} \frac{X_i^S}{g_i(X_i^S)}.$$

Use the fact that $g_i(X_i^N) = (X_i^N - X_i^S)\frac{\partial g_i(X_i^{'})}{\partial X_i} + g_i(X_i^S) \geq (X_i^N - X_i^S)\frac{\partial g_i(X_i^S)}{\partial X_i} + g_i(X_i^S)$

(A3)

for $X_i^S < X_i^{'} < X_i^N$.

Hence, a sufficient condition for a positive environmental impact is:

$$C_H\left[(X_i^N - X_i^S)\frac{\partial g_i(X_i^S)}{\partial X_i} + g_i(X_i^S)\right] - \left(C_L + m_i\right) g_i(X_i^S) > 0. \qquad \text{(A4)}$$

Rearranging terms and multiplying by $\mathrm{X}_\mathrm{i}^\mathrm{S}$ gives

$$\frac{\partial g_i(X_i^S)}{\partial X_i}\frac{X_i^S}{g_i(X_i^S)} > \frac{\left(C_L + m_i\right) - C_H}{C_H}\frac{X_i^S}{X_i^N - X_i^S} \quad . \qquad \text{(A5)}$$

Rewriting using the definition of the elasticity of the accident probability function gives:

$$\varepsilon_{gX} < \frac{C_H - \left(C_L + m_i\right)}{C_H}\frac{X_i^S}{X_i^N - X_i^S}. \qquad \text{Q.E.D.}$$

<u>Proof of Proposition 3b</u>:

For all agents $j \neq i$ to want to exclude agent *i* the following condition has to hold:

$$\pi_j(X_j^{M_j}) - g_j(X_j^{M_j})\left(C_L + m_j - S_j\right) - \sum_{k \neq j \neq i} g_k(X_k^{M_k})\left(\frac{S_k}{N-2}\right) \geq$$
$$\pi_j(X_j^{M_j}) - g_j(X_j^{M_j})\left(C_L + m_j - S_j\right) - \sum_{k \neq j \neq i} g_k(X_k^{M_k})\left(\frac{S_k}{N-1}\right) - g_i(X_i^M)\left(\frac{S_i}{N-1}\right) \geq$$
$$\pi_i(X_i^K) - \sum_{i=1}^{N} g_i(X_i^K) K$$

$$\Leftrightarrow \quad \sum_{k \neq j \neq i} g_k(X_k^{M_k})\left[-\frac{S_k}{N-2}\right] + g_i(X_i^M)S_i \geq 0 \tag{A6}$$

For a given *N* and *j*, when $m_i \to \infty$, the above expression $\to \infty$.

Condition (A6) that indicates when a group of agents want to exclude an agent with significantly higher self-reporting costs can be rewritten as a condition on the heterogeneity of self-reporting costs:

$$g_i(X_i^M)\left(C_L + m_i - K\right) \geq \frac{1}{N-2}\sum_{k \neq j \neq i} g_k(X_k^{M_k})\left(C_L + m_k - K\right). \qquad \text{Q.E.D}$$

<u>Proof of Proposition 4</u>:

(iii):

Under collective action with side-payment S_i, $X_i^M = X_i^K$, and the agents will organize self-reporting if and only if $NK \geq C_L + m_i$

(i) and (ii):

If $\sum_{j \neq i} g_j(X_j^K)\left[K - \frac{\left(C_L + m_j - K\right)}{N-1}\right] \geq 0 \quad \forall i$ and $NK < C_H$, then

$$\sum_{j \neq i} g_j(X_j^K)\left[C_H - C_L - m_j\right] > 0 \qquad \forall i \text{, and}$$

$$\sum_{i=1}^{N} g_i(X_i^K)\left[C_H - C_L - m_i\right] > 0. \qquad \text{Q.E.D.}$$