Optimal Law Enforcement with Self-Reporting of Behavior

(Article begins on next page)

Optimal Law Enforcement with Self-Reporting of Behavior

Louis Kaplow and Steven Shavell

Harvard Law School and National Bureau of Economic Research

Self-reporting—the reporting by parties of their own behavior to an enforcement authority—is a commonly observed aspect of law enforcement, such 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, and we characterize the optimal scheme. Self-reporting offers two advantages over schemes without self-reporting: enforcement resources are saved because individuals who report their harmful acts need not be detected, and risk is reduced because individuals who report their behavior bear certain rather than uncertain sanctions.

I. 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 environmental and safety violations, individuals often report accidents they cause to the police, and even those who commit crimes sometimes confess their acts to the authorities. Presumably, parties voluntarily report their behavior because they fear more severe treatment if they do not.[1]

[1] Notably, under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), 42 U.S.C., sec. 9603(b), failure to report the release of hazardous substances may result in fines or imprisonment, apart from any penalty associated

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 (1920), 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 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 to report their harmful acts without materially affecting their incentives to refrain from committing the acts.[3] This can be accomplished by allowing those who report committing a harmful act to pay a sanction equal to (or slightly less than) the certainty equivalent of the sanction they would face if they did not report the act. Then individuals will be led to report their acts, but deterrence of acts 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

---

with the release itself. Under federal law and some state laws, corporations and individuals are required to report violations of product safety statutes and face penalties for not doing so (see, e.g., Consumer Product Safety Act, 15 U.S.C., sec. 2064). Federal criminal sentencing guidelines specify that organizations that do not report their offenses suffer harsher treatment than if they do. Federal and state laws also stipulate that individuals who do not report a variety of crimes and harmful acts suffer greater sanctions than those who do. For instance, hit-and-run statutes generally make a driver's failure to report an accident a crime.

[2] 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), Wagenhofer (1987), and Mookherjee and Png (1989). See also Townsend (1979), Baiman and Demski (1980), Baron and Besanko (1984), Reinganum and Wilde (1985), Dye (1986), Scotchmer (1987), and Mookherjee and Png (1990). We comment on this literature in n. 25. In the literature on law enforcement, the only article analyzing self-reporting is Malik (1993), which was written independently of our original working paper (Kaplow and Shavell 1991). See also Kaplow (1992), which incorporates self-reporting when analyzing other issues.

[3] For concreteness, we speak here of individuals who report acts that cause harm with certainty. But our analysis applies as well to situations in which individuals' actions result in harm only with a probability and what they report is the occurrence of harm. For example, a firm might take precautions to reduce the risk of discharge of a pollutant and report discharges when they occur.

when actors are risk-averse), because those who commit harmful acts report their behavior and pay a certain amount.[4] By contrast, under law enforcement systems without self-reporting, those who commit harmful acts bear the risk of sanctions.

In Section II, we analyze a model of self-reporting in which risk-neutral individuals choose whether to commit a single type of harmful act; in Section III, we examine a generalization of the model that allows for many types of harmful acts. In these models, we characterize optimal enforcement schemes with self-reporting and show that they are superior to schemes without self-reporting because the former allow enforcement costs to be reduced. We also demonstrate that a positive level of enforcement is always desirable with self-reporting even though it is not necessarily desirable without self-reporting. In Section IV, we consider risk aversion, imprisonment as a sanction, error in examination of behavior, and administrative costs of self-reporting. In Section V, we observe that, depending on the method of enforcement, the advantages of self-reporting may be greater than in our model or nonexistent, and we offer further concluding remarks.

II. The One-Act Model

Risk-neutral individuals choose whether or not to commit an act that causes a harm h. (Our analysis would be essentially the same if harm were probabilistic; see n. 12.) 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 $F(\cdot)$.[5] The size of the population is normalized to one.

We now define and analyze 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 accurately determines whether an individual committed the harmful act, and each examina-

---

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

[5] 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, 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.

tion costs c.[6] 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.[7] The 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 an asterisk.

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

$$W = \int_{ps}^{\infty} (b - h) f(b)\, db - pc. \tag{1}$$

The first term is the benefits minus the harm from commission of the act. The second term is the enforcement cost, because 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 $p^* > 0$.[8] Were $s^* < \bar{s}$, s could be raised and p lowered such that ps remained constant. Then the first term in (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 (1) with respect to p, using $s^* = \bar{s}$, to obtain

$$\frac{dW}{dp} = \bar{s}(h - p\bar{s}) f(p\bar{s}) - c. \tag{2}$$

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) is negative at $p = 1$. An interior solution for p^* must satisfy the first-order condition that $dW/dp = 0$.[9] In this case, the optimal probability is determined by

$$p^* = \frac{h - c/[\bar{s}f(p^*\bar{s})]}{\bar{s}}, \tag{3}$$

[6] Other methods of enforcement are discussed in Sec. V.

[7] 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 n. 21, our main results do not depend on this assumption.

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

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

and the optimal expected sanction is determined by

$$p^*\bar{s} = h - \frac{c}{\bar{s}f(p^*\bar{s})}. \tag{4}$$

The left side of (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.

In summary, we have the following proposition.

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 (3), and the optimal sanction is the maximum feasible sanction $\bar{s}$.

B. *Enforcement with Self-Reporting*

In the scheme with self-reporting, if an individual admits to committing a harmful act, he pays an ex ante sanction r ($r \leq \bar{s}$), and he is not examined.[10] 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 committed it. Individuals who do commit the act will report this if and only if $r \leq ps$.[11] 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 it, and welfare is as given in (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 it, and social welfare is

$$W = \int_r^\infty (b - h)f(b)\,db - pcF(r). \tag{5}$$

Expressions (5) and (1) differ in two respects. First, the lower limit of integration in (5) is r rather than ps, because individuals who commit the act report it and bear the certain sanction r rather than the expected sanction ps. Second, the examination cost in (5) is $pcF(r)$

---

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

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

rather than pc, 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 a lower enforcement 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 (1) and (5) are equal. But enforcement costs are lower with self-reporting by $[1 - F(ps)]pc$, because those who commit the act report this and are not examined. Thus, we have the following proposition.[12]

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 it, 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 it and pay r, so the integral in (5) would not change. But the reduction in p would reduce the second term, increasing welfare.[13] Second, the optimal ex post sanction is $\bar{s}$: as in the case without self-reporting, this sanction economizes enforcement resources.

To find the optimum, we may substitute $p\bar{s}$ for r in (5) and differentiate with respect to p to obtain

$$\frac{dW}{dp} = \bar{s}(h - p\bar{s})f(p\bar{s}) - pc\bar{s}f(p\bar{s}) - cF(p\bar{s}). \qquad (6)$$

---

[12] The arguments leading to this result would be virtually the same if harm occurred with probability π and individuals were required to report harm caused by their acts. Given that harm occurs, individuals would report this if and only if $r \le ps$, just as when harm occurs with certainty. They would decide to commit the act if and only if $b \ge \min(\pi r, \pi ps)$ rather than when $b \ge \min(r, ps)$. Thus, if $r \le ps$, (5) becomes

$$W = \int_r^\infty (b - h)f(b)\,db - pc[1 - \pi + \pi F(r)],$$

and proposition 2 follows. Similarly, the analysis and conclusions in the rest of our paper would be essentially unchanged if harm were assumed to be probabilistic.

[13] 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.

At $p = 0$, the derivative equals $\bar{s}hf(0)$, which is positive, so that p^* must be positive. By 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 subject to examination, 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 subject to examination; 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

$$p^* = \frac{h - cF(p^*\bar{s})/[\bar{s}f(p^*\bar{s})]}{\bar{s} + c}, \tag{7}$$

or

$$p^*\bar{s} = h - p^*c - \frac{cF(p^*\bar{s})}{\bar{s}f(p^*\bar{s})}. \tag{8}$$

Equation (8) is analogous to (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; and $cF(p^*\bar{s})/\bar{s}f(p^*\bar{s})$, 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 (8), as the optimal Pigovian tax for committing the harmful act, because this is the amount individuals pay with certainty when they commit the act (since 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—because of enforcement costs.[14]

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

[14] However, the optimal Pigovian tax may exceed the harm for some acts in the n-act model of Sec. III.

individuals from committing the harmful act: the optimal proba-
bility of examination is positive.

c)   The optimal probability p^* is given by equation (7), the optimal
     ex post sanction is the maximum feasible sanction $\bar{s}$, and the opti-
     mal 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 equivalent for $p = 0$ with self-reporting,
but $p = 0$ is not optimal with self-reporting by part b of proposition
3. Hence, we have the following proposition.

PROPOSITION 4. The optimal self-reporting scheme is superior to
the optimal scheme without self-reporting.

Finally, we compare the optimal probabilities of examination with
and without self-reporting, using equations (4) and (8). The optimal
probabilities generally differ because the costs of deterring the mar-
ginal individual differ. On one hand, this marginal enforcement cost
tends to be lower with self-reporting because an increase in the proba-
bility 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.[15]

III.   The n-Act Model

The model in Section II can be generalized as follows. There are n
harmful acts, and act i causes harm 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.[16] For convenience, not acting is sometimes referred to as commit-

---

[15] To illustrate, consider the case in which $f(\cdot)$ is uniform on $[0, 1]$. Subtracting the
right side of (3) from the right side of (7) yields

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

The numerator obviously can be positive or negative.

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

ting act 0. When the authority conducts an examination, it learns both whether a harmful act was committed and, if so, what type of act was committed. 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; and 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

$$W = \sum_{i=1}^{n} \theta_i \left[\int_{ps_i}^{\infty} (b - h_i) f(b)\, db - pc \right]. \tag{9}$$

Thus,

$$\frac{dW}{ds_i} = \theta_i p (h_i - ps_i) f(ps_i). \tag{10}$$

Assume that $p^* > 0$. It then follows from (10) 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 acts subject to this sanction are underdeterred. Let I denote the set of i for which $s_i^* = \bar{s}$. Then we have

$$\frac{dW}{dp} = \sum_{i \in I} \theta_i \bar{s}(h_i - p\bar{s}) f(p\bar{s}) - c. \tag{11}$$

Thus, as in the one-act model, the optimum involves $p^* = 0$ if c is sufficiently large. If p^* is positive, (11) is zero. In this case, the set I cannot be empty (it must include n), for otherwise (11) is negative. We now can state the following analogue of proposition 1.[17]

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 (11) equal to zero; 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 $\{0, 1, \ldots, n\}$. Individuals who report i pay an ex ante sanction

[17] This result appears in Shavell (1991).

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.[18] The social authority chooses an enforcement mechanism—a set of r_i, p_i, and s_{ij}—to maximize social welfare.

Because the revelation principle applies,[19] attention may be confined to mechanisms in which individuals who commit act i truthfully report i. We then can use the fact that the optimal scheme induces truth telling at minimum enforcement cost to establish three results (see the Appendix). First, the optimal ex post sanctions may be taken to involve the maximal penalty for lying and no penalty for telling the truth. That is,

$$s_{ij}^* = \bar{s} - r_i, \quad \text{for } i \neq j, \tag{12}$$

and

$$s_{ii}^* = 0. \tag{13}$$

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

$$r_0^* = 0. \tag{14}$$

Third, the probabilities of examination for a given set of r_i must obey

$$p_i = \frac{\bar{r} - r_i}{\bar{s} - r_i}, \tag{15}$$

where $\bar{r}$ denotes the highest of the r_i. It follows from (15) 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 as

$$W = \sum_{i=1}^{n} \theta_i \left(\int_{r_i}^{\infty} (b - h_i) f(b) \, db - c\{p_i[1 - F(r_i)] + p_0 F(r_i)\} \right). \tag{16}$$

Note that the second expression in large parentheses measures examination costs: the fraction $1 - F(r_i)$ of those in group i commit act i and report i, so they are examined with probability p_i; the remaining

---

[18] The analysis in the Appendix demonstrating that the 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 Border and Sobel (1987) did, 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.)

[19] We verify this in our discussion paper (Kaplow and Shavell 1991).

fraction of those in group i do not commit act i, so they report 0 and are examined with probability p_0.

We now prove the analogue 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 (12) and (13) and the p_i as in (15). The decision whether to commit acts will be the same as it was without self-reporting (because $r_i = ps_i$), so the integrals in (9) and (16) will be equal. To compare enforcement costs, observe from (15) 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, from (15), $p_i \leq p$ for all i, and $p_i < p$ for all $i > 0$ such that $s_i > 0$.[20] Thus, the enforcement cost term in (16) is strictly less than that in (9) (because only the deterred are examined with the highest probability rather than the entire population). This establishes the following proposition.

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 (16) over the r_i, where the p_i are determined by (15). For any $r_i < \bar{r}$, we have

$$\frac{dW}{dr_i} = \theta_i \left((h_i - r_i) f(r_i) - c \left\{ (p_0 - p_i) f(r_i) + [1 - F(r_i)] \frac{\bar{r} - \bar{s}}{(\bar{s} - r_i)^2} \right\} \right). \quad (17)$$

The first term in large parentheses in (17) is the direct social benefit from deterring the marginal individual in group i from committing act i: harm 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 [15] 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 [15], the optimal p_i falls as r_i rises).[21] Be-

---

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

[21] This follows because the numerator of the derivative of (15) with respect to r_i is $\bar{r} - \bar{s}$, which is negative because we assume that $\bar{s} \geq h_n$ and demonstrate below that $\bar{r}^* < h_n$. Without our assumption, the numerator could equal zero; i.e., $\bar{r}^* = \bar{s}$ is possible. (If c is sufficiently small and $\bar{s} < h_n$, $dW/d\bar{r}$ in [19] 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 [15]). In this case, $r_i^* = h_i$ for all $h_i < \bar{r}^*$. (In [17], $p_0 = p_i$ and $\bar{r} = \bar{s}$, so $dW/dr_i = \theta_i(h_i -$

cause these two components have opposite signs, $h_i - r_i$ may be positive or negative at an interior optimum when $dW/dr_i = 0$.[22] Thus, r_i^* may be such that there is either underdeterrence or overdeterrence relative to first-best behavior.

Using (17), 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 large parentheses is $h_i f(0)$, which is positive; the first component of the second term is zero (for, from [15], $p_i = p_0$ at $r_i = 0$); and the second component of the second term is positive. The explanation is that increasing r_i from zero has deterrence benefits and reduces the rate at which individuals of type i must be examined.

It also follows from (17) that, for any r_i^* and r_j^* (other than r_0^*) less than $\bar{r}^*$,

$$h_i > h_j \Rightarrow r_i^* > r_j^*. \tag{18}$$

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 I be the set of positive i such that $r_i = \bar{r}$. Then, varying $\bar{r}$ for all $i \in I$, we obtain[23]

$$\frac{dW}{d\bar{r}} = \sum_{i \in I} \theta_i \left\{ (h_i - \bar{r}) f(\bar{r}) - c \left[p_0 f(\bar{r}) + \frac{F(\bar{r})}{\bar{s}} \right] \right\}$$

$$- c \sum_{i \notin I} \theta_i \left[\frac{1 - F(r_i)}{\bar{s} - r_i} + \frac{F(r_i)}{\bar{s}} \right]. \tag{19}$$

The summation in (19) over $i \in I$ has an interpretation similar to that of (17), except for the last component: in (17), 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 (19), those who are deterred (the fraction $F(\bar{r})$

$r_i) f(r_i)$.) 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.

[22] Because the first component is a marginal effect and the second an inframarginal effect, the relationship of the two will depend on, among other things, 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.

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

$$\frac{dW}{d\bar{r}} = (h_1 - \bar{r}) f(\bar{r}) - c p_0 f(\bar{r}) - \frac{c}{\bar{s}} F(\bar{r}),$$

which is what is obtained in the one-act model by differentiating W in (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.)

of group i) must be examined at a higher rate, because p_0 increases as $\bar{r}$ increases. The summation over the $i \notin I$ is the cost of examining individuals not subject to $\bar{r}$ more frequently (from [15], all the p_i, including p_0, rise with r).

Observe that all the terms except the first component of the first summation, $(h_i - \bar{r})f(\bar{r})$, are negative. Thus, if $dW/d\bar{r} = 0$ at the optimum, the sum over $i \in I$ 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.[24]

Now let us show that some degree of enforcement is optimal; that is, $\bar{r}^* > 0$. Suppose that $\bar{r} = 0$, so that $I = \{1, \ldots, n\}$, and consider raising all the r_i (except r_0) uniformly. Evaluating the expression for marginal welfare (19) at zero yields

$$\frac{dW}{d\bar{r}} = \sum_{i=1}^{n} \theta_i h_i f(0) > 0. \tag{20}$$

Thus, $\bar{r}^* > 0$. From (15), this implies $p_0^* > 0$, so some enforcement effort is applied at the optimum. The reason is essentially that given for part b of proposition 3 in Section II: when the sanctions r_i for all the acts are raised 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.

Finally, 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.[25]

PROPOSITION 7. Under the optimal self-reporting scheme in the n-act model, the following conclusions 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

---

[24] For an overdeterred act, say act k, to be subject to $\bar{r}$ at the optimum, dW/dr_k in (17) must be positive when evaluated at $\bar{r}$. This possibility cannot be ruled out. Although overdeterrence implies that the first term in (17) is negative and (15) implies that the second term is negative ($p_k = 0$ at $r_k = r$), the third term is positive.

[25] In the mechanism design literature on auditing cited in n. 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 Sec. IV) is not contained in the auditing literature, since it does not address the optimal control of harmful externalities.

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, that is, $0 = r_0^* < r_1^* < \ldots < r_j^* = \ldots = r_n^* = \bar{r}^*$; $\bar{r}^*$ and the lesser r_i^* are determined by setting (19) and (17), respectively, equal to zero. (ii) The most harmful act is underdeterred relative to first-best behavior, that is, $r_n^* < h_n$; other acts may be underdeterred or overdeterred, although acts subject to the highest ex ante sanction 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 $\bar{r}$), that is, $p_0^* > p_1^* > \ldots > p_j^* = \ldots = p_n^* = 0$. The p_i^* are given by (15).

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 II. There we assumed that those who commit the harmful act (which is trivially the most harmful act) are not examined, those who do 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 note 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.[26] 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.[27] Elsewhere, we sketch an argument estab-

---

[26] Stigler (1970) first used 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.

[27] 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.

lishing 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 (see Kaplow and Shavell 1991).

IV. Extensions

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

A. Risk Aversion

Our main results 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) (see Kaplow and Shavell 1991). Achievable welfare is greater with self-reporting than without it, but now there are two reasons: enforcement costs are reduced and risk-bearing costs are eliminated. Risk is not borne by those who commit harmful acts because they are induced to report and thus pay a sanction r with certainty. Finally, because no risk is borne, the optimal ex post sanction is $\bar{s}$ (as it was in Sec. II). By contrast, without self-reporting the optimal enforcement scheme may change substantially when individuals are risk-averse; in particular, the optimal sanction may be less than maximal (see Polinsky and Shavell 1979).

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.[28]

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

---

[28] For a discussion of the optimal use of fines and imprisonment, see Polinsky and Shavell (1984).

is desirable for society to employ monetary sanctions to their limit $\bar{s}_1$ before resorting 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.

Suppose 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 it and are not examined. Now, however, there is the further advantage of reducing the use of imprisonment. Specifically, define 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 $r \le \bar{s}_1$, 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 argument discussed above may be expressed informally as follows. With self-reporting, the ex ante sanctions that are actually imposed on an individual are lower in magnitude by a factor of $1 - p$ 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 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. 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 $pq_1 s$ 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

$$b \geq ps(1 - q_0 - q_1). \tag{21}$$

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 $pq_1 s$. Thus, individuals will commit the harmful act if and only if (21) 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 (21) 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. The savings in enforcement costs under self-reporting, however, 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 do not commit the harmful act might mistakenly be deemed to have committed it, the imposition of ex post sanctions is not entirely avoided under self-reporting.[29] 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, notably by lowering sanctions. But the advantages of self-reporting

[29] Observe that all who are sanctioned for false reporting are indeed innocent, but the reason is that all the guilty are induced to admit their guilt initially.

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 in schemes without self-reporting, greater sanctioning costs must be incurred for 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 involve 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.[30]

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 (1) is reduced by $pd[1 - F(ps)]$, because those who commit the harmful act and are examined make payments. With self-reporting, the level of welfare previously given by (5) is reduced by $d[1 - F(r)]$, because those who commit the act make payments with certainty. In this case, the argument of proposition 2—that the same behavior can be induced with self-reporting as without it but at a 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 are now

$$[1 - F(r)][pc - (1 - p)d]. \tag{22}$$

---

[30] A related effect would arise if imprisonment were the only sanction and individuals were risk-averse with respect to this sanction. Then the ex ante sanction required to keep deterrence unchanged would exceed the expected amount of imprisonment imposed with ex post, probabilistic sanctions, making self-reporting more expensive. Our conjecture, however, is that individuals' preferences with respect to imprisonment are probably convex: the first year of a sentence imposes more disutility than an additional year.

The savings depend 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.[31]

V. Concluding Remarks

A. Methods of Enforcement and the Value of Self-Reporting

In the situations with which we were concerned, an enforcement authority learns both whether a harmful act occurred and who committed it through *examination* of an individual. A somewhat different situation arises when an enforcement authority knows at the outset about the occurrence of a harmful act and undertakes enforcement effort solely to determine who committed it. For example, the police may know that a crime has occurred, but not who committed it, before they proceed to determine who that person is. We shall refer to enforcement effort in such contexts as *investigation*.

When enforcement occurs by investigation, self-reporting does not merely reduce enforcement costs, it eliminates them: once someone confesses, others need not be investigated. (With examination, 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 confess equal to the expected ex post sanction. Thus, the reduction in the sanction for admitting one's act should be greater the lower the probability of apprehending the person through investi-

---

[31] 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).

gation would have been.[32] 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.

Another enforcement method is *monitoring*: the posting of enforcement agents to observe violations among any of a population, such 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 enforcement method, there may be no achievable cost savings 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 posted officers required to maintain the probability of apprehension for other drivers who might commit violations.

B.   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. There are three plausible explanations for this. First, an optimal mechanism may not be in place: the reduction in the sanction for admitting harmful acts may be too low to induce individuals to self-report in light of the probability of apprehension. Second, even if informed individuals would be induced to report their acts, some may underestimate the probability or magnitude of the sanction. Similarly, some individuals may not be aware of the nature of the acts that they have committed (see Kaplow 1992). (For example, a firm may not know that a chemical it uses is a regulated toxic substance.) Third, some individuals may know from their circumstances that detection is relatively unlikely.

C.   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 sanction costs are saved by inducing people to come forward with information about their con-

---

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

duct.[33] 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 (such as 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, 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. The incentives to report one's conduct frequently seem weak, because 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 remark that although we have focused on public enforcement of law, our discussion is more broadly relevant, notably to enforcement of incentive schemes in private contractual arrangements. For example, employers may have a policy of treating employees who admit to drug use or pilferage more favorably 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).

Appendix

Because the revelation principle applies (see Kaplow and Shavell 1991), we may assume that reports are truthful. Thus, the following incentive compatibility constraints, denoted by IC_{ij}, must hold for all i and j:

$$r_i + p_i s_{ii} \leq r_j + p_j s_{ji}. \tag{A1}$$

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 (12), which states that, for $i \neq j$, $s_{ij}^* = \bar{s} - r_i$ is an optimal sanction for lying. If (12) did not hold, one could alter the mechanism by raising s_{ij} to the point at which (12) does hold. Since IC_{ji} is satisfied under the original mechanism, it would be satisfied under the altered mechanism, because 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}, since

---

[33] Another reason self-reporting may be used in some contexts is to reduce harm: early identification of a toxic spill may facilitate mitigation of its effects.

all would report truthfully. Thus, social welfare under the altered mechanism would be the same as under the original one.

Next we show (13), that $s_{ii}^* = 0$. If (13) did not hold, one could alter the mechanism by lowering s_{ii} to zero 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 and the left side would be unaffected; thus, p_i could be lowered, which would reduce enforcement costs.[34] Finally, because the expected sanction for truthful reports 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 (A1) to be rewritten as[35]

$$r_i - r_j \leq p_j(\bar{s} - r_j). \tag{A2}$$

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 (A2), we can establish (14), 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 (12). This change would not affect the left side of IC_{ij} in (A2) 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_j, so enforcement costs could be reduced. Finally, it is apparent 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; 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 (A2): 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 (A2) will continue to hold for all i. Also, this reduction in

---

[34] 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.

[35] Note that (A2) holds trivially if $i = j$.

p_j does not alter individuals' choices of acts (because the r_i are the same). But reducing p_j saves enforcement costs, so welfare is higher. Because the constraints (A2) are binding when $r_i = \bar{r}$, the optimal probabilities are given by (15).

Finally, we demonstrate the relationships between the h_i and the optimal r_i. First, we prove (18), 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 W is. In particular, $(1/\theta_i)W$ is maximized with respect to r_i at r_i^* and $(1/\theta_j)W$ is maximized with respect to r_j at r_j^*. Now, from (17), 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.[36] 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 (17), so the argument demonstrating (18) establishes that $r_i^* > r_j^*$, a contradiction. Now consider the case in which only act j is subject to $\bar{r}^*$. Using expression (19), observe that, at $\bar{r}^*$,

$$\theta_j[(h_j - r_j^*)f(r_j^*) - cp_0 f(r_j^*)] > 0, \tag{A3}$$

because all the other terms in (19) are negative. This implies that

$$\theta_n[(h_n - r_n^*)f(r_n^*) - cp_0 f(r_n^*)] > 0, \tag{A4}$$

because $h_n > h_j$ and $r_n^* < r_j^*$. But then, from (17), $dW/dr_n > 0$ at r_n^*, because (A4) is the first two components of (17) and all the other components are also positive. This contradicts the optimality of r_n^*.

References

Baiman, Stanley, and Demski, Joel S. "Economically Optimal Performance Evaluation and Control Systems." *J. Accounting Res.* 18 (suppl., 1980): 184–220.

Baron, David P., and Besanko, David. "Regulation, Asymmetric Information, and Auditing." *Rand J. Econ.* 15 (Winter 1984): 447–70.

Becker, Gary S. "Crime and Punishment: An Economic Approach." *J.P.E.* 76 (March/April 1968): 169–217.

Border, Kim C., and Sobel, Joel. "Samurai Accountant: A Theory of Auditing and Plunder." *Rev. Econ. Studies* 54 (October 1987): 525–40.

Dye, Ronald A. "Optimal Monitoring Policies in Agencies." *Rand J. Econ.* 17 (Autumn 1986): 339–50.

Kaplow, Louis. "A Model of the Optimal Complexity of Rules." Working Paper no. 3958. Cambridge, Mass.: NBER, January 1992.

---

[36] This result relies on our assumption that $f(\cdot)$ is independent of the type of act. Otherwise, (18) might not hold, because inframarginal effects on examination costs could be greater for the less harmful act.

Kaplow, Louis, and Shavell, Steven. "Optimal Law Enforcement with Self-Reporting of Behavior." Discussion Paper no. 95. Cambridge, Mass.: Harvard Law School, Program Law and Econ., August 1991.

Malik, Arun S. "Self-Reporting and the Design of Policies for Regulating Stochastic Pollution." *J. Environmental Econ. and Management* 24 (May 1993): 241–57.

Mookherjee, Dilip, and Png, Ivan. "Optimal Auditing, Insurance, and Redistribution." *Q.J.E.* 104 (May 1989): 399–415.

———. "Enforcement Costs and the Optimal Progressivity of Income Taxes." *J. Law, Econ., and Organization* 6 (Fall 1990): 410–31.

Pigou, Arthur C. *The Economics of Welfare*. London: Macmillan, 1920.

Polinsky, A. Mitchell, and Shavell, Steven. "The Optimal Tradeoff between the Probability and Magnitude of Fines." *A.E.R.* 69 (December 1979): 880–91.

———. "The Optimal Use of Fines and Imprisonment." *J. Public Econ.* 24 (June 1984): 89–99.

Reinganum, Jennifer F., and Wilde, Louis L. "Income Tax Compliance in a Principal-Agent Framework." *J. Public Econ.* 26 (February 1985): 1–18.

Scotchmer, Suzanne. "Audit Classes and Tax Enforcement Policy." *A.E.R. Papers and Proc.* 77 (May 1987): 229–33.

Shavell, Steven. "Specific versus General Enforcement of Law." *J.P.E.* 99 (October 1991): 1088–1108.

Stigler, George J. "The Optimum Enforcement of Laws." *J.P.E.* 78 (May/June 1970): 526–36.

Townsend, Robert M. "Optimal Contracts and Competitive Markets with Costly State Verification." *J. Econ. Theory* 21 (October 1979): 265–93.

Wagenhofer, Alfred. "Investigation Strategies with Costly Perfect Information." In *Agency Theory, Information, and Incentives*, edited by Günter Bamberg and Klaus Spremann. Berlin: Springer-Verlag, 1987.