Monitoring vis-à-vis Investigation in Enforcement of Law

By Dilip Mookherjee and I. P. L. Png*

Enforcement by monitoring cannot be conditioned on the severity of an offense while enforcement by investigation can be. If some degrees of the offense are not adequately reported or if investigation is too costly, the regulator must monitor and treat offenses of different severity quite differently. Smaller offenses should not be investigated; they should be deterred by monitoring alone, coupled with graduated fines. To deter larger offenses, the regulator should vary the investigation rate while setting maximal fines. (*JEL* D82, K42, L51)

The Singapore government recently published regulations to control noise from construction sites. Commenting on enforcement of the new regulations, the Environment Minister remarked that, initially, his ministry would inspect noise levels at building sites, but he hoped that, with time, the government could rely on complaints from the public to identify excessively noisy work sites.[1]

We give the name *monitoring* to enforcement activity in which the regulator must commit resources before receiving information about the offense, if any. By contrast, in enforcement by what we call *investigation*, the regulator can condition the resources committed on information (e.g., from victims) about the severity of the offense. This is the key difference between the two enforcement mechanisms.[2]

Assuming that both monitoring and investigation are feasible, how *should* an environmental regulator allocate resources between them? Similar issues arise in tax compliance and internal organization. A revenue collector could condition audit policies on taxpayers' reports or, alternatively, inspect taxpayers without regard for information from taxpayers. To regulate employees' effort, an employer could hire supervisors to monitor the workers or, instead, rely on reports from dissatisfied customers.[3]

We postulate that the principal (whether law enforcer, regulator, or employer) aims to minimize the cost of enforcing a given schedule of expected fines, which specifies an expected fine for each degree of the offense. We show that investigation allows fines to be used to the maximum at all levels of the offense, whereas enforcement by monitoring must reserve the largest fine for the most serious offense. Since fines merely transfer wealth, investigation is endogenously more cost-effective than monitoring. In addition, however, the efficient mix of monitoring and investigation must also de-

*Mookherjee: Indian Statistical Institute, 7 S.J.S. Sansanwal Marg, New Delhi 110016, India; Png: Anderson Graduate School of Management, University of California, Los Angles, CA 90024-1481. We gratefully acknowledge thoughtful advice from Y. N. Chen, F. Easterbrook, L. Kornhauser, M. Polinsky, S. Shavell, the referees, and seminar participants at UC Berkeley, UCLA, the University of Southern California, and the American Economic Association 1990 meetings.

[1]See "New Rules To Curb Noise from Building Sites," *The Straits Times Weekly*, Overseas Edition, 22 December 1990, p. 6. In the Los Angeles area, drivers are encouraged to report vehicles emitting excessive pollution to the South Coast Air Quality Management District through a toll-free telephone number.

[2]To monitor, the authorities may need to spend resources even before the offender acts, while the decision to investigate follows the crime. Such a temporal distinction echoes prior work by Donald Wittman (1977), Steven Shavell (1984), and Charles D. Kolstad et al. (1990) that compared direct regulation of offenders' actions with liability rules as alternative ways to control externalities.

[3]For instance, Johnson & Johnson dental floss carries a toll-free telephone number for customer complaints. Many hotels distribute postage-paid forms with which guests may comment on the quality of service.

pend on the direct costs of the two enforcement methods.

We show that, if the direct cost of investigation is sufficiently low and the offense is adequately reported at all levels, then efficient enforcement involves only investigation. For example, Orley Ashenfelter and Robert Smith (1979) measured about a 65-percent rate of compliance with the U.S. minimum wage in 1973, although it was enforced by only 880 personnel. This could well be explained by the incentive of workers paid less than the minimum wage to report offending employers (see John R. Lott, Jr., and Russell D. Roberts, 1989).

If, on the other hand, some degrees of the offense are not reported at a sufficient rate, or if investigation is too costly, then efficient enforcement involves monitoring as well. Since monitoring applies to all levels of the offense, it is efficient to divide offenses into two ranges. Reports of occurrences in the lower range should not be investigated at all; these small offenses are deterred through monitoring alone, coupled with graduated fines. Reports of offenses in the upper range must be investigated at positive rates, while fines, whether enforced by investigation or monitoring, should be set at the maximum. The principal varies the investigation rate to deter these larger offenses.

Our analysis also helps to resolve a puzzle in the economics of law enforcement. Since fines are transfer payments, while enforcement is costly, Gary Becker (1968) argued that society should set all fines at the maximum possible.[4] George J. Stigler (1970), however, contended that more serious offenses must be punished more severely; otherwise, minor offenders would switch to bigger crimes. However, Richard A. Posner (1986 p. 208 [footnote 2]) observed that "even if all crimes were punished with the same severity, some marginal deterrence could be preserved by varying the probability of punishment with the gravity of the crime... ."

In this paper, we show that marginal deterrence requires graduated fines only if more serious offenses cannot be punished with higher probability. Louis Wilde (1989), C. Y. Cyrus Chu (1990), and Steven Shavell (1992) obtain related results, as does Shavell (1991) in a different setting.[5] Posner's argument and the formal analyses of Chu, Shavell (1991, 1992), and Wilde, however, leave open the important normative issue: should criminals be deterred by graduating fines or enforcement rates?

In the key contribution of our paper, we show how the trade-off between two enforcement technologies—one that *requires* graduated fines (monitoring) and another that *allows* graduated enforcement rates (investigation)—depends on their relative cost and the rates at which offenses are reported. We also discuss the effect of changes in the maximum permissible fine on the efficient mix of investigation and monitoring.[6]

I. The Model

Let $s \in [0, \bar{s}]$ represent the severity of the noise created by a construction site. The cost to the authorities of monitoring the site 24 hours each day is c_M. Assume that the cost of monitoring (randomly) with probability μ is $c_M \mu$. By the very nature of monitoring, the authorities cannot condition the rate μ on the severity of the offense, s.[7] Let

[4]Becker formalized Jeremy Bentham's (1931 p. 338) intuition: "Pecuniary punishments are highly economical, since all the evil felt by him who pays turns into an advantage for him who receives."

[5]In Shavell (1991), each potential offender chooses between no offense and committing the offense at a (different) exogenous level.

[6]Our results also apply to the employment relation. To deter shirking, an employer could require employees to post bonds (Edward Lazear, 1981). William T. Dickens et al. (1989), however, argue that contract law and social norms limit the size of bonds that an employer may seize. By contrast, our results imply that, if an employer enforces by monitoring, penalties should rise with the degree of shirking; hence, the employer often *should* not confiscate the entire bond.

[7]As emphasized by Wilde (1989) and Shavell (1991, 1992), the monitoring cost is a joint cost across the various offense levels.

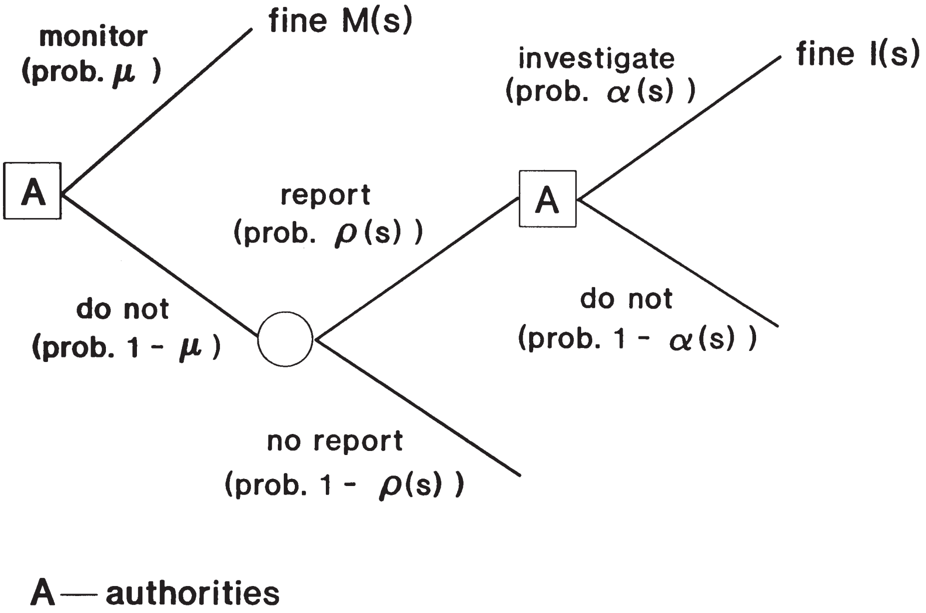

FIGURE 1. STRUCTURE OF ENFORCEMENT

the authorities levy a fine $M(s)$, according to the noise level, on any builder apprehended through monitoring.

We suppose that if the builder creates noise s, a victim will (truthfully) report the offense s to the authorities with exogenous probability $\rho(s)$. To ensure that investigation is always feasible, we assume the following.

ASSUMPTION 1:[8]

$$\rho(s) > 0 \qquad \textit{for all } s > 0.$$

Suppose that c_I is the cost of investigating the report to the extent sufficient to apprehend the culprit. We assume that this cost does not vary with the severity of the offense or the corresponding fine and that the cost of investigating at rate $\alpha(s)$ is $c_I \alpha(s)$. Let the fine enforced by investigation be $I(s)$.

Accordingly, an enforcement policy consists of a monitoring rate μ, investigation rates $\{\alpha(s)\}$, and fines enforced by monitoring and investigation, respectively, $\{M(s), I(s)\}$. We assume that both monitoring and investigation produce information without error.[9]

An individual who commits the offense at level s will be monitored and fined $M(s)$ with probability μ. With probability $(1-\mu)$, he will not be monitored, but then with probability $\rho(s)\alpha(s)$, he will be reported, investigated, and fined $I(s)$ (see Fig. 1).[10] Hence, in expectation, he will be fined a combined amount of

$$(1) \quad E(s) \equiv \mu M(s) + (1-\mu)\rho(s)\alpha(s)I(s).$$

[8]This assumption may be difficult to satisfy for (i) offenses, such as atmospheric pollution, that are difficult to detect or that spread harm among many victims (in the latter case, a free-rider problem in reporting arises), and (ii) so-called victimless crimes such as bribery, prostitution, and sale of illegal drugs.

[9]Like Becker (1968), Stigler (1970), and others, we implicitly assume that the authorities can set the enforcement rate independently of the fines. See James Andreoni (1991) for a critique of this premise.

[10]Figure 1 shows a report being possible only if the offender is not detected through monitoring. Of course, a victim might report an offender who has already been so detected. This report, however, would not add to the authorities' information; hence, it would not affect the efficient enforcement strategy.

Potential offenders may differ in their private benefit from crime. We assume that all are risk-neutral and that each may commit at most one offense. Accordingly, each will choose the level, s, of the offense that maximizes the difference between his private benefit and the expected fine, $E(s)$.[11] Let the rate of offenses at level s be $\lambda(s)$.

Generally, the socially optimal schedule of fines $E(s)$ and rate of offenses $\lambda(s)$ balance the benefits of crime against the harm caused and the enforcement cost.[12] Since neither the harm caused nor the enforcement cost C depends on the enforcement method per se (monitoring vis-à-vis investigation),[13] the socially optimal policy must minimize C relative to the optimal choices of $E(s)$ and $\lambda(s)$.

Accordingly, in focusing here on how to minimize the enforcement cost, it is legitimate to treat $E(s)$ and $\lambda(s)$ as exogenous. The conditions that we derive below are necessary for a policy to maximize welfare; if a policy violates these conditions, it will be dominated by another that enforces the same $E(s)$ and $\lambda(s)$ at lower cost.

To simplify exposition, we confine attention to settings that satisfy the following.

ASSUMPTION 2: $\lambda(s)>0$, *all* s, *and is continuous.* $E(s)$ *is continuous and increases strictly with* s;[14] $E(0)=0$.

For simplicity, all potential offenders have identical wealth Y. To enable the authorities to deter all crime completely through fines if they choose to do so, we assume that the offenders' wealth exceeds the required expected fine at every level of the offense.

ASSUMPTION 3: $Y>E(s)$, *all* s.

Alternatively, Y may be interpreted to be the maximum permissible fine.

II. Efficient Enforcement Policy

From the authorities' standpoint, an enforcement policy $(\mu,\{\alpha(s), M(s), I(s)\})$ will involve expenditure of $c_M\mu$ on monitoring, and $(1-\mu)c_I\int_0^{\bar{s}}\rho(s)\alpha(s)\lambda(s)ds$ on investigation. We postulate that the authorities' problem is to choose a policy to minimize the total cost,

$$(2)\quad C\equiv c_M\mu + (1-\mu)c_I\int_0^{\bar{s}}\rho(s)\alpha(s)\lambda(s)ds$$

of enforcing by monitoring and investigation a schedule of expected fines,

$$(1)\quad \mu M(s)+(1-\mu)\rho(s)\alpha(s)I(s) = E(s)$$

subject to the following constraints:

$$(3)\quad 0\le\mu\le 1$$

$$(4)\quad 0\le\alpha(s)\le 1$$

$$(5)\quad 0\le M(s)\le Y$$

$$(6)\quad 0\le I(s)\le Y$$

for all s.[15] Any policy that meets (1) and (3)–(6) is said to be *feasible*.

Consider levels of the offense at which the authorities investigate reports at a positive rate, $\alpha(s)>0$. Suppose that, at one such level, the accompanying fine is less than maximal, $I(s)<Y$. Then, following the logic of Becker (1968) and R. A. Carr-Hill and N. H. Stern (1979), the law-enforcers may reduce $\alpha(s)$ and simultaneously raise $I(s)$ without affecting the expected fine from investigation conditional on a re-

[11] A risk-averse offender's choice will depend on the fines separately enforced by monitoring and investigation, and not merely on the combined expected fine.

[12] Society, of course, may attach different weights to the benefits, the harm, and the enforcement cost.

[13] Essentially because potential offenders are risk-neutral.

[14] If the schedule $E(s)$ were flat over some interval $[s_1,s_2]$, offenses at level s_2 would dominate all offenses within the interval, violating the assumption that $\lambda(s)>0$ for all s.

[15] Since this problem consists of maximizing a continuous function over a compact set of variables, it has a solution.

port, $\alpha(s)I(s)$. This reduces the enforcement cost, C, while preserving constraints (1) and (3)–(5). Hence, it is efficient to set $I(s)=Y$.[16]

For levels of the offense at which the authorities do not investigate reports, $\alpha(s)=0$, the value of $I(s)$ is irrelevant and, without loss of generality, may be set as $I(s)=Y$. This establishes the following proposition.

PROPOSITION 1: *Fines enforced by investigation should be maximal, that is, $I(s)=Y$ for all s.*

Substituting in (1), the requirement on the expected fine simplifies to

$$E(s)=\mu M(s)+(1-\mu)\rho(s)\alpha(s)Y. \tag{7}$$

Stigler (1970) contended that marginal deterrence *requires* fines to increase with the size of the offense. Proposition 1 shows that Stigler's argument applies only if the enforcement rate cannot be conditioned on the level of the crime, as in the case of monitoring. We next turn to consider how fines should be set in this case.

Suppose that the authorities do monitor, so that $\mu>0$. An increase in the fine, $M(s)$, will not affect enforcement costs, while a cut in the investigation rate, $\alpha(s)$, will reduce costs. Accordingly, whenever they may do so while preserving the expected fine, $E(s)$, the authorities should raise $M(s)$ and reduce $\alpha(s)$.

Two constraints, however, limit such an adjustment. First, the rate of investigation cannot be reduced below zero; that is, $\alpha(s)\geq 0$, which by (7), implies $\mu M(s)\leq E(s)$. Secondly, no individual may be fined more than the maximum permissible, $M(s)\leq Y$. Thus, the fines enforced by monitoring should be set as large as possible, subject to these two constraints.

LEMMA 1: *Suppose that the efficient enforcement policy involves some monitoring, $\mu>0$. Then, fines enforced by monitoring, $\{M(s)\}$, should be set according to*

$$M(s)=\min\left[Y,\frac{1}{\mu}E(s)\right]. \tag{8}$$

Whenever $E(s)/\mu<Y$, fines enforced by monitoring will fall short of Y. Since monitoring cannot be targeted to specific levels of the offense, marginal deterrence through monitoring requires fines to rise with s. By contrast, investigation allows maximal use of fines at all s (Proposition 1); hence, investigation is (endogenously) more cost-effective than monitoring in enforcing marginal deterrence. We next characterize the fines enforced by monitoring and rates of investigation as follows.

PROPOSITION 2: *Suppose that the efficient enforcement policy involves some monitoring, $\mu>0$. Then, there exists offense level s^* such that*

$$\mu=\frac{E(s^*)}{Y} \tag{9}$$

and the efficient policy will comprise: (a) for offenses $s\leq s^*$, investigation rate $\alpha(s)=0$, and fine enforced by monitoring,

$$M(s)=\frac{1}{\mu}E(s)<Y \tag{10}$$

while (b) for offenses $s>s^*$, investigation rate

$$\alpha(s)=\frac{E(s)-\mu Y}{(1-\mu)\rho(s)Y}>0 \tag{11}$$

and fine $M(s)=Y$, for all s.[17]

Proposition 2 shows that, if the authorities monitor at all, the efficient enforcement

[16]See, however, Frank Easterbrook (1983), Arun S. Malik (1990), and Andreoni (1991) for limits on the extent to which the authorities may and should substitute larger sanctions for lower enforcement rates.

[17]Proofs of this and subsequent propositions are presented in the Appendix.

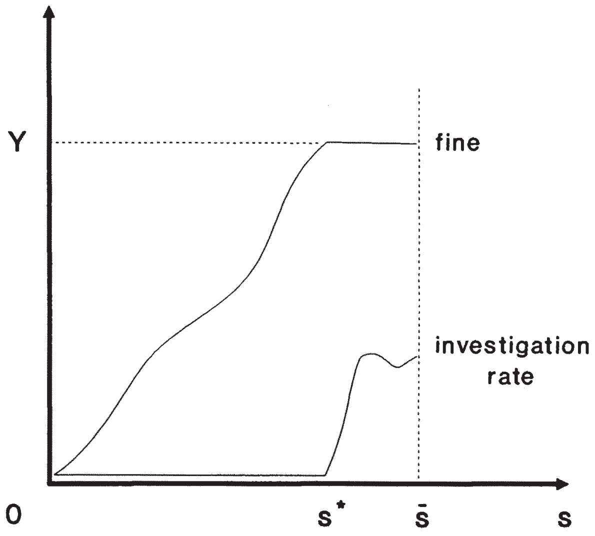

FIGURE 2. EFFICIENT ENFORCEMENT POLICY WHEN MONITORING RATE $\mu > 0$

policy varies with the severity of the offense. Enforcement by monitoring resembles a public good in the sense that it applies equally to offenses of all severities. Over a range of less serious offenses, monitoring alone coupled with strictly increasing fines suffices to provide the required marginal deterrence. The authorities should not investigate reports of these small offenses, because to do so would involve inefficient expenditure of additional resources.

Offenses more serious than s^* should be treated very differently. Since monitoring alone does not effect the required marginal deterrence, the efficient enforcement policy must combine investigation with monitoring. Therefore, reports of all occurrences within this range should be investigated with positive probability, graduated according to the severity of the offense. Over this range, fines—whether enforced by monitoring or investigation—should be set at the maximum.[18] Figure 2 illustrates the main results of Proposition 2.

We posed a setting that allowed fines to depend on the method of enforcement; that is, we did not constrain $I(s) = M(s)$. Proposition 2, however, shows that it is not efficient to vary the penalty by the enforcement method. Over the range $s \leq s^*$, reports are not investigated; hence $I(s)$ is irrelevant, and the only relevant fine is $M(s) > 0$. The authorities do investigate larger offenses ($s > s^*$) at a positive rate, but then $I(s) = M(s) = Y$. These normative results conform with the actual legal practice of setting fines without regard for the enforcement method.

Proposition 2 yields another interesting implication. From the viewpoint of a potential offender, the likelihood of being investigated is the probability of the offense being reported, multiplied by the investigation rate,

$$\rho(s)\alpha(s) = \frac{E(s) - \mu Y}{(1-\mu)Y}$$

by (11), which rises with severity, s. If $\rho(s)$ increases sufficiently faster than $E(s)$ in s, then the rate of investigation $\alpha(s)$ should *fall* with severity, s.

Rape, for instance, may be considered to be a lower degree of battery than murder. Many rape victims are reluctant to report offenses. If the reporting rate is sufficiently lower for rape than murder, then the authorities should spend more resources on investigating rape than murder. Likewise, in the case of environmental regulation, offenses that cause minor widespread harm are seldom reported and hence may require more investigation than serious violations.

We now turn to the trade-off between monitoring and investigation, and in particular, to a consideration of when the principal *should* choose $\mu > 0$. There are two cases to consider.

The Singapore Environment Minister suggested that affected residents could identify excessively noisy building sites more cheaply than government inspectors. Similarly, a hotel manager seeking to encourage courtesy among her staff may find it more convenient to rely on reports from unhappy customers than to monitor courtesy herself. In the specific context of our model, investigation is less costly than monitoring in a direct sense when the following condition is

[18] Note that we have not ruled out either $s^* = 0$ or $s^* = \bar{s}$.

satisfied:

$$(12) \qquad c_{\mathrm{M}} \geq c_{\mathrm{I}}.$$

Since, in addition, investigation allows full use of fines, whereas monitoring does not (Proposition 1 and Lemma 1), enforcement costs would surely be minimized if all monitoring were replaced by investigation. Investigation, however, will suffice to effect the required marginal deterrence only if *all* levels of the offense are adequately reported. For instance, if the Air Quality Management District in Los Angeles relies solely on reports from civic-minded motorists to identify motor vehicles that emit excessive pollution, it would not be able to deter invisible pollution. When reporting is insufficient at some level of the offense, the regulator must also invest resources in monitoring.

By contrast, for a manufacturer seeking to control production-line effort, it is probably cheaper to hire supervisors than to identify shirkers by investigating reports of defective products from dissatisfied customers. Suppose that the manufacturer does not monitor. Then, reports at all levels must be investigated at a positive rate, and all fines must be set to the maximum. Since monitoring is cheaper, however, the manufacturer could reduce enforcement costs by replacing some investigation with a slight increase in monitoring. These arguments motivate the following proposition.

PROPOSITION 3: *The efficient enforcement policy involves monitoring at a positive rate, $\mu > 0$, if and only if either (a) reporting is inadequate at some level,*

$$(13) \qquad \rho(s)Y < E(s)$$

for some s, or (b) the direct costs of investigation are sufficiently high, $c_{\mathrm{M}} < c_{\mathrm{I}}$.

Suppose that it is relatively cheaper to enforce pollution regulations by investigation and that invisible pollution corresponds to some $s < s^*$ (i.e., causes relatively less harm). As argued above, if reporting of invisible pollution is inadequate, the regulator must also use monitoring. Once the regulator does monitor, however, he ironically will not investigate reports of invisible pollution —because monitoring alone provides sufficient deterrence. Further, the enforcement policy will involve investigation with certainty at some level, that is, $\alpha(s) = 1$ for some s (for otherwise, the regulator could reduce enforcement costs by switching some resources from monitoring toward investigation).

Faced with insufficient reporting, the pollution regulator has two other alternatives. He could try to encourage reporting and, in this case, should focus "whistle-blower" programs on those offenses for which reporting is weakest relative to the requisite deterrence, that is, s at which (13) holds. Secondly, if the inadequately reported level of the offense is not very harmful, society might be better off legalizing it. This will permit the regulator to reduce enforcement and substitute investigation for monitoring.

If the maximum permissible fine, Y, is higher, the regulator should raise fines. Again, this will not only allow the regulator to reduce enforcement, but will also enable him to switch toward a more cost-effective method of enforcement (from monitoring to investigation). Thus, the larger the maximum permissible fine, the closer will be the efficient enforcement policy to Becker's (1968) prescription.[19] Note, however, that we assumed constant returns to both investigation and monitoring. In reality, raising either enforcement rate from 10 percent to 20 percent is probably much less costly than enhancing it from 90 percent to 100 percent. Accordingly, the tendency toward maximal penalties for all offenses will be attenuated.

III. Concluding Remarks

Our model could be extended to consider nonpecuniary penalties. Imprisonment and other nonpecuniary penalties, however, are not pure transfers. Hence, even in enforcement by investigation, imprisonment should

[19]See Mookherjee and Png (1990 section 4) for comparative-statics results in detail. There, we show, in addition, how the efficient mix of investigation and monitoring varies with changes in the enforcement costs and reporting rates.

not be maximal at all levels of the offense. Rather, the regulator must strike a balance between graduating the enforcement rate vis-à-vis the penalty.

We have asked how a regulator should combine graduated fines (effected through monitoring) with graduated enforcement rates (effected through investigation) to deter noise or chemical pollution at a particular site. Environmental regulators face a second problem: how to allocate enforcement resources across potential sources of pollution. To the extent that the regulator knows that some parties are inherently more likely to emit toxic pollutants and to pollute in larger volumes than others (e.g., chemical plants vis-à-vis bakeries), allocation of expenditures on monitoring can be conditioned on this information.[20]

Suppose that the maximum pollution from the chemical plant exceeds the maximum pollution from the bakery. Then, applying the Becker (1968) logic, enforcement costs would be reduced if the regulator could fine the bakery more than the chemical plant for the same amount of pollution. If the regulator, however, is constrained to set identical fines across polluters for the same offense, our results suggest that he should inspect chemical plants more frequently than bakeries.

We have ignored several other factors that are relevant when designing enforcement schemes. In a setting where offenses do not vary in severity, A. Mitchell Polinsky and Shavell (1979) have shown that maximal penalties are not optimal if offenders are risk-averse. Patrick Bolton (1985) and Mookherjee (1992) show that, under certain circumstances, the same is true if mistakes occur in enforcement. Errors could arise from false reports by victims or third parties. For instance, since government investigations can be very onerous, a firm may drag down a competitor by falsely accusing it of violating some law.[21] Generally, reporting of offenses will depend on the authorities' enforcement policy. In particular, the more likely the authorities are to investigate, the more frequently victims will report.[22] Future work should consider the incentives of victims and other related parties to report offenses and how the authorities can influence the rate and accuracy of such reports.

APPENDIX

PROOF OF PROPOSITION 2:

(a) Let s^* be the largest s such that $\alpha(s)=0$. Substituting in (7), $E(s^*)=\mu M(s^*)$, and by (8), $M(s^*)=E(s^*)/\mu \leq Y$. Since $E(s)$ strictly increases with s, for all $s<s^*$, $E(s)/\mu < E(s^*)/\mu \leq Y$; hence, by (8), $M(s)=E(s)/\mu$, and by (7), $\alpha(s)=0$. (b) Similarly, if $s>s^*$, then $E(s)/\mu > Y$; hence, by (8), $M(s)=Y$. Substituting in (7), we have (11).

LEMMA 2: *Suppose that $c_M \geq c_I$. Then, if $\mu>0$ and $M(s)<Y$ on a set of positive measure, an efficient enforcement policy will have $\alpha(s)=1$, for at least one s.*

PROOF:

Suppose otherwise, that $c_M \geq c_I$, $\mu>0$, and $M(s)<Y$ on a set of positive measure, but $\alpha(s)<1$ for all s. Consider a variation to reduce μ and raise the $\alpha(s)$ slightly while preserving (7), that is,

$$0=[M(s)-\rho(s)\alpha(s)Y]\Delta\mu + (1-\mu)\rho(s)Y\Delta\alpha(s)$$

or

$$(1-\mu)\rho(s)\Delta\alpha(s) = \left[-\frac{M(s)}{Y}+\rho(s)\alpha(s)\right]\Delta\mu.$$

[20] We thank a reviewer for suggesting this issue.

[21] Also, Suzanne Scotchmer (pers. comm.) has suggested that victims will report in a way so as to maximize the likelihood of investigation and hence the probability of recovering their property.

[22] In the National Crime Survey for 1987, 8.1 percent of all households who did not report a larceny smaller than \$50 selected "police would not want to be bothered" as one reason; this rate was 11.9 percent for failures to report larcenies exceeding \$50 (U.S. Department of Justice, 1989 table 103).

For $\Delta\mu < 0$ sufficiently close to zero, the variation will satisfy (3) and (4).

Substituting in (2), the variation will change the cost of enforcement by

$$\Delta C = \left[c_M - c_I \int_0^{\bar{s}} \rho(s)\alpha(s)\lambda(s)\,ds\right]\Delta\mu$$

$$+ (1-\mu)c_I \int_0^{\bar{s}} \rho(s)\Delta\alpha(s)\lambda(s)\,ds$$

$$= \left[c_M - c_I \int_0^{\bar{s}} \frac{M(s)}{Y}\lambda(s)\,ds\right]\Delta\mu.$$

Now, $M(s) < Y$ on a set of positive measure; hence,

$$c_M - c_I \int_0^{\bar{s}} \frac{M(s)}{Y}\lambda(s)\,ds > c_M - c_I \geq 0$$

since $c_M \geq c_I$. Since $\Delta\mu < 0$, we conclude that $\Delta C < 0$, which is a contradiction.

PROOF OF PROPOSITION 3:

Sufficiency.—(a) Suppose that $\rho(s)Y < E(s)$ at some s. Then, even if $\alpha(s) = 1$, $\alpha(s)\rho(s)Y < E(s)$; hence, the authorities must also monitor, $\mu > 0$, to meet the expected fine constraint, (7). (b) Given $c_M < c_I$, suppose otherwise that $\mu = 0$. Then by (7), $E(s) = \rho(s)\alpha(s)Y$, for all s; hence, $\alpha(0) = 0$ and $\alpha(s) > 0$, for all $s > 0$. Since $\mu = 0$, we may set $M(0) = 0$ and $M(s) = Y$, for all $s > 0$ without loss of generality. Consider a variation to raise μ and reduce $\alpha(s)$ for all $s > 0$, while preserving (7), that is,

$$0 = [M(s) - \rho(s)\alpha(s)Y]\Delta\mu$$

$$+ (1-\mu)\rho(s)Y\Delta\alpha(s).$$

For sufficiently small $\Delta\mu$, the variation will meet (3), and it also meets (4) for $s > 0$. The increase in μ will not affect the expected fine at $s = 0$ since $M(0) = \alpha(0) = 0$. Thus, the variation is feasible.

Following the argument in the proof of Lemma 2, by (2), the variation will change the enforcement cost by

$$\Delta C = \left[c_M - c_I \int_0^{\bar{s}} \frac{M(s)}{Y}\lambda(s)\,ds\right]\Delta\mu$$

$$= (c_M - c_I)\Delta\mu$$

since $M(s) = Y$, for all $s > 0$. Now $c_M < c_I$ and $\Delta\mu > 0$; hence, we conclude that $\Delta C < 0$, which is a contradiction.

Necessity.—Suppose that $c_M \geq c_I$ and

$$\text{(A1)} \qquad \rho(s)Y \geq E(s)$$

for all s. Then, a policy with $\mu = 0$ and $\alpha(s) = E(s)/\rho(s)Y \leq 1$ would meet (7) and be feasible. We next show that it is not efficient to monitor. Suppose otherwise, that $\mu > 0$.

Case (i): $M(s) < Y$ on a set of positive measure. Then, by Lemma 2, $\alpha(s) = 1$, for some s. We will show that $M(s) < Y$ at any s where $\alpha(s) = 1$. Suppose otherwise, $M(s) = Y$; then,

$$\mu M(s) + (1-\mu)\rho(s)\alpha(s)Y$$

$$= \mu Y + (1-\mu)\rho(s)Y$$

$$> \mu E(s) + (1-\mu)E(s) = E(s)$$

by Assumption 3 and (A1). Since this contradicts (7), we conclude that if $\alpha(s) = 1$, then $M(s) < Y$. Hence for all such s, we may raise $M(s)$ and reduce $\alpha(s)$ and thereby lower the cost of enforcement while maintaining the expected fine.

Case (ii): $M(s) = Y$ almost everywhere. Then, by (7),

$$E(s) = \mu Y + (1-\mu)\rho(s)\alpha(s)Y \geq \mu Y > 0$$

almost everywhere, which is a contradiction, since $E(s)$ is continuous at $s = 0$ and $E(0) = 0$.

REFERENCES

Andreoni, James, "Reasonable Doubt and the Optimal Magnitude of Fines: Should the

Penalty Fit the Crime?" *Rand Journal of Economics*, Autumn 1991, *22*, 385–95.

Ashenfelter, Orley and Smith, Robert, "Compliance with the Minimum Wage Law," *Journal of Political Economy*, April 1979, *87*, 333–50.

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

Bentham, Jeremy, *Theory of Legislation*, New York: Harcourt Brace, 1931.

Bolton, Patrick, "Random Inspection and Optimal Penalties in the Principal–Agent Relationship," Ph.D. dissertation, University of London, 1985.

Carr-Hill, R. A. and Stern, N. H., *Crime, the Police, and Criminal Statistics*, London: Academic Press, 1979.

Chu, C. Y. Cyrus, "Crime and Punishment with Possible Post-offense Compromise," *Academia Economic Papers*, 1990, *17* (2), 131–67.

Dickens, William T., Katz, Lawrence F., Lang, Kevin and Summers, Lawrence H., "Employee Crime and the Monitoring Puzzle," *Journal of Labor Economics*, July 1989, *7*, 331–47.

Easterbrook, Frank, "Criminal Procedure as a Market System," *Journal of Legal Studies*, June 1983, *12*, 289–332.

Kolstad, Charles D., Ulen, Thomas S. and Johnson, Gary V., "*Ex Post* Liability for Harm vs. *Ex Ante* Safety Regulation: Substitutes or Complements?" *American Economic Review*, September 1990, *80*, 888–901.

Lazear, Edward, "Agency, Earnings Profiles, Productivity, and Hours Restrictions," *American Economic Review*, September 1981, *71*, 606–20.

Lott, John R., Jr. and Roberts, Russell D., "Why Comply? One-Sided Enforcement of Price Controls and Victimless Crime Laws," *Journal of Legal Studies*, June 1989, *18*, 403–14.

Malik, Arun S., "Avoidance, Screening and Optimum Enforcement," *Rand Journal of Economics*, Autumn 1990, *21*, 341–53.

Mookherjee, Dilip, "The Economics of Enforcement," in M. K. Rakshit, eds., *Essays in Honour of Tapas Majumdar*, New Delhi: Oxford University Press, 1992 (forthcoming).

Mookherjee, Dilip and Png, I. P. L., "Monitoring versus Investigation in Law Enforcement and Regulation," Business Economics working paper, UCLA, May 1990.

Polinsky, A. Mitchell and Shavell, Steven, "The Optimal Trade-off Between the Probability and Magnitude of Fines," *American Economic Review*, December 1979, *69*, 880–91.

Posner, Richard A., *Economic Analysis of Law*, 3rd Ed., Boston: Little, Brown, 1986.

Shavell, Steven, "Liability for Harm Versus Regulation of Safety," *Journal of Legal Studies*, June 1984, *13*, 357–78.

______, "Specific versus General Enforcement of Law," *Journal of Political Economy*, October 1991, *99*, 1088–1108.

______, "A Note on Marginal Deterrence," *International Review of Law and Economics*, 1992 (forthcoming).

Stigler, George J., "The Optimum Enforcement of Laws," *Journal of Political Economy*, May/June 1970, *78*, 526–36.

Wilde, Louis, "Criminal Choice, Nonmonetary Sanctions and Marginal Deterrence: A Normative Analysis," unpublished manuscript, California Institute of Technology, May 1989.

Wittman, Donald, "Prior Regulation versus Post Liability: The Choice Between Input and Output Monitoring," *Journal of Legal Studies*, January 1977, *6*, 193–211.

The Straits Times Weekly, Overseas Edition, 22 December 1990.

United States Department of Justice, *Criminal Victimization in the United States, 1987*, NCJ-115524, Washington, DC: Bureau of Justice Statistics, June 1989.