NBER WORKING PAPERS SERIES

OPTIMAL LAW ENFORCEMENT WITH SELF-REPORTING OF BEHAVIOR

Louis Kaplow

Steven Shavell

Working Paper No. 3822

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1991

Both of Harvard University and the National Bureau of Economic Research. We are grateful to A. Mitchell Polinsky for helpful comments, the John M. Olin Foundation for financial support, and Brian Timmons for research assistance. This paper is part of NBER's research program in Law and Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3822
August 1991

OPTIMAL LAW ENFORCEMENT WITH SELF-REPORTING OF BEHAVIOR

ABSTRACT

Self-reporting -- the reporting by parties of their own behavior to an enforcement authority -- is a commonly observed aspect of law enforcement, as in the context of environmental and safety regulation. We add self-reporting to the model of the control of harmful externalities through probabilistic law enforcement. Optimal self-reporting schemes are characterized and are shown to offer two advantages over schemes without self-reporting: enforcement resources are saved because individuals who are led to report harmful acts need not be identified; risk is reduced because individuals bear certain sanctions when they report their behavior, rather than face uncertain sanctions.

Louis Kaplow
Harvard Law School
Griswold 402
Harvard University
Cambridge, MA 02138
and NBER

Steven Shavell
Harvard Law School
Langdell 357
Harvard University
Cambridge, MA 02138
and NBER

1. Introduction

A commonly observed feature of law enforcement is what we shall call self-reporting of behavior -- the reporting by parties of their own harm-producing actions to an enforcement authority. For example, firms frequently report on their behavior in the context of environmental and safety regulation, individuals involved in accidents causing injury to others often report this to the police, and even those who commit crimes sometimes confess their acts to the authorities. Presumably, parties voluntarily report on their behavior because they fear more severe treatment if they do not.[1]

What are the social advantages of self-reporting that may help to explain its use in law enforcement? More broadly, how does self-reporting fit in the theory of the control of harmful externalities? The literature on controlling externalities, dating from Pigou (1918), suggests that activities that create harm be taxed, but does not emphasize the costs of identifying parties who cause harm. The more recent literature on law enforcement, however, investigates the control of harmful activities when it is costly to identify the parties responsible for causing harm. This literature begins with Becker (1968), who stresses that, because of these enforcement costs, it is not socially advantageous to identify those who cause harm all the time, but rather to do so only with a probability (and to raise the level of sanctions accordingly).

In this article, we add self-reporting to the model of probabilistic law enforcement.[2] Under a scheme with self-reporting, individuals can be induced

[1] To illustrate, under CERCLA, 42 U.S.C. §9603(b), failure to report the release of hazardous substances may result in fines or imprisonment, apart from any penalty associated with the release itself. Also, under the pending federal criminal sentencing guidelines for organizations, reporting of violations would lead to substantial reductions in the fines that otherwise would apply.

[2] Self-reporting has not previously been considered in the literature on law enforcement. However, a number of articles in the mechanism design literature addressed to risk-sharing contracts, tax collection, regulation, and the principal-agent model are relevant to our analysis because they examine the costly auditing of reports. See, in particular, Border and Sobel (1987),

to report their harmful acts without materially affecting their incentives whether or not to commit the acts. This can be accomplished by allowing a person who reports committing a harmful act to pay a sanction equal to (or slightly less than) the expected sanction he would face if he did not report his act. Then, he will be led to report his act, but his willingness to commit the act will be unchanged. As a consequence, enforcement schemes with self-reporting offer society two advantages. First, enforcement resources are saved; because those who commit harmful acts are induced to report their behavior, enforcement effort need not be spent identifying them. Second, risk-bearing costs are eliminated (a benefit when actors are risk-averse), for those who commit harmful acts report their behavior and pay a certain amount.[3] By contrast, under law enforcement systems without self-reporting, those who commit harmful acts bear the risk of sanctions.

Now let us describe the organization of the article. In section 2, we analyze a model of self-reporting in which risk-neutral individuals choose whether to commit a single type of harmful act, and in section 3 we examine a generalization of the model allowing for many types of harmful acts. In these models, we characterize optimal enforcement schemes with self-reporting and show how they differ from schemes that do not allow self-reporting. In particular, we demonstrate that self-reporting schemes are superior to schemes without self-reporting because the former allow enforcement costs to be reduced. We also show that a positive level of enforcement is always desirable with self-reporting, even when it is not necessarily desirable without self-reporting. In section 4, we consider extensions of our analysis, allowing for risk aversion, imprisonment as a sanction, error in examination of behavior, and administrative costs of self-reporting. In section 5, we offer concluding remarks.

Mookherjee and Png (1989), and Wagenhofer (1987). See also Baiman and Demski (1980), Baron and Besanko (1984), Dye (1986), Mookherjee and Png (1990), Reinganum and Wilde (1985), Scotchmer (1987), and Townsend (1979). We comment on this literature in note 21.

[3] A related advantage is that schemes with self-reporting reduce the need to impose imprisonment, as we discuss in subsection 4B.

2. The One-Act Model

Risk-neutral individuals choose whether or not to commit an act that causes a harm h. If an individual commits the harmful act, he obtains a benefit $b \in [0, \infty)$; b differs among individuals and has positive continuous density $f(\cdot)$ with cumulative distribution $F(\cdot)$.[4] The size of the population is normalized to one.

We now define and analyze the two schemes of enforcement: enforcement without self-reporting and enforcement with self-reporting.

A. Enforcement Without Self-Reporting

In the scheme without self-reporting, the social authority examines individuals with probability p. An examination determines with certainty whether an individual committed the harmful act, and each examination costs c.[5] Individuals found to have committed the act pay a monetary sanction s, which is assumed to be socially costless to impose. The maximum level of the sanction is $\bar{s}$, where $\bar{s} \geq h$; $\bar{s}$ may be interpreted as an individual's wealth.[6] The social authority chooses the probability of examination and the sanction to maximize social welfare, defined as the sum of individuals' benefits minus the harm due to their acts and examination costs. Socially optimal values of variables will be denoted by a "*".

An individual will commit the harmful act if and only if $b \geq ps$, so that social welfare is

[4] The assumption that b has positive density on $[0, \infty)$ rules out the possibility that it is desirable to deter all individuals from committing the harmful act. If, however, b were distributed on $[0, \bar{b}]$ and $h > \bar{b}$, then it would be desirable for no one to commit the harmful act. In this case, however, complete deterrence may not be optimal (because of the high enforcement costs that would be required), in which event our analysis and results would not change.

[5] Other methods of enforcement are discussed in section 5.

[6] The assumption that $\bar{s} \geq h$ is used to rule out the corner solution in which the optimal probability equals one. As explained in note 17, our results do not depend on this assumption.

$$(2.1) \qquad W = \int_{ps}^{\infty} (b - h)f(b)db - pc.$$

The first term is the benefits minus the harm from commission of the act. The second term is the enforcement cost, as the entire population (which, recall, is normalized to one) is examined with probability p and each examination costs c.

The optimal s must be $\bar{s}$ if $p^* > 0$.[7] Were $s^* < \bar{s}$, s could be raised and p lowered such that ps remained constant. Then the first term in (2.1) would be unchanged but enforcement costs, pc, would fall; welfare would thus be higher, contradicting the optimality of s^*. (This is the argument of Becker (1968).)

To determine p^*, differentiate (2.1) with respect to p, using $s^* = \bar{s}$, to obtain

$$(2.2) \qquad \frac{dW}{dp} = \bar{s}(h - p\bar{s})f(p\bar{s}) - c.$$

This expression will be negative for all $p \in [0,1]$ if c is sufficiently large, so $p^* = 0$ is possible. However, $p^* = 1$ is not possible because the assumption that $\bar{s} \geq h$ implies that (2.2) is negative at $p = 1$. An interior solution for p^* must satisfy the first-order condition that $dW/dp = 0$.[8] In this case, the optimal probability is

$$(2.3) \qquad p^* = \frac{h - c/\bar{s}f(p\bar{s})}{\bar{s}},$$

and the optimal expected sanction is

$$(2.4) \qquad p^*\bar{s} = h - c/\bar{s}f(p\bar{s}).$$

The left side of (2.4) is the social loss from deterring the marginal individual, because he would have obtained a benefit of $p^*\bar{s}$ had he committed the act. The right side is the net social gain from deterring the marginal individual, the harm avoided minus the enforcement cost of deterring him.

[7] If $p^* = 0$, s^* clearly can be taken to equal $\bar{s}$.

[8] Here and below, we do not discuss the possibility of multiple optima, as this does not affect our analysis.

In summary, we have

Proposition 1: When there is no self-reporting:

a. The optimal probability of examination p^* may be zero.

b. If p^* is positive, it is given by equation (2.3), and the optimal sanction is the maximum feasible sanction $\bar{s}$.

B. Enforcement With Self-Reporting

In the scheme with self-reporting, individuals have the option of admitting that they committed the harmful act. If an individual states that he did so, he pays an ex ante sanction r, where $r \leq \bar{s}$, and he is not examined.[9] If an individual does not report that he committed the act, he is treated as he was in the scheme without self-reporting: he is examined with probability p and, if the examination reveals that he committed the act, he pays an ex post sanction s.

Individuals who do not commit the harmful act clearly will not report having done so. Individuals who do commit the act will report this if and only if $r \leq ps$.[10] Hence, individuals commit the act if and only if $b \geq \min(r, ps)$. There are thus two cases. If $r > ps$, individuals who commit the act do not report this, and welfare is as given in (2.1). That is, enforcement without self-reporting is a special case of enforcement with self-reporting. If $r \leq ps$, individuals who commit the act report this, and social welfare is

$$(2.5) \qquad W = \int_{r}^{\infty} (b - h)f(b)db - pcF(r).$$

Expressions (2.5) and (2.1) differ in two respects. First, the lower limit of integration in (2.5) is r rather than ps, because individuals who commit the act report this and bear the certain sanction r rather than the expected sanction ps. Second, the examination cost in (2.5) is pcF(r) rather than pc,

[9] The sanction r is called an ex ante sanction because it is paid before an individual might be examined. However, in the sequence of events that we describe, r is paid after an individual commits the act (although we could as well imagine r to be paid before an individual commits the act).

[10] As is the convention, we assume that when individuals are indifferent between reporting the truth and not -- when $r = ps$ -- they tell the truth.

because only individuals who do not commit the act (and thus do not report committing it) -- those with benefits less than r -- are examined.

We can now make precise the argument sketched in the introduction that enforcement with self-reporting can induce the same behavior as enforcement without self-reporting but at lower cost. Let $p > 0$ and s apply without self-reporting. With self-reporting, use the same p and s and set $r = ps$. Then it is apparent that the same individuals commit the act with self-reporting as without self-reporting, so the integrals in (2.1) and (2.5) are equal. But enforcement costs are lower with self-reporting by $(1 - F(ps))pc$, because those who commit the act and report this are not examined. Thus, we have

Proposition 2: Given any enforcement scheme (involving p > 0) without self-reporting, there exists a scheme with self-reporting under which behavior is the same but enforcement costs are lower.

The comparison made in this proposition understates the advantage of the optimal self-reporting scheme over the optimal scheme without self-reporting, because the optimal probabilities under the two schemes generally differ.

We now characterize the optimal enforcement scheme with self-reporting. First, the optimum will involve $r = ps$. If $r > ps$, individuals who commit the act would not report this, which proposition 2 implies cannot be optimal. If $r < ps$, p could be lowered slightly, maintaining the inequality. Then individuals who commit the act would continue to report this and pay r, so the integral in (2.5) would not change, but the reduction in p would reduce the second term, increasing welfare.[11] Second, the optimal ex post sanction is $\bar{s}$; as in the case without self-reporting, this sanction economizes on enforcement resources.

To find the optimum, we now may substitute $p\bar{s}$ for r in (2.5) and differentiate with respect to p, to obtain

[11] We have implicitly assumed that $p^* > 0$ in making this argument. But if $p^* = 0$ (a possibility that we rule out below), r can be taken to equal zero (which is p^*s) since welfare will be independent of r.

(2.6) $$\frac{dW}{dp} = \bar{s}(h - p\bar{s})f(p\bar{s}) - pc\bar{s}f(p\bar{s}) - cF(p\bar{s}).$$

At $p = 0$, the derivative equals $\bar{s}hf(0)$, which is positive, so that p^* must be positive. In contrast, $p^* = 0$ was possible without self-reporting if examination costs were sufficiently high. The reason for the difference is that, without self-reporting, the entire population needs to be examined, so that the marginal cost of increasing p is c. With self-reporting, only those individuals who do not report having committed the act need to be examined, but there are no such people when $p = 0$ ($F(r) = 0$ since $r = ps = 0$), so the marginal enforcement cost at that point is zero. The possibility that $p^* = 1$ is again ruled out by the assumption that $\bar{s} \geq h$. Since an interior solution obtains, p^* satisfies $dW/dp = 0$, which implies that

(2.7) $$p^* = \frac{h - cF(p\bar{s})/\bar{s}f(p\bar{s})}{\bar{s} + c},$$

or

(2.8) $$p^*\bar{s} = h - p^*c - cF(p^*\bar{s})/\bar{s}f(p^*\bar{s}).$$

Equation (2.8) is analogous to (2.4). The left side is the social loss from deterring the marginal individual (for his benefit from the act is r^*, which equals $p^*\bar{s}$ at the optimum). The right side is the net social gain from deterring the marginal individual, the harm avoided minus the enforcement cost of deterring him. The latter has two components in this case: p^*c, the expected cost of examining the marginal individual who, because he has been deterred, joins the pool of those who do not commit the act and thus might be examined; $cF(p^*\bar{s})/\bar{s}f(p^*\bar{s})$, the inframarginal cost of examining with a higher probability those who do not commit the act.

We can interpret r^*, which equals $p^*\bar{s}$ in (2.8), as the optimal Pigouvian tax for committing the harmful act, because this is the amount individuals pay with certainty when they commit the act (as all who commit the act are induced to report this and pay r^*). The optimal tax is less than the harm -- the externality is not fully internalized -- due to enforcement costs.[12]

[12] However, the optimal Pigouvian tax may exceed the harm for some acts in the n-act model of section 3.

Let us summarize.

Proposition 3: When there is self-reporting:

a. *In the optimal scheme, all individuals who commit the harmful act report having acted and no individuals who do not commit the act report having acted.*

b. *It is optimal to expend enforcement resources to deter some individuals from committing the harmful act -- the optimal probability of examination is positive.*

c. *The optimal probability p^* is given by equation (2.7), the optimal ex post sanction is the maximum feasible sanction $\bar{s}$, and the optimal ex ante sanction r^* equals $p^*\bar{s}$.*

We may now conclude that the optimal self-reporting scheme is superior to that without self-reporting. Proposition 2 establishes that welfare is higher with self-reporting than without it for any common positive probability of enforcement. If $p^* - 0$ without self-reporting, which is possible, welfare is obviously equivalent for $p - 0$ with self-reporting, but $p - 0$ is not optimal with self-reporting by proposition 3b. Hence, we have *Proposition 4: The optimal self-reporting scheme is superior to the optimal scheme without self-reporting.*

Finally, it is interesting to compare the optimal probabilities of examination with and without self-reporting, using equations (2.4) and (2.8). The optimal probabilities generally differ because the costs of deterring the marginal individual differ. On one hand, this marginal enforcement cost tends to be lower with self-reporting because an increase in the probability of examination applies only to deterred individuals. On the other hand, the marginal enforcement cost tends to be higher with self-reporting because an increase in the probability enlarges the pool of individuals subject to examination by deterring more individuals (an effect not present without self-reporting because all individuals are in the pool in any event). Either of these tendencies could be dominant, so that the optimal probability with self-reporting could be either higher or lower than the optimal probability without self-reporting.[13]

3. The N-Act Model

The model of section 2 can be generalized as follows. There are n harmful acts, where act i causes harm of h_i, $0 < h_1 < \ldots < h_n$. The population is divided into groups of size θ_i; individuals in group i choose between not acting and committing the act that causes harm h_i.[14] (For convenience, not acting is sometimes referred to as committing act 0.) Otherwise, the assumptions are as before: individuals obtain a benefit b if they commit a harmful act, where b is distributed according to $f(\cdot)$; examinations cost c; the maximum feasible sanction is $\bar{s} \geq h_n$.

A. Enforcement Without Self-Reporting

Because an individual in group i will commit a harmful act if and only if $b \geq ps_i$, where s_i is the sanction for committing act i, social welfare is

$$(3.1) \qquad W = \sum_{i=1}^{n} \theta_i \left[\int_{ps_i}^{\infty} (b - h_i) f(b)db - pc \right].$$

Thus,

$$(3.2) \qquad \frac{dW}{ds_i} = \theta_i p(h_i - ps_i)f(ps_i).$$

Assume that $p^* > 0$. It then follows from (3.2) that the optimum is $s_i^* = h_i/p^*$ if this is feasible -- that is, if $h_i/p^* \leq \bar{s}$. Otherwise, $s_i^* = \bar{s}$. In other words, optimal sanctions rise with the level of harm and lead to first-best behavior (individuals commit harmful acts if and only if their benefit exceeds the harm) until the maximum feasible sanction is reached; any

[13] To illustrate, consider the case where $f(\cdot)$ is uniform. Subtracting (2.3) from (2.7) yields

$$\frac{c(2c/\bar{s} + 1 - 2h)}{\bar{s}(\bar{s} + 2c)}.$$

The numerator obviously can be positive or negative. (It will be positive, implying that the optimal probability with self-reporting will be higher, if c is large enough relative to $\bar{s}$ or h is low enough -- conditions indicating that in schemes with and without self-reporting it is optimal to deter only a small fraction of the population.)

[14] At the end of this section, we consider the case in which each individual may choose any of the n acts.

acts subject to this sanction are underdeterred. Let I denote the set of i for which $s_i^* = \bar{s}$. Then we have

$$(3.3) \qquad \frac{dW}{dp} = \sum_{i \in I} \theta_i \bar{s}(h_i - p\bar{s})f(p\bar{s}) - c.$$

Thus, as in the one-act model, the optimum involves $p^* = 0$ if c is sufficiently large. If p^* is positive, (3.3) is zero. In this case, the set I cannot be empty (it must include n), for otherwise (3.3) is negative. We now can state the following analog of proposition 1.

Proposition 5: When there is no self-reporting in the n-act model:

- ***a. The optimal probability of examination p^* may be zero.***
- ***b. If p^* is positive, its level is determined by setting (3.3) equal to zero, and the optimal sanction s_i^* for acts of type i is h_i/p^* if this is feasible and is the maximum feasible sanction $\bar{s}$ otherwise.***

B. Enforcement With Self-Reporting

Now suppose that individuals must report a type of act: a number in the set {0, 1,..., n}. Individuals who report i pay an ex ante sanction r_i; they are then examined with probability p_i and, if examined, pay an ex post sanction s_{ij}, where j is their true act. The maximum amount that an individual may be sanctioned, $r_i + s_{ij}$, is $\bar{s}$, and all sanctions are nonnegative.[15] The social authority chooses an enforcement mechanism -- a set of r_i, p_i, and s_{ij} -- to maximize social welfare.

In an appendix, we indicate why the revelation principle applies, so attention may be confined to mechanisms in which individuals who commit act i truthfully report i. We then use the fact that the optimal scheme induces truth-telling at minimum enforcement cost to establish several results. First, the optimal ex post sanctions involve the maximal penalty for lying and no penalty for telling the truth. That is,

[15] The analysis in the appendix demonstrating that s_{ii}^* and r_0^* equal zero suggests that welfare could be further increased if negative sanctions -- rewards -- for telling the truth and for not doing harm were permitted (contrary to actual practice). However, if we allowed for rewards of up to some limit, as did Border and Sobel (1987), our results would not change. (Mookherjee and Png (1989) impose no such constraint and use the assumption of risk aversion to limit the optimal size of rewards.)

$$s_{ij}^* = \bar{s} - r_i, \text{ for } i \neq j, \text{ and} \tag{3.4}$$

$$s_{ii}^* = 0. \tag{3.5}$$

Also, those who report not having acted are not sanctioned, so

$$r_0^* = 0. \tag{3.6}$$

Finally, we show that the probabilities of examination for a given set of r_i must obey

$$p_i = \frac{\bar{r} - r_i}{\bar{s} - r_i}, \tag{3.7}$$

where $\bar{r}$ denotes the highest of the r_i. It follows from (3.7) that if $r_i = \bar{r}$, then $p_i = 0$; also, if $r_i > r_j$, then $p_i < p_j$.

Because individuals report the truth and, when doing so, bear only the ex ante sanction, their (expected) sanction for committing act i is simply r_i. Thus, an individual in group i commits act i if and only if $b \geq r_i$, so social welfare can be written

$$W = \sum_{i=1}^{n} \theta_i \left[\int_{r_i}^{\infty} (b - h_i)f(b)db - c[p_i(1 - F(r_i)) + p_0F(r_i)] \right]. \tag{3.8}$$

Note that the second expression in brackets measures examination costs: the fraction $1 - F(r_i)$ of group i commit act i and report i, so they are examined with probability p_i; the remaining fraction of group i do not commit act i, so they report 0 and are examined with probability p_0.

We now prove the analog of proposition 2, that any behavior resulting under a scheme without self-reporting can be induced with self-reporting at a lower enforcement cost. Let $p > 0$ and s_i $(i = 1, \ldots, n)$ apply without self-reporting. With self-reporting, set $r_i = ps_i$ and $r_0 = 0$; also, set the s_{ij} as in (3.4) and (3.5) and the p_i as in (3.7). The decision whether to commit acts will be the same as it was without self-reporting (as $r_i = ps_i$), so the integrals in (3.1) and (3.8) will be equal. To compare enforcement costs, observe from (3.7) that, with self-reporting, for all i, $p_i \leq p_0 = \bar{r}/\bar{s}$. Moreover, $\bar{r} = ps_j$ for some j, so $p = \bar{r}/s_j \geq \bar{r}/\bar{s}$. Therefore, $p \geq p_i$ for all i, and $p_i < p$ for all $i > 0$ such that $s_i > 0$.[16] Thus, the enforcement cost term in (3.8) is strictly less than that in (3.1) (because only the undeterred are examined with the highest probability rather than the entire population). This establishes *Proposition 6: In the n-act model, given any enforcement scheme (involving $p > 0$) without self-reporting, there exists a scheme with self-reporting under which behavior is the same but enforcement costs are lower.*

At this point, we can determine the optimal ex ante sanctions by maximizing (3.8) over the r_i, where the p_i are determined by (3.7). For any $r_i < \bar{r}$, we have

$$(3.9)\quad \frac{dW}{dr_i} = \theta_i\left[(h_i - r_i)f(r_i) - c[(p_0 - p_i)f(r_i) + (1 - F(r_i))\frac{\bar{r} - \bar{s}}{(\bar{s} - r_i)^2}]\right].$$

The first term in brackets in (3.9) is the direct social benefit from deterring the marginal individual in group i from committing act i: harm of h_i is avoided, but his benefits of r_i are lost (the marginal individual's benefit equals the sanction r_i). The remainder of the expression is the change in examination costs. The first component is a cost arising because individuals who are deterred are examined at rate p_0 rather than at rate p_i (and (3.7) implies $p_0 > p_i$ for $r_i > 0$ because $r_0* = 0$). The second component is a benefit arising because those who commit act i (the fraction $1 - F(r_i)$ of group i) are examined less frequently (from (3.7), the optimal p_i falls as r_i rises[17]). Because these two components are of opposite sign, $h_i - r_i$ may be positive or negative at an interior optimum, when $dW/dr_i = 0$.[18] Thus, r_i* may be such that

[16] If $s_i = 0$ for all i, $r_i = 0$, which implies $p_i = 0$, for all i, so the result that enforcement costs can be lowered with self-reporting follows trivially.

[17] This follows because the numerator of the derivative of (3.7) with respect to r_i is $\bar{r} - \bar{s}$, which is negative since we assume that $\bar{s} \geq h_n$ and demonstrate below that $\bar{r}* < h_n$. Without our assumption, the numerator could equal zero -- that is, $\bar{r}* = \bar{s}$ is possible. (If c is sufficiently small and $\bar{s} < h_n$, $dW/d\bar{r}$ in (3.11) can be positive at $\bar{r} = \bar{s}$.)

It should be noted that if $\bar{r}* = \bar{s}$, then $p_i* = 1$ for all i such that $r_i* < \bar{r}*$ (see (3.7)). In this case, $r_i* = h_i$ for all $h_i < \bar{r}*$. (In (3.9), $p_0 = p_i$ and $\bar{r} = \bar{s}$, so $dW/dr_i = \theta_i(h_i - r_i)f(r_i)$.) Also, our results concerning the relationship among the r_i* and the advantage of self-reporting (those subject to $\bar{r}$ need not be examined) would hold.

there is either underdeterrence or overdeterrence relative to first-best behavior.

Using (3.9), we can show that $r_i* > 0$ for any $r_i* < \bar{r}*$ (other than r_0*), because dW/dr_i at $r_i = 0$ is positive. Specifically, the first term in brackets is $h_i f(0)$, which is positive; the first component of the second term is zero (for, from (3.7), $p_i = p_0$ at $r_i = 0$); the second component of the second term is positive. The explanation is that increasing r_i from 0 has deterrence benefits and reduces the rate at which individuals of type i must be examined.

It also follows from (3.9) that, for any r_i* and r_j* (other than r_0*) less than $\bar{r}*$,

$$(3.10) \qquad h_i > h_j \Rightarrow r_i* > r_j*.$$

That is, the ex ante sanction (which, after all, equals the expected sanction) increases with the harm for those not subject to the highest sanction. The proof is in the appendix.

We now determine the optimal level of $\bar{r}$. Let J be the set of positive j such that $r_j = \bar{r}$. Then, varying $\bar{r}$ for all $j \in J$, we obtain[19]

[18] Because the first component is a marginal effect and the second an inframarginal effect, the relationship of the two will depend, among other things, on the shape of the distribution $f(\cdot)$. It can be demonstrated that there exist parameters and distributions consistent with our assumptions such that either component may dominate at the optimum.

[19] If $n = 1$, (3.11) reduces to

$$\frac{dW}{d\bar{r}} = (h_1 - \bar{r})f(\bar{r}) - cp_0 f(\bar{r}) - \frac{c}{s}F(\bar{r}),$$

which is what is obtained in the one-act model by differentiating W in (2.5) with respect to r, using the relationship $r = p\bar{s}$. (To facilitate the comparison to a scheme without self-reporting, we had differentiated W with respect to p rather than r in the one-act model.)

$$(3.11)\qquad \frac{dW}{d\bar{r}} = \sum_{j\in J} \theta_j \left[(h_j - \bar{r})f(\bar{r}) - c[p_0 f(\bar{r}) + \frac{F(\bar{r})}{\bar{s}}]\right]$$

$$- c \sum_{i\notin J} \theta_i \, [\frac{1 - F(r_i)}{\bar{s} - r_i} + \frac{F(r_i)}{\bar{s}}].$$

The summation in (3.11) over $j\in J$ has an interpretation similar to that of (3.9), except for the last component: in (3.9), those who commit the act (the fraction $1 - F(r_i)$) are examined at a lower rate because p_i falls as r_i increases; in (3.11), those who are deterred (the fraction $F(\bar{r})$ of group j) must be examined at a higher rate, because p_0 increases as $\bar{r}$ increases. The summation over the $i\notin J$ is the cost of examining individuals not subject to $\bar{r}$ more frequently (from (3.7), all the p_i, including p_0, rise with $\bar{r}$). Observe that all the terms except the first component of the first summation, $(h_j - \bar{r})f(\bar{r})$, are negative. Thus, if $dW/d\bar{r} = 0$ at the optimum, the sum over $j\in J$ of the first components must be positive; that is, on average, there must be underdeterrence of individuals subject to the highest sanction. It need not be the case, however, that all acts subject to $\bar{r}$ are underdeterred at the optimum.[20]

Now let us show that some degree of enforcement is optimal -- that is, $\bar{r}^* > 0$. Suppose that $\bar{r} = 0$, so that $J = \{1,\ldots, n\}$, and consider raising all the r_i (except r_0) uniformly. Evaluating the expression for marginal welfare (3.11) at 0 yields

$$(3.12)\qquad \frac{dW}{d\bar{r}} = \sum_{i=1}^{n} \theta_i h_i f(0) > 0.$$

Thus, $\bar{r}^* > 0$. From (3.7), this implies $p_0^* > 0$, so some enforcement effort is applied at the optimum. The reason is essentially that given for proposition 3b in section 2: when the sanctions r_i for all the acts are raised

[20] For an overdeterred act, say act k, to be subject to $\bar{r}$ at the optimum, it must be that dW/dr_k in (3.9) is positive when evaluated at $\bar{r}$. This possibility cannot be ruled out. Although overdeterrence implies that the first term in (3.9) is negative and (3.7) implies that the second term is negative ($p_k = 0$ at $r_k = \bar{r}$), the third term is positive.

simultaneously from zero, there is a first-order social benefit due to deterrence, but no first-order examination cost is borne because no one is being examined.

Last, we prove in the appendix that the acts subject to the largest ex ante sanction, $\bar{r}^*$, are the most harmful acts -- that is, if $r_i^* < r_j^* = \bar{r}^*$, then $h_i < h_j$.

Our results about the optimal self-reporting mechanism are summarized in the following two propositions.[21]

Proposition 7: Under the optimal self-reporting scheme in the n-act model, the following hold:

- *a. All individuals report their behavior truthfully.*
- *b. Individuals who commit act i pay a certain ex ante sanction r_i^* and no more, for there is no ex post sanction for having told the truth -- $s_{ii}^* = 0$; also, the ex post sanction for lying can be taken to be maximal -- $s_{ij}^* = \bar{s} - r_i^*$ for $i \neq j$.*
- *c. Some individuals are deterred from committing each of the harmful acts.*
  - *i. The ex ante sanction is positive for all harmful acts and zero for not committing a harmful act; the r_i^* rise with the level of harm until reaching a maximum $\bar{r}^*$ at some h_j and are $\bar{r}^*$ thereafter -- $0 = r_0^* < r_1^* < \ldots < r_j^* = \ldots = r_n^* = \bar{r}^*$; $\bar{r}^*$ and the lesser r_i^* are determined by setting (3.11) and (3.9) respectively equal to zero.*
  - *ii. The most harmful act is underdeterred relative to first-best behavior -- $r_n^* < h_n$; other acts may be underdeterred or overdeterred, although acts subject to the highest ex ante sanction $\bar{r}^*$ are underdeterred on average.*
- *d. The probability of examination is highest for those who report not having committed a harmful act; for those who report having committed harmful acts, the p_i^* fall with the level of harm until reaching zero for the most harmful acts (those subject to the highest ex ante sanction $\bar{r}$) -- $p_0^* > p_1^* > \ldots > p_j^* = \ldots = p_n^* = 0$; the p_i^* are given by (3.7).*

Proposition 8: In the n-act model, the optimal self-reporting scheme is superior to the optimal scheme without self-reporting.

Observe that proposition 7 justifies implicit assumptions made in the one-act model in section 2. There we assumed that those who commit the harmful act (which is trivially the most harmful act) are not examined, those who do

[21] In the mechanism design literature on auditing cited in note 2, it also generally is true that efficient auditing involves maximal penalties for lying, no penalties for telling the truth, and greater audit probabilities for reports associated with lower payments. But our characterization of the optimal r_i and their relationship to the h_i (as well as our extensions in section 4) are not in the auditing literature, because our model is addressed to the optimal control of harmful externalities.

not report having committed the act pay no ex ante sanction, and those who truthfully report not having committed the act bear no ex post sanction. Proposition 7 states that each of these restrictions on the enforcement mechanism is in fact a feature of the optimal mechanism.

Finally, let us consider the consequences of relaxing the assumption that each individual chooses between committing one harmful act and not acting. Instead we can allow individuals to choose among any of the n harmful acts or not acting, which presents the issue of marginal deterrence.[22] In this case, an individual will choose the act for which the excess of the benefit over the expected sanction is largest, unless the net benefit of that act is negative, in which event he will not commit a harmful act.[23] In the appendix, we sketch the argument establishing that our results continue to hold in the case of marginal deterrence, except that the first-order conditions determining the optimal r_i must be modified.

4. Extensions

In this section, we discuss a number of extensions of our analysis, illustrating them in the one-act model for convenience.

A. Risk Aversion

As we suggested at the outset, an important feature of self-reporting schemes is that individuals need not bear any risk of sanctions: they can be induced to report their true behavior and pay sanctions with certainty. Without self-reporting, by contrast, individuals who commit harmful acts must bear the risk of sanctions, which is socially costly if individuals are risk-averse.

[22] Stigler (1970) chose the term marginal deterrence because an individual's choice between two harmful acts depends on the difference, or margin, between expected sanctions for the two acts.

[23] Specifically, an individual is assumed to obtain a benefit b_i if he commits act i, where the b_i are independently and identically distributed according to the previously described density $f(\cdot)$. Thus, individuals choose the act i that maximizes $b_i - r_i$, unless this maximum is negative.

To elaborate, we modify the one-act model to incorporate risk aversion. Assume that an individual's utility consists of three separable components: a strictly concave function $u(\cdot)$ of his net wealth, plus the benefit of committing the harmful act if it is committed, minus the harm suffered due to other individuals committing the harmful act.[24] An individual's net wealth is his initial wealth, w, minus any sanctions paid and a lump-sum tax t equal to net per capita enforcement expenses (examination costs minus fine revenue). Social welfare is the sum of individuals' expected utilities.

If there is no self-reporting, individuals will commit the harmful act if and only if

$$(4.1) \qquad (1-p)u(w - t) + pu(w - t - s) + b \geq u(w - t).$$

Let $\rho(p,s)$ be the certainty equivalent of being subject to a sanction of s with probability p. Individuals will thus commit the harmful act if and only if

$$(4.2) \qquad b \geq u(w - t) - u(w - t - \rho(p,s)).$$

Define the right side of (4.2) as $\beta(\rho)$. The enforcement authority's problem is to maximize social welfare

$$(4.3) \qquad W = \int_{\beta(\rho)}^{\infty} [u(w - t - \rho(p,s)) + b - h]f(b)db + u(w - t)F(\beta(\rho)),$$

subject to

$$t = pc - ps(1-F(\beta(\rho))).$$

If there is self-reporting, individuals who commit the harmful act and report this obtain utility (abstracting from harm suffered) of

[24] The relevance of the assumption that the benefit is additively separable is explained in note 25.

$$(4.4)\qquad u(w - t - r) + b,$$

and those who commit the act but do not report it obtain utility of

$$(4.5)\qquad (1-p)u(w - t) + pu(w - t - s) + b.$$

Thus, an individual who commits the harmful act will report it when[25]

$$(4.6)\qquad u(w - t - r) \geq (1-p)u(w - t) + pu(w - t - s).$$

We assume that (4.6) applies -- that individuals who commit the act are induced to report it. Otherwise, as in section 2, the situation is as if there is no self-reporting. Moreover, at the optimum (4.6) holds as an equality: if not, as before, p could be lowered, maintaining (4.6) and thus individuals' behavior, but reducing enforcement costs. Let r(p,s) denote the r that makes (4.6) an equality given p and s. Individuals consequently will commit the harmful act if and only if

$$(4.7)\qquad b \geq u(w - t) - u(w - t - r(p,s)).$$

Define the right side of this expression to be $\beta(r)$. The problem is to maximize social welfare

$$(4.8)\qquad W = \int_{\beta(r)}^{\infty} [u(w - t - r(p,s)) + b - h]f(b)db + u(w - t)F(\beta(r)),$$

subject to

$$t = pcF(\beta(r)) - r(1-F(\beta(r))).$$

We can demonstrate that self-reporting schemes are superior to schemes without self-reporting by comparing expressions (4.8) and (4.3) for any positive p and s. Let us evaluate both expressions at the t given in (4.3).

[25] Note that (4.6) does not depend on b. If we had assumed that b were included in wealth, rather than entering utility in an additively separable manner, then whether individuals report acts would generally depend on b because the certainty equivalent of the ex post sanction depends on total wealth. Self-reporting would, however, remain advantageous as long as some individuals would report their acts under the optimal scheme.

Observe first that $\rho(p,s)$ in (4.3) must equal $r(p,s)$ in (4.8), because each measures the certainty equivalent of being subject to sanction s with probability p, while initial net wealth, $w - t$, is the same in each case. As a result, the values of expressions (4.3) and (4.8) are the same. Second, let us show that the t in (4.3) -- which by assumption makes the government break even without self-reporting -- produces a surplus with self-reporting. This will mean that achievable welfare must be higher with self-reporting.[26] The government will be in surplus for two reasons, as is evident from comparing the expressions for t in (4.3) and (4.8). First, enforcement costs with self-reporting are $pcF(\beta)$ rather than pc; this savings, due to examining only those who do not report having committed the act, is that identified in the risk-neutral case. Second, revenues from payments of sanctions are $r(1 - F(\beta))$ rather than $ps(1 - F(\beta))$, because those who are not deterred pay r for certain rather than ps on average. And $r > ps$: $r - ps$ is the risk premium for avoiding exposure to s with probability p (see (4.6)). The increase in revenue from sanctions reflects the social benefit of eliminating risk under self-reporting.

It can be demonstrated that our other results continue to hold when individuals are risk-averse (although the characterizations of the optimal probability and ex ante sanctions differ because the certainty equivalent of sanctions rather than their expected value determines behavior). Specifically, with self-reporting, the optimum involves all individuals reporting truthfully, and optimal enforcement effort is positive.[27]

[26] With the surplus, t can be reduced and welfare raised (even though behavior might change due to wealth effects, requiring an offsetting increase in p to keep behavior constant).

[27] When $r = 0$, individuals all have wealth equal to w and thus equal marginal utilities of wealth, so raising r has no distributive effects, while, as in the risk-neutral case, it produces first-order benefits from deterrence but no first-order examination costs. Formally,

$$\frac{dW}{dr} = h\beta' f(\beta) - [(1-F(\beta))u'_a + F(\beta)u'_{na}][(pc + r)\beta' f(\beta) + p'cF(\beta)]$$

$$- (u'_a - u'_{na})F(\beta)(1-F(\beta)),$$

where u'_a and u'_{na} denote the marginal utility of wealth for those who act and those who do not act. When $r = 0$, the second and third terms are both zero, and the first term is positive.

Achievable welfare is greater with self-reporting than without it, but now there are two reasons: reduction of enforcement costs and elimination of risk-bearing costs. Finally, because no risk is borne, the optimal ex post sanction is $\bar{s}$ (as it was in section 2). In contrast, without self-reporting the optimal enforcement scheme may change substantially because of risk aversion; in particular, the optimal sanction may be less than maximal.[28]

B. Imprisonment as a Sanction

Suppose that imprisonment, a socially costly sanction, may be employed as a supplement to monetary sanctions, which we have assumed to be socially costless to impose when individuals are risk-neutral (as we assume is true in this subsection). Then schemes with self-reporting have the additional advantage that society can enjoy the deterrence benefits of imprisonment without imposing any imprisonment or imposing it to a lesser extent than in schemes without self-reporting.

Let us demonstrate this advantage in the one-act model. Denote the ex post monetary sanction by s_1, where $s_1 \leq \bar{s}_1$ (the maximum monetary sanction, perhaps equal to wealth), and the ex post sanction of imprisonment by s_2, where $s_2 \leq \bar{s}_2$ (the maximum term of imprisonment). The disutility of sanctions to individuals is s, where $s = s_1 + s_2$; the social cost of imposing s_2 is ηs_2, where $\eta > 0$. Observe that it is desirable for society to employ monetary sanctions to their limit $\bar{s}_1$ before resort to imprisonment: otherwise, s_2 could be lowered and s_1 raised, keeping s (and thus behavior) the same but reducing the social costs of using imprisonment.

Now assume that, without self-reporting, imprisonment is employed probabilistically -- that is, $s = \bar{s}_1 + s_2$, where $s_2 > 0$, and $0 < p < 1$. With self-reporting, choose $r = ps$. As before, individuals' behavior will be the same as without self-reporting and there will be the usual advantage of conserving on examination costs, because those who commit the harmful act are induced to report this and are not examined. Now, however, there is the further advantage of reducing the use of imprisonment. Specifically, define

[28] See Polinsky and Shavell (1979).

r_1 and r_2 as the ex ante monetary sanction and the ex ante term of imprisonment, respectively. Then, $r = ps$ is equivalent to $r_1 + r_2 = p(\bar{s}_1 + s_2)$. If $r \leq \bar{s}_1$, set $r_1 = r$ and $r_2 = 0$; hence, there is no imprisonment, producing a savings of $p\eta s_2$. If $r > \bar{s}_1$, set $r_1 = \bar{s}_1$ and $r_2 = p(\bar{s}_1 + s_2) - \bar{s}_1$; then the savings in imprisonment costs is $(1-p)\eta\bar{s}_1$. The advantage of self-reporting in this latter case is that $\bar{s}_1$ is imposed with certainty rather than only with probability p, so that the use of imprisonment is diminished by $(1-p)\bar{s}_1$.

The idea underlying the above argument may be expressed informally as follows. With self-reporting, the ex ante sanctions that are actually imposed are lower in magnitude by a factor of p than those that are necessary to impose ex post without self-reporting. Because costless monetary sanctions are used before imprisonment, this reduction in the magnitude of imposed sanctions with self-reporting allows society to reduce or eliminate the actual imposition of imprisonment.

The conclusion that imprisonment costs can be saved with self-reporting is relevant whenever imprisonment would be desirable to impose without self-reporting. But even when imprisonment would not be desirable to employ without self-reporting, the threat of imprisonment as an ex post sanction for those who fail to report their harmful acts always enhances the advantages of self-reporting schemes. A given level of deterrence -- a given ex ante sanction r -- can be achieved more cheaply, with a lower probability of examination, because those who would report falsely face a greater ex post sanction than otherwise. Furthermore, because ex post sanctions are never actually imposed, no social costs of imprisonment are incurred.

C. Errors in Examinations

We assumed throughout that individuals' true behavior would be accurately determined in examinations by the enforcement authority. Suppose instead that their behavior is sometimes assessed erroneously. This will decrease achievable welfare in schemes with and without self-reporting, but (perhaps surprisingly) will increase the relative advantage of self-reporting schemes.

Assume that if a person does not commit the harmful act and is examined, he will mistakenly be found to have committed the act with probability q_1; if he commits the harmful act and is examined, he will erroneously be found not to have committed the act with probability q_0.

Without self-reporting, an individual who does not commit the harmful act bears an expected sanction of pq_1s rather than zero, and a person who does commit the act bears an expected sanction of $p(1-q_0)s$ rather than ps. Thus, individuals will commit the act if and only if

$$(4.9) \qquad b \geq ps(1-q_0-q_1).$$

Now let us demonstrate that with self-reporting the same behavior can be achieved as without self-reporting, but at lower enforcement cost. Keep p and s at the same levels as without self-reporting and set $r = p(1-q_0)s$. If a person commits the harmful act and does not report it, the expected sanction will be $p(1-q_0)s$, so he will report it; if he does not commit the act, the expected sanction will be pq_1s. Thus, individuals will commit the harmful act if and only if (4.9) holds, the same condition as without self-reporting. Although behavior is the same under both schemes, enforcement costs with self-reporting are lower by $(1 - F(ps(1-q_0-q_1)))pc$, because those who commit the act are not examined. Moreover, observe from (4.9) that as the magnitude of errors q_0 and q_1 increases, the level of p necessary to achieve a given level of deterrence increases with and without self-reporting (by the same amount), so achievable welfare is reduced in both schemes. Also, the savings in enforcement costs under self-reporting are greater: the benefit from not examining those who report committing the act rises when p must be increased on account of error to maintain deterrence.

Because individuals who truthfully report not having committed the harmful act might mistakenly be deemed to have committed it, the imposition of ex post sanctions is not entirely avoided under self-reporting. Consequently, when individuals are risk-averse or imprisonment is used, social costs associated with the ex post imposition of sanctions are incurred, and it may be optimal to adjust the enforcement scheme, as by lowering sanctions. But the advantages of self-reporting with regard to saving risk-bearing costs and imprisonment costs are still present. Consider a given p and s. Individuals who do not commit the harmful act are in the same situation with and without self-reporting: they are exposed to the same chance of bearing ex post sanctions through error. But individuals who commit the act are subject to ex post sanctions only when there is no self-reporting; with self-reporting, such individuals report committing the act and are subject to ex ante sanctions alone. Thus, schemes with self-reporting continue to have the benefit of reducing sanctioning costs for those who commit harmful acts. Moreover, to achieve a given level of deterrence, greater sanctioning costs must be imposed on those who commit harmful acts when errors sometimes are made, so this advantage of self-reporting is enhanced.

D. Administrative Costs of Self-Reporting

We assumed in the model that the only social costs associated with enforcement were the costs of examining individuals' behavior. However, processing reports and collecting payments involves administrative costs. This is a disadvantage of self-reporting because, when an individual reports his behavior and pays a sanction, society bears administrative costs with certainty, whereas without self-reporting society bears administrative costs only with a probability.

To illustrate, assume that collecting a positive payment, whether ex ante or ex post, involves a fixed administrative cost d. Without self-reporting, the level of social welfare previously given by expression (2.1) is reduced by $pd(1 - F(ps))$, since only those who commit the harmful act and are examined make payments. With self-reporting, the level of welfare previously given by (2.5) is reduced by $d(1 - F(r))$, because those who commit the act make payments with certainty. Thus, the argument of proposition 2 -- that the same behavior can be induced with self-reporting as without it but at lower social cost -- may no longer hold. If r is set equal to ps, the same behavior is produced under self-reporting, but the savings in enforcement and administrative costs is now

$$(1 - F(r))(pc - (1-p)d). \tag{4.10}$$

The savings depends on the fraction of the population who commit the act and report this under self-reporting, $1 - F(r)$, because individuals who do not commit the act are subject to the same treatment under both schemes. For those who commit the act, self-reporting schemes save pc because examinations need not be conducted for those who commit the act, whereas examinations otherwise would be conducted with probability p at unit cost c. But self-reporting schemes involve the additional cost $(1-p)d$ because payments must be collected at unit cost d from those who would not have been examined without self-reporting (the fraction $1-p$). Whether self-reporting remains preferable depends on whether c and p are sufficiently large relative to d.[29]

5. Concluding Remarks

Methods of Enforcement. Our analysis demonstrated that enforcement costs could be saved when enforcement took the form of examination, by which we meant that the enforcement authority randomly selected individuals and determined whether they committed the harmful act. We briefly discuss here two other methods of enforcement.

First, consider investigation -- determining who committed a particular harmful act that the enforcement authority already knows occurred. This type of enforcement is typical in the area of crime, where victims often inform the police of harms that they have suffered. (By contrast, examination is used in areas such as environmental regulation, where the occurrence of harmful acts may not be immediately apparent and later attempts to trace harm to particular sources may be impossible.)

[29] In the n-act model, it may not be optimal to have individuals report and pay positive ex ante sanctions for acts whose harm is below some threshold (because raising r_i from zero requires that the administrative cost of reporting be incurred). These acts would be subject only to ex post sanctions (set to optimize deterrence, as when there is no self-reporting).

When enforcement is by investigation, self-reporting does not merely reduce enforcement costs, it eliminates them: once someone confesses, others need not be investigated. (In our model, by comparison, one person's admission that he committed a harmful act does not rule out the possibility that others may also have committed harmful acts.) To realize this savings in enforcement costs, individuals must be induced to admit committing harmful acts, and (as in our model) this can best be accomplished by setting the ex ante sanction for those who admit committing harmful acts equal to the expected ex post sanction. Thus, the reduction in the sanction for admitting one's act should be greater the lower would have been the probability of apprehending the person through investigation.[30] Accordingly, if a person confesses when the police have little evidence (such as immediately after a crime is committed), the reduction in his sanction should be large, but if a person confesses when the police have already gathered substantial evidence against him, the reduction should be small.

Second, consider monitoring -- the posting of enforcement agents to observe violations among any of a population of parties, as when police are stationed at the roadside. Monitoring is useful when a single agent is readily able to spot any violations that occur within sight of his post. (Monitoring is not enough, and examination or investigation is necessary, when extra effort is required to detect any particular individual's violation.)

When monitoring is the method of enforcement, there may be no cost savings achievable under self-reporting. For example, even if individuals who wish to speed or make illegal left turns were to report this in advance to the police, there would be little if any reduction in the number of officers that would have to be posted in order to maintain the probability of apprehension for other drivers who might commit violations.

Why Individuals Might Not Report Truthfully. In the model (as well as in the extensions of it), individuals report truthfully given socially optimal enforcement. We do not, however, observe all individuals reporting the truth.

[30] The necessary reduction is $s - r = s - ps = s(1-p)$.

There are two plausible explanations for this. First, the perceived or actual expected sanctions for failing to report or lying may be relatively low for certain individuals. While examination rates could be increased in an attempt to induce these individuals to report truthfully, this would be costly. And to the extent that individuals are not induced to report the truth, the savings in examination costs and the reduction in risk-bearing and in the use of imprisonment will be lower.

Second, some individuals may not be aware of the nature of the act that they have committed. The social authority could attempt to remedy this problem by adjusting the enforcement mechanism to increase the expected cost of false reports, thereby inducing uninformed individuals to acquire information so that they could report truthfully. This, however, would be costly. Also, it may not be desirable, because the savings in examination costs (incurred probabilistically) when such individuals report the truth may be less than individuals' information acquisition costs (incurred ex ante, with certainty).

The Use of Self-Reporting. That self-reporting is a frequently observed feature of law enforcement is consistent with our analysis, for it seems that in many contexts significant enforcement resources or costs of imposing sanctions can be saved by inducing people to come forward with information about their conduct. At the same time, it is not surprising that self-reporting is not observed in some instances. With regard to the example of driving violations (like improper left turns) that are not reported by those who commit them, two of the limitations of self-reporting are relevant. The administrative cost of processing reports of many types of driving violations would be large relative to the expected harm they cause, and the number of police necessary to maintain a given level of deterrence would not be much reduced if some violations were reported.

It does not appear, however, that the benefits of self-reporting are fully realized in practice. Notably, the incentives to report one's conduct frequently seem weak, as the reduction in penalties for parties who admit harmful behavior is often modest even when the probability of punishment for those not reporting their violations is substantially less than one. When this is the case, increasing incentives for reporting harmful acts would induce more reporting and raise welfare.

Finally, we should remark that although our discussion throughout has focused on public enforcement of law, it is more broadly relevant, to enforcement through private suit and to enforcement of incentive schemes in private contractual arrangements. For example, employers may have a policy of treating more favorably employees who admit to drug use or pilferage than employees who are found out. Inducing employees to report their own misconduct reduces the employer's need to police employee behavior and also the need to impose costly sanctions (such as dismissal).

References

Baiman, Stanley and Joel S. Demski, 1980, Economically Optimal Performance Evaluation and Control Systems, Journal of Accounting Research 18, 184-220.

Baron, David P. and David Besanko, 1984, Regulation, Asymmetric Information, and Auditing, Rand Journal of Economics 15, 447-470.

Becker, Gary S., 1968, Crime and Punishment: An Economic Approach, Journal of Political Economy 76, 169-217.

Border, Kim C. and Joel Sobel, 1987, Samurai Accountant: A Theory of Audit and Plunder, Review of Economic Studies 54, 525-540.

Dye, Ronald A., 1986, Optimal Monitoring Policies in Agencies, Rand Journal of Economics 17, 339-350.

Mookherjee, Dilip and Ivan Png, 1989, Optimal Auditing, Insurance, and Redistribution, Quarterly Journal of Economics 104, 399-415.

Mookherjee, Dilip and Ivan Png, 1990, Enforcement Costs and the Optimal Progressivity of Income Taxes, Journal of Law, Economics, and Organization 6, 411-431.

Pigou, A.C., 1918, The Economics of Welfare (London: Macmillan).

Polinsky, A. Mitchell and Steven Shavell, 1979, The Optimal Tradeoff Between the Probability and Magnitude of Fines, American Economic Review 69, 880-891.

Reinganum, Jennifer F. and Louis L. Wilde, 1985, Income Tax Compliance in a Principal-Agent Framework, Journal of Public Economics 26, 1-18.

Scotchmer, Suzanne, 1987, Audit Classes and Tax Enforcement Policy, American Economic Review 77, 229-233.

Stigler, George J., 1970, The Optimum Enforcement of Laws, Journal of Political Economy 78, 526-536.

Townsend, Robert M., 1979, Optimal Contracts and Competitive Markets with Costly State Verification, Journal of Economic Theory 21, 265-293.

Wagenhofer, Alfred, 1987, Investigation Strategies with Costly Perfect Information, in Gunter Bamberg and Klaus Spremann, eds., Agency Theory, Information, and Incentives (Berlin: Springer-Verlag), 347-377.

</ref_section>

Appendix

Proof of Proposition 7

Here we prove the claims of proposition 7 that are not demonstrated in the text.

First, we observe that an individual who chooses act i and reports j bears an expected sanction of $r_j + p_j s_{ji}$, so he will report j(i), the report that minimizes the expected sanction for those who commit act i. Therefore, an individual in group i will commit act i rather than not act (commit act 0) if and only if

$$\text{(A.1)} \qquad b \geq (r_{j(i)} + p_{j(i)} s_{j(i)i}) - (r_{j(0)} + p_{j(0)} s_{j(0)0}).$$

The socially optimal mechanism can be assumed to be such that individuals report their acts truthfully. In other words, the revelation principle applies, even though individuals report only their acts, not their underlying types.[31] To explain (we omit details), suppose that an optimal mechanism involves an individual choosing act i but reporting j different from i. Then it must be that any other individual who chooses act i also reports j, for an individual will select the report that minimizes his expected sanction, and this report depends only on one's act because expected sanctions depend only on one's act and report. Now, alter the mechanism by relabeling j as i -- so that the treatment of a person who reports i is exactly what it had been under the original mechanism when he reported j. This altered mechanism will induce anyone who commits act i to report truthfully. Moreover, anyone who commits act i will be subject to the same expected sanction, so his behavior will be unchanged. By this method, we generally can construct a mechanism that is equivalent to the original one but which involves truthful reporting.

[31] We assume that individuals do not report their benefits. A justification is that verifying an individual's benefit would be prohibitively costly in many contexts.

Because we may assume reports are truthful, the following incentive compatibility constraints, denoted by IC_{ij}, must hold for all i and j:

(A.2) $r_i + p_i s_{ii} \leq r_j + p_j s_{ji}$.

This constraint requires that the expected sanction if a person commits act i and tells the truth does not exceed the expected sanction if he reports j instead.

We now demonstrate equation (3.4), which states that $s_{ij}* = \bar{s} - r_i$, for $i \neq j$. If (3.4) did not hold, one could alter the mechanism by raising s_{ij} to the point where (3.4) does hold. Since IC_{ji} is satisfied under the original mechanism, it would satisfied under the altered mechanism, as the right side of IC_{ji} would be greater and the left side would be unaffected. Moreover, individuals' choices of acts would be unaffected by raising any s_{ij}, because no individual would bear s_{ij}, as all would report truthfully. Thus, social welfare under the altered mechanism would be the same as under the original one.

Next, we show (3.5), that $s_{ii}* = 0$. If (3.5) did not hold, one could alter the mechanism by lowering s_{ii} to 0 and raising r_i by $p_i s_{ii}$. This alteration would not affect IC_{ij}: the left side would have the same value, and the right side would be unaffected. For the IC_{ji}, $j \neq i$, however, a higher r_i would increase the right side while the left side would be unaffected, so p_i could be lowered, which would reduce enforcement costs.[32] Finally, since the expected sanction for telling the truth under this altered mechanism would be the same as under the original mechanism, individuals' choices of acts would be unaffected. Thus, social welfare would be higher under the altered mechanism.

These two simplifications allow IC_{ij} in (A.2) to be rewritten as[33]

[32] If p_i in the optimal mechanism were zero, the argument in the text would not hold. But the level of s_{ii} would not affect behavior, so s_{ii} could be taken to be zero.

[33] Note that (A.3) holds trivially if $i = j$.

(A.3) $\quad r_i - r_j \leq p_j(\bar{s} - r_j)$.

That is, any savings in the ex ante sanction gained by reporting j rather than the true act i cannot exceed the expected ex post sanction for lying.

Using (A.3), we can establish (3.6), that $r_0* = 0$. First, we show that the lowest of the r_i, which we denote $\underline{r}$, must equal zero. If $\underline{r} > 0$, one could alter the mechanism by reducing each of the r_i by $\underline{r}$ and raising the s_{ij}, for $i \neq j$, as indicated by (3.4). This change would not affect the left side of IC_{ij} in (A.3) and would raise the right side, so the IC_{ij} would continue to hold. Moreover, each of the IC_{ij} could be satisfied with lower p_i, so enforcement costs could be reduced. Finally, it is apparent from (A.1) (which reduces to $b \geq r_i - r_0$) that individuals' choices of acts would be unaffected, so welfare would be higher under the altered mechanism. Second, $r_0* = \underline{r}$. If not, $r_k = 0$ for some $k > 0$, since $\underline{r} = 0$. Let $K = \{k|\ r_k = 0\}$ and alter the mechanism by increasing all the r_k, $k \in K$, by the same small amount -- in particular small enough that $r_k < r_m$ for all $m \notin K$ (including $m = 0$). The constraints IC_{ik} will continue to hold. (If $i \notin K$, the increase in r_k will relax the constraint, and if $i \in K$ the constraint will still be satisfied because the left side equals zero and the right side is nonnegative.) The constraints IC_{kj} will continue to hold as well. (The left side will be negative if $j \notin K$ and zero otherwise.) Finally, increasing the r_k will not affect behavior: $r_k - r_0$ is negative before and after the alteration, so all individuals in group k, for $k \in K$, commit their act regardless. Thus, if $r_0 > 0$, the optimal $\underline{r}$ need not equal zero, a contradiction.

Next, let us demonstrate that the incentive compatibility constraints IC_{ij} are binding for reports of acts subject to $\bar{r}$ and for no others -- that is, IC_{ij} is binding if and only if $r_i = \bar{r}$. First, the constraints for reports not subject to $\bar{r}$ are not binding. This is apparent from (A.3): the left side is greater the greater is r_i and the right side is independent of r_i, so the constraint can be binding only if $r_i = \bar{r}$. To prove that the constraints are binding for reports subject to $\bar{r}$, suppose instead that $\bar{r} - r_j < p_j(\bar{s} - r_j)$ for some j. Alter the mechanism by lowering p_j such that this inequality continues to hold. The constraint IC_{ij} in (A.3) will continue to hold for all i. Also, this reduction in p_j does not alter individuals' choices of acts (as the r_i are the same). But reducing p_j saves enforcement costs, so welfare is higher. Because the constraints (A.3) are binding when $r_i = \bar{r}$, the optimal probabilities are given by (3.7).

Finally, we demonstrate the relationships between the h_i and the optimal r_i. First, we prove (3.10), that for any r_i* and r_j* (other than r_0*) less than $\bar{r}$*, $h_i > h_j$ implies $r_i* > r_j*$. Observe that for any positive constant λ, the function λW is maximized at the same r_i as is W. In particular, $(1/\theta_i)W$ is maximized at r_i* and $(1/\theta_j)W$ is maximized at r_j*. Now, using (3.9), for any $r \in [0, \bar{r}]$, $(1/\theta_i)dW/dr_i - (1/\theta_j)dW/dr_j = (h_i - h_j)f(r)$, which is positive.[34] Thus, $r_i* = r_j*$ is ruled out. Also, $r_i* < r_j*$ is impossible: because r_j* maximizes $(1/\theta_j)W(r_j)$, this expression's value at r_j* cannot be exceeded by its value at r_i*; but then the value of $(1/\theta_i)W(r_i)$ must be greater at r_j* than at r_i* because the difference in the derivatives is positive over the interval $[r_i*, r_j*]$, which contradicts the optimality of r_i*.

Second, we show that, if $r_i* < r_j* = \bar{r}*$, then $h_i < h_j$. To see this, assume otherwise, that $h_i > h_j$. Consider first the case in which there is more than one type of act subject to $\bar{r}$*. Then, the derivative of welfare with respect to both r_i and r_j in the interval $[0, \bar{r}*]$ is given by expression (3.9), so the argument demonstrating (3.10) establishes that $r_i* > r_j*$, a contradiction. Now consider the case in which only act j is subject to $\bar{r}$*. Using expression (3.11), observe that, at $\bar{r}$*,

$$\text{(A.4)} \qquad \theta_j[(h_j - r_j*)f(r_j*) - cp_0f(r_j*)] > 0,$$

because all the other terms in (3.11) are negative. This implies that

[34] This result relies on our assumption that $f(\cdot)$ is independent of the type of act. Otherwise, (3.10) might not hold, because inframarginal effects on examination costs could be greater for the less harmful act.

$$(A.5) \qquad \theta_n[(h_n - r_n^*)f(r_n^*) - cp_0f(r_n^*)] > 0,$$

because $h_n > h_j$ and $r_n^* < r_j^*$. But then, from (3.9), $dW/dr_n > 0$ at r_n^*, because (A.5) is the first two components of (3.9) and all the other components are also positive. This contradicts the optimality of r_n^*.

Marginal Deterrence

Here we sketch the argument that our results extend to the case in which individuals may choose among any of the n harmful acts or not acting, except that the first-order conditions determining the optimal r_i must be modified. First, the revelation principle still applies, because all individuals who choose a given act will make the same report, the one that minimizes the expected sanction. Second, the analysis of the incentive compatibility constraints is unaffected, because the relevant arguments held expected sanctions for each act constant. Thus, the ex post sanction for telling the truth is zero and for lying is maximal; also, the ex ante sanction for those not committing any harmful act is zero and the examination probabilities are given by expression (3.7).

Using these results, social welfare can be expressed as

$$(A.6) \qquad W = \sum_{i=1}^{n} \int_{r_i}^{\infty} (b_i - h_i - cp_i)g_i(b_i)db_i - cp_0 \prod_{i=1}^{n} F(r_i).$$

The integral is the benefit net of harm and examination costs for those who commit harmful act i, where $g_i(b_i)$ is the fraction of individuals who commit act i and obtain benefit b_i. Observe that if $b_i < r_i$, then $g_i(b_i) = 0$; otherwise,

$$(A.7) \qquad g_i(b_i) = f(b_i) \prod_{j \neq 0,i} F(r_j + b_i - r_i).$$

This is because, for a person to choose act i, $b_i - r_i$ must exceed $b_j - r_j$ or, equivalently, b_j must be less than $r_j + b_i - r_i$ for all $j \neq i$. The second term in (A.6) is examination costs for those who do not commit a harmful act (which will be those for whom $b_i < r_i$ for all i).

The derivative with respect to the $r_i < \bar{r}$, analogous to (3.9), can be verified to equal

$$\text{(A.8)} \qquad \frac{dW}{dr_i} = \alpha_i\Big[(h_i - B_i) - \gamma_{i0}c(p_0 - p_i) + \sum_{j\neq 0,i} \gamma_{ij}[c(p_i - p_j) + B_{ij} - h_j]\Big]$$

$$- c(1 - G_i(r_i))\frac{\bar{r} - \bar{s}}{(\bar{s} - r_i)^2}.$$

The new variables in (A.8) are defined as follows:

$$\text{(A.9)} \qquad \alpha_i = g_i(r_i) - \int_{r_i}^{\infty} \frac{dg_i(b_i)}{dr_i}db_i,$$

which represents the fraction of the population who are just deterred from committing act i.

$$\text{(A.10)} \qquad \gamma_{i0} = \frac{g_i(r_i)}{\alpha_i},$$

which is the proportion of those who do not commit a harmful act among those who are just deterred from committing act i.

$$\text{(A.11)} \qquad \gamma_{ij} = \frac{1}{\alpha_i}\int_{r_j}^{\infty} \frac{dg_j(b_j)}{dr_i}db_j, \; j \neq 0, \; i,$$

which is the proportion of those who commit act j among those who are just deterred from committing act i.

$$\text{(A.12)} \qquad B_i = \frac{1}{\alpha_i}[r_i g_i(r_i) - \int_{r_i}^{\infty} b_i\frac{dg_i(b_i)}{dr_i}db_i],$$

which is the average benefit from committing act i of individuals who are just deterred from committing act i.

$$\text{(A.13)} \qquad B_{ij} = \frac{1}{\alpha_i\gamma_{ij}}\int_{r_j}^{\infty} b_j\frac{dg_j(b_j)}{dr_i}db_j, \; j \neq 0, \; i.$$

which is the average benefit from committing act j of individuals who are induced to commit act j when just deterred from committing act i.

$$\text{(A.14)} \qquad G_i(r_i) = \int_0^{r_i} g_i(b_i)db_i,$$

so that $1 - G_i(r_i)$ represents the fraction of the population who commit act i and report i.

The first two terms in brackets and the last term in (A.8) are analogous to the three terms of (3.9). The third term in brackets in (A.8) is the effect on welfare associated with individuals who, when deterred from committing act i, are induced to commit other harmful acts j (in the model of section 3, those who are deterred from committing i do not commit a harmful act). This term thus involves for each j a difference in expected examination costs, benefits from committing act j, and the harm caused by act j.

We now indicate why the r_i* (except r_0) are positive. Assume otherwise, that $r_j^* = 0$ for some $j \neq 0$. Consider the change in welfare resulting from raising each of the r_i except r_0 by the same small amount. Because the r_i were optimal, the change in welfare from raising any positive r_i must be zero. But the effect of raising r_j is positive, which contradicts the assumption that the r_i* were optimal. To see this, observe that the first term of dW/dr_j in (A.8) is positive because B_j equals zero. (All the r_i are increased by the same amount, so the only individuals who commit different acts are those who switch to act 0; at $r_j = 0$, these will be individuals with benefits of zero.) The second term is zero because, by (3.7), $p_j = p_0$. The third term is zero because no individuals switch to any acts except act 0. The fourth term is positive.

Next, we show that the optimal $r_i < \bar{r}$ increase with h_i by applying the argument used to demonstrate (3.10) -- that is, we compare dW/dr_i and dW/dr_j for $r_i = r_j$. Observe that, because the distributions of benefits for acts i and j are the same, when $r_i = r_j$ considerations of symmetry imply the following: $\alpha_i = \alpha_j$; $g_i(r_i) = g_j(r_j)$; $\gamma_{ik} = \gamma_{jk}$ for $k \neq i, j$; $\gamma_{ij} = \gamma_{ji}$; $B_{ik} = B_{jk}$ for $k \neq 0, i, j$; $B_{ij} = B_{ji}$; and $B_i = B_j$. Also, from (3.7), $p_i = p_j$. Thus, the difference in marginal welfare is

$$\text{(A.15)} \qquad \frac{dW}{dr_i} - \frac{dW}{dr_j} = \alpha_i(h_i - h_j)(1 + \gamma_{ij}),$$

which has the sign of $h_i - h_j$. Hence, the argument demonstrating (3.10) applies.

We can also show that only the most harmful acts will be subject to $\bar{r}$, by modifying the argument from section 3 in a manner analogous to that just employed.

To demonstrate that $\bar{r} > 0$, we can use the same argument as in section 3. This is because increasing all the r_i (other than r_0) by the same small amount from zero has the same effects as before and no others, as no individuals are led to switch among harmful acts.

Finally, the argument from section 3 establishing that any behavior that results without self-reporting can be induced at lower cost with self-reporting applies here without modification, for expected sanctions were held constant in that argument. This implies that self-reporting is superior.