JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT 11, 327–346 (1984)

Enforcement Costs and Regulatory Reform: The Agency and Firm Response[1]

STEPHEN H. LINDER[2]

Center for Public Policy Studies, Tulane University, New Orleans, Louisiana 70118

AND

MARK E. MCBRIDE

Department of Economics, Miami University, Oxford, Ohio 45056

Received June 7, 1982; revised February 1983 and July 1983

Influenced by models of optimal law enforcement, several authors have recently revised the work on efficient levels of regulatory control to accommodate the realities of underenforcement and imperfect compliance. However, most of these efforts have centered on either the enforcement agency or the firm and have largely ignored the decentralized nature of the enforcement process. This paper extends these results by modelling both the firm and the local agency and by incorporating detection uncertainty and concealment activity. Each model is then evaluated with respect to the alternative regulatory regimes of direct controls and emission taxes.

Since the early work on society's optimal level of law enforcement [1, 10] a number of scholars have attempted to recast familiar models of social regulation to accommodate the realities of enforcement costs and imperfect compliance. By and large, attention has focused upon the kinds of adjustments made by various actors once enforcement difficulties have been raised. McKean [7], for example, considers how the analyst might revise his or her evaluation of regulatory initiatives in light of potential enforcement requirements. Shifting to the relative effectiveness of enforcement schemes, Polinsky [8] ties the concern with setting an appropriate fine to the problem of choosing the best enforcer. Other work deals not so much with the problems and feasibility of enforcement as with the implications of including enforcement and compliance costs in determining an optimal level of regulatory control.

The implications of imperfect enforcement given the most attention are those connected with the behavior of the firm. Downing and Watson [3] investigate how the control actions of the firm respond both to several levels of control and to different enforcement schemes. Going one step further, Harford [4] expands on the nature of the firm's reaction to imperfect enforcement in order to predict levels of noncompliance. Taking a similar perspective, Viscusi and Zeckhauser [11] evaluate the significance of expected noncompliance for arriving at an optimal level of

[1]A preliminary version of this paper was presented by the authors at the Association for Public Policy Analysis and Management Meetings, October 1981. The authors wish to acknowledge the helpful comments of an anonymous reviewer who prompted the clarification of several important points.

[2]Author to whom all correspondence should be sent. Current address: School of Public Health, The University of Texas, Health Science Center at Houston, Houston, Texas 77225.

0095-0696/84 $3.00

control. Despite these recent refinements in our understanding of social regulation, several important issues remain.

First, there has been no attempt to integrate the work on the enforcement difficulties faced by government regulators with an examination of the firm's behavior. Clearly, the choice of an enforcement scheme and how it is implemented should affect the firm's compliance decision. Following Polinsky, the enforcement agent's objectives are likely to differ from those defining social welfare. Furthermore, we can expect enforcement by local authorities to differ from that typically attributed to the Federal Government. In a decentralized regime then, as public choice theorists are quick to point out, each local enforcement agency will maximize its own objective function, seldom reaching a social optimum and seldom coinciding with the objectives of higher authorities.

What the firm responds to, then, is a mixed set of enforcement signals generated by a *decentralized* and often fragmented system of enforcement. Under this system, the firm's compliance calculus is likely to be more complex than generally assumed in previous work relying on simpler, unified schemes of enforcement. Not only will expected costs expand, but the number of government-chosen parameters influencing firm behavior will increase as well. Secondly, as McKean points out, there have been few attempts to incorporate enforcement issues into the evaluation of alternative regulatory regimes. Our intent is to offer a comparative assessment of direct control and tax regimes in the context of decentralized enforcement, tracing the implications of each regime for the behavior of both the local enforcer and the firm.

THE CENTRAL ENFORCEMENT AGENCY

Most of the approaches which examine the firm's response to controls under imperfect enforcement begin by respecifying the social objective function to include the social costs of enforcement. The conventional rule that pollution should be eliminated only to the point where marginal damages are equal to the marginal costs of pollution control no longer holds. Instead, socially optimal control is found to occur at the point where marginal damages equal marginal control costs plus the marginal costs of enforcing these controls. This provides the central control agency with information on optimal enforcement, as well as the level of the optimal standard. However, the social objective function should go further than the objectives attributed to the agency; it should account for the social gains from pollution relative to its social costs. Nevertheless, assuming that the central agency specifies some definitive standard, the interesting question is whether taxes or direct controls are the most appropriate means for reaching it.

If we consider only marginal treatment costs, then a tax regime clearly outperforms direct controls. But the choice becomes less one-sided when relative damage and enforcement costs are considered. Under imperfect compliance, Polinsky argues that the costs of damages given a private, incentive-based enforcement scheme (not unlike those proposed for tax regimes) exceed similar costs under relatively more expensive, public enforcement. Likewise, Harford points out that the marginal social costs of enforcement will also be relatively higher under taxes than under a direct control regime. Thus, it is plausible that the least-cost control advantages of a tax on waste discharges may be outweighed by a corresponding increase in marginal damage and enforcement costs. In short, the relative advantages of a taxing regime may have been overstated, once the impact of each regime on agency and firm behavior is scrutinized. We will return to this proposition once our modelling of the firm and agency objectives is complete.

Our first major assumption is that the central agency attempts to influence firms' behavior by setting a standard w^* for desired levels of pollution control, binding under either direct control or tax regimes, and by setting an expected penalty for noncompliance to encourage firms to attain w^*. While the central agency chooses some probability of detecting firms in noncompliance, the local agency revises it and consequently defines the threshold for subsequent action against violators. However, because of statutory and resource limitations, the local agency has no discretion over the level of fines or the likelihoods of litigation and conviction established by the central agency. Like the firm, the local agency must view them as fixed. The next section examines the local agency's behavior under these premises.

THE LOCAL ENFORCEMENT AGENCY

The Local Context

The local enforcement agency with jurisdiction over one of several control areas within a particular state is subject to several political cross-pressures which shape its incentives. While the central agency sets the ultimate control objectives, the state supervises the implementation of any control program and directly influences the level of enforcement by rationing the local agency's resources. As part of the Federal enforcement effort, and as enforcement professionals, local authorities are typically sympathetic to the central agency's objectives, and yet as state employees they must be sensitive to the state's commitment to control. Moreover, as line personnel, local authorities deal with firms on a day-to-day basis and have a substantial stake in maintaining smooth working relations. They depend on data from the firms to satisfy the central agency's reporting requirements and on the firms' sufferance so as not to antagonize state officials concerned with the area's business climate. Since most of these local agencies are under-funded given their operating responsibilities, they are also forced to rely on industry cooperation in achieving control targets. Negotiation is generally preferred to formal sanction as a means of settling compliance problems. Without state or central agency intervention, local authorities have neither the resources nor the expertise to offer litigation as a credible deterrent to noncompliance.

Too aggressive an enforcement effort may undermine state support and industry cooperation; on the other hand, too lenient an effort may bring sanctions from the central agency and fail to induce any firms to comply with control requirements. With the mix of loyalties and pressures facing local authorities, determining an appropriate level of enforcement is a difficult task. Adding to the difficulty, however, is the problem of uncertainty. All enforcement actions are based upon the determination of whether a violation has occurred; in fact, much of the local agency's time is taken up with surveillance and monitoring to find violators, rather than with convincing violators to comply. Assuming that the definition of a violation is clear and can be applied to individual cases, a problem in many states, local authorities must detect violations from very limited and often ambiguous information. Even if the amount of information were increased, there are limits to how much of it could be usefully integrated into the agency's detection decision. And with uncertain information comes the chance of errors, some more costly to the agency than others.

The agency, then, can be viewed as a fallible observer, evaluating information on noncompliance and observations of firm behavior in order to detect any violations that might occur. However, because of the uncertainty involved, the decision over whether or not a violation is indicated will prove to be wrong at least part of the time. Thus, there are four possible outcomes to the agency's detection decision: (1) allege a violation when, in fact, there is one ("hit"); (2) find no violation, although one is occurring ("miss"); (3) allege a violation, but none occur ("false alarm"); (4) find no violation, when there is none ("correct rejection").

In effect, the agency is choosing between two statistical hypotheses on the basis of limited information with the outcome determined by which of two states of nature, violation or no violation, applies. This, of course, assumes that the state or central agency has some means of establishing the true state of nature in order to verify the local enforcer's allegations. "Misses," for example, may be brought to the state's attention by the media or concerned citizens, while "false alarms" are likely to be vigorously objected to by firms, wrongly accused. Detection errors then can presumably be uncovered through subsequent investigations or formal proceedings motivated by those bearing the costs of improper enforcement. We assume that most errors are eventually brought to light, even if none of the actors has perfect knowledge of the state of nature, since the capacity of the state to validate the local enforcer's choice of the true state of nature is a prerequisite to any state-imposed penalties for factual error.

The two possible errors, "misses" and "false alarms," are analogous to Type I and Type II statistical errors, respectively. For the agency, this implies that efforts to reduce "misses" by favoring the violation hypothesis, also increases the chances of a "false alarm." Conversely, efforts to diminish the frequency of "false alarms" by being more conservative in alleging violations and favoring the nonviolation hypothesis is likely to result in more "misses." The optimal behavior for the agency in this context is suggested by statistical decision theory [6]. One should fashion a decision rule which balances these outcomes, based upon their relative value to the agency and the efficient use of available information.

Modeling Optimal Behavior

To model the agency's optimal behavior, we will first consider the simple stochastic properties of the information bearing on violations, then introduce certain biases particular to each agency, and finally consider the agency's choice of a decision rule.

We represent the information on the firm's compliance status as a random variable x drawn from one distribution in the case of violations and from a second when no violation has occurred. The value of this variable will affect the local enforcer's confidence about the true state of the firm's compliance. Accordingly, we assume that a separate probability density is associated with x under each of the two alternative hypotheses, $f_v(x)$ for violations and $f_{nv}(x)$ for nonviolation. The ratio of these two densities then expresses the relative likelihood $l(x) = f_v(x)/f_{nv}(x)$ that the information represents a violation rather than no violation. Because there are only two alternative hypotheses under consideration, the likelihood ratio will summarize information into a single, unidimensional index, regardless of the dimensionality of that information. We can assume that our compliance information variable has k dimensions, each dimension corresponding to a separate source of evidence on the firm's compliance. For example, sources might include private citizen and media observations, the firm's own reporting, data from monitoring equipment, on-site inspections, and more cursory surveillance. The dimensions may vary in salience, and some may take on a zero value when no information is available from particular sources. Nonetheless, the local enforcer must integrate this information, by some unspecified process, in order to fashion a likelihood ratio.

Following a practice in signal detection theory [5], we can assume that both probability densities can be mapped on a unidimensional axis. Since the presence of a violation simply adds a certain increment of information to the data arising from nonviolation, we can consider the density function for violations as a translation of the function for nonviolations and assume that both are of equal variance. Further, by restricting the distribution of x to be Gaussian under each hypothesis, we can guarantee that x will be monotonic with the likelihood ratio. This assumption permits us to treat the unidimensional axis of x values as an ordinal scale of the likelihood ratio. The local enforcer then can choose a criterion x_c and a corresponding likelihood ratio $l(x_c)$ that will optimize her detection decisions. Thus, she will choose the nonviolation hypothesis whenever $x < x_c$ and the violation hypothesis whenever $x_c < x$. This unidimensional construction is illustrated in Fig. 1.

The chances of error mentioned earlier are indicated by the area of overlap between the distributions of Fig. 1. As the distributions move closer together, discrimination between them becomes more difficult and the chances of "misses" and "false alarms" increase. Conversely, as the distributions move apart, the chances of errors diminish and discrimination becomes more accurate. In the context of testing statistical hypotheses, this distance is directly related to sample size; as the size increases, discrimination improves. But for our purposes, neither the number of observations nor their dimensionality matter nearly as much as their quality. The more sensitive the agency's detection network, the further apart the distributions will appear and the smaller the likelihood of detection error. Of course in the instance of a substantial violation the sensitivity of the monitor becomes far less important for detection. Accordingly, detectability d^0, represented by the distance between the centers of the two distributions, will be taken as a product of the relative magnitude of violations, as well as of the sensitivity of the observer.

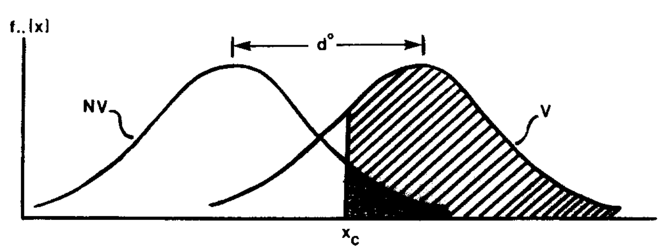

FIG. 1. The probability density functions of nonviolations (NV) and violations (V).

The Optimal Criterion

A decision rule for the agency in this context establishes the conditions for identifying a given observation as representing a violation. Referring back to Fig. 1, the agency selects a criterion point from the range of observations which serves as a threshold. Any observation exceeding this point provides support for the violation hypothesis. In other words, any observation x_i falling to the right of x_c in Fig. 1 suggests the presence of a violation. Alternatively, observations falling to the left of x_c support the assumption of no violation. From the perspective of statistical theory, x_c identifies the critical region. Observations falling in this region serve as evidence for the violation hypothesis. The next question is how the agency goes about choosing a particular criterion.

As noted earlier, the agency can allege a violation or find no violation, and in either case may be in error. The likelihood of correctly identifying a violation, a "hit," depends upon the joint occurrence of an observation falling within the critical region and the agency choosing to support the violation hypothesis. For a "miss," the choice of the violation hypothesis occurs jointly with an observation falling outside the critical region, and so on. The interesting feature of this construction is that, for fixed distributions like those in Fig. 1 and a given a priori probability of a violation, the agency's selection of a criterion value x_c completely determines the probabilities of all four possible decision outcomes. Likewise, for any balance among the outcome probabilities desired by the agency, there is a corresponding optimal location for the criterion. For example, an agency might desire a balance that maximizes the number of correct decisions. If both "hits" and "correct rejections" are of equal value to the agency and if both kinds of errors are also considered of equal value, the optimal location of the criterion value will be at the intersection of the two distributions in Fig. 1. But as the values of any of these outcomes change, the optimal criterion value will fall to the left or right of this intersection. Choice of the optimal criterion, then, will be affected not only by the relative likelihood of a violation, but also by the values the agency attaches to each decision outcome.

The Agency Objective Function

Consider how the detection goal might differ across local enforcement agencies. Just as there are areas with special climatic problems where the enforcement practices mandated by the state seem surprisingly lax, there are also areas whose economic and political climates make these practices appear unreasonably onerous. While in the former instance the value of a "hit" and the cost of a "miss" may be especially high, in the latter instance avoiding "false alarms" and reinforcing the cooperative efforts of compliers and near-compliers may be the predominant concern. To accommodate these different priorities within a single model, we will postulate a well-known rule [2] for decision making under uncertainty which captures local enforcement behavior: The local agency maximizes the expected value of its detection decision.

Using the joint probabilities of choosing a particular hypothesis and being correct or in error to weigh the values of the four possible outcomes, the agency's objective function appears as

$$\text{max expected value} = p(D \& V)B_{D,V} + p(ND \& NV)B_{ND,NV} - p(ND \& V)C_{ND,V} - p(D \& NV)C_{D,NV}, \quad (1)$$

where B signifies a benefit and C, a cost. The D stands for detection, or alleging a violation, and V for a violation, with the prefix N designating no detection (ND) or no violation (NV). The p elements represent the joint probabilities of the events indicated within the braces. Note that without a decision rule the agency would be forced to evaluate the benefits and costs of every decision. Whether to allege a violation, for example, would depend upon the expected value of that decision relative to the value of not making an allegation:

$$p(D \& V)B_{D,V} - p(D \& NV)C_{D,NV} \lesseqgtr p(ND \& NV)B_{ND,NV} - p(ND \& V)C_{ND,V}. \quad (2)$$

The use of a criterion, of course, economizes on both evaluation efforts and information requirements.

To simplify matters, we can restate the probabilities of joint occurrences used in these two expressions as the product of an a priori probability and a conditional probability. For example, $p(D \& V)$ can be replaced by $p(V) \cdot p(D|V)$, where $p(V)$ is the a priori probability that a violation will occur and $p(D|V)$ is the probability that the agency will allege a violation given that one has actually occurred. The other important conditional is $p(D|NV)$, the probability that the agency will allege a violation when, in fact, no violation has taken place. Borrowing our earlier designations, these two conditionals are the "hit" and "false alarm" rates for the agency and are sufficient to capture agency performance, since the probabilities of the other two outcomes, "miss" and "correct rejection," are their complements. Substituting these conditionals and a priori probabilities into expression (1), maximizing the expected value is equivalent to maximizing:

$$p(D|V) - bp(D|NV) \quad (3)$$

where

$$b = \frac{p(NV)}{p(V)} \cdot \frac{(B_{ND,NV} + C_{D,NV})}{(B_{D,V} + C_{ND,V})}. \quad (4)$$

Further, it can be shown [5] that b in expression (4) designates the value of the optimal criterion. In order to maximize the expected value of its decisions, then, the agency must locate its criterion x_c at the point within the set of observations in Fig. 1 which corresponds to the value of b. The optimal location of x_c can now be seen to depend solely upon the a priori probability representing the agency's expectations about the chances of violations before any observation is made, and the relative values assigned to the decision outcomes.

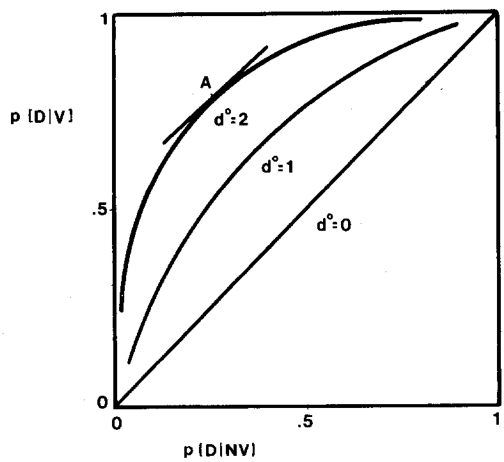

FIG. 2. Curves showing the "false alarm" rate versus the "hit" rate for several levels of detectability (d^0).

As expression (3) illustrates, the value of b offers a compromise between missed detections and false ones based upon agency values and expectations. More specifically, b specifies the agency's optimal weighting of "hits" relative to "false alarms." The trade-off between "hits" and "false alarms" can be expressed more directly if we examine the possible combinations of probabilities on a unit square with $p(D|V)$ on the ordinate axis and $p(D|NV)$ on the abscissa. This plot shown in Fig. 2 is simply another way of representing the information in Fig. 1 for a range of criterion values. It should be pointed out that $p(D|V)$ and $p(D|NV)$ correspond to the areas under the respective V and NV curves of Fig. 1 which fall to the right of the criterion x_c.

The curves in Fig. 2 represent the trade-offs between $p(D|V)$ and $p(D|NV)$ as x_c ranges over its possible values. The shape and location of these curves are determined by the parameters of the density functions displayed in Fig. 1. As x_c moves from the far left to the right in Fig. 1, we move from the upper right toward the lower left along the curves in Fig. 2. The problem of choosing a criterion value now becomes a problem of locating the particular point which represents the optimal weighting of "misses" and "false alarms." Not surprisingly, the optimal point A appears at the tangency between the tradeoff curve and the linear expression given by expression (3). Once again b appears, this time as the slope of the tangent at the optimum. Even if b happens to change, the appropriate "hit" and "false alarm" probabilities will continue to appear at the point of tangency.

The Impact of Changing Parameter Values

Unlike earlier theories of optimal enforcement [9, 10] our theory is premised on a decentralized system of enforcement characterized by wide disparities in how the states implement Federal goals and adaptive behavior on the part of local agencies. When given discretion by the state, the local agency responds not by raising the probability of detection to the highest level its resources will sustain (or even to the level desired by the central agency) but by trading off its likely error and success rates to utilize the available information in the most effective way. These trade-offs provide an effective solution to the problems of balancing technical and political feasibility. The significance of such a strategy becomes more apparent as we turn to the implications of changing the various components of optimal agency behavior and examine how these alter the character of local enforcement.

Consider the baseline case mentioned earlier with "even" a priori odds of a violation and equally valued outcomes. The criterion value b in this instance will equal one; accordingly, x_c will be positioned at the intersection of the curves in Fig. 1, where the two density functions assume the same value. This particular criterion is known as the "minimax" and is optimal for minimizing errors. Assume for the moment that the parameters of the density functions remain fixed. The a priori odds express the agency's expectations about the incidence of violations. Any change in the a priori odds then signifies a change in the level of evidence required to confirm or deny a violation. If a particular jurisdiction has had few violations in the past or if noncompliers seem to be making a "good faith" effort to comply, the odds are likely to favor the no-violation hypothesis.

For a given set of outcome values, if the odds are against a violation, the value of b will increase and x_c will shift to the right reflecting the agency's reluctance to allege a violation in all but the most extreme cases. On the other hand, if the odds favor a violation, the value of b will decrease, shifting x_c to the left. In this instance, the agency classifies even ambiguous cases as violations since it expects noncompliance to be the rule rather than the exception. Although the a priori odds are shaped by the experiences of the agency, they will also reflect the agency's relations with firms in their jurisdiction. Maintaining good relations with the agency may assure the firm a "benefit of the doubt" in detection decisions.

Alternatively, for a fixed level of a priori odds, the detection of violations may increase in importance because of pressure from the central agency or emergency conditions, effectively raising the cost of every "miss" or the value of every "hit." Referring back to expression (4), such a change increases the size of the denominator of the term on the right, lowering the value of b. The position of x_c in Fig. 1 shifts to the left, increasing the proportion of "hits" relative to "misses." Conversely, if the importance of detection becomes secondary to the cultivation of good industry relations, the value of acknowledging compliance and avoiding unsubstantiated allegations increases. These changes affect the numerator of the term on the right side of expression (4), raising the value of b and shifting the optimal criterion in Fig. 1 to the right. In effect, the agency becomes less predisposed to allege a violation for any particular observation of firm activity. This bias toward restraint can result from the fear of increased pressure on agency resources generated by firms' unwillingness to cooperate voluntarily or by their inclination to contest allegations through costly formal proceedings. Pressure can also be applied by firms indirectly by complaining to state authorities. If the state favors the industry position, the local agency can be punished by cutting its resources, restraining its authority, or curtailing its salaries and promotions.

Finally, for a given set of agency values and expectations, the distance d^0 between the distributions in Fig. 1 (the area under each curve in Fig. 2) might be altered. As noted earlier, increasing this distance improves the agency's ability to discriminate violations and nonviolations. For a given "hit" rate, for example, as the distance between distributions increases, the "false alarm" rate must decrease. In Fig. 2, the curve would be displaced upward and to the left for any increase in this distance, simultaneously reducing both "miss" and "false alarm" rates. Generally, any measures heightening the surveillance capability of the agency will reduce error rates and consequently improve the accuracy of detection decisions. Of course, greater accuracy can also be a product of an increase in the severity of violations. Thus, the quality of detection is likely to differ from jurisdiction to jurisdiction not only as the result of agency biases and the state's investment in monitoring technology, but also due to disparities in the scope and magnitude of noncompliance.

For the firm committed to avoiding detection, our model suggests that a strategy based on changing the local agency's expectations, the values of decision outcomes, or the detectability of violations, can be successful in lowering the agency's "hit" rate. On the other hand, it is always possible for the agency to respond to such overt gaming behavior by moving to a higher likelihood of detection, despite the costs of "false alarms." Unfortunately, addressing these interactions requires a model of the participants' reaction functions in what amounts to a noncooperative game. We will leave these questions to a separate paper.

Changes under a Tax Regime

The shift from a direct control to an emission tax regime is an external source of change in the local agency's decision parameters which, like the intervention of particular actors, will alter its optimal decision rule. Nonetheless, the local agency will continue to choose a rule based upon its goal of maximizing the expected value of its decisions, and the basic character of violations will remain the same.

Under either regime the difference between the firm's level of reported wastes and the central agency's ultimate control standard w^* must be accounted for in some way. For a tax regime, any difference must be covered by the level of fees t the firm pays. Under direct controls, differences must be authorized by a schedule of compliance specifying a required level of control w_s for any given period. Both w_s and t are imposed on the local agency from above and are subject to periodic adjustment to assure that w^* is eventually attained. Once a firm is bound by a particular w_s, any excess of its reported wastes w_r over the level mandated could be considered a violation. Violations of this sort, however, are typically resolved either by convincing the local agency to accept a larger w_r based on intended progress toward w_s or by appealing to the central agency for a variance to raise w_s. Consequently, the only form of violation causing the agency a detection problem in either regime occurs when the firm's actual level of wastes exceeds the level it reports.

As with violations, the burden of detection facing the local agency will appear the same regardless of the regime. The agency is responsible for detecting departures from authorized levels of emissions, whether the authorization takes the form of a tax receipt or control requirement. Under either regime, monitoring for continuous compliance is a desirable practice; and the more accurate the monitoring, the better able the agency is to detect violations. However, the two regimes differ substantially in the *flexibility* each permits the local agency and the firm.

Under direct controls the local agency is able to barter flexibility over the firm's w_r for an accommodation of its error costs. Such cooperative arrangements can also reduce the agency's surveillance and data burdens, while providing an opportunity to persuade firms to comply. Although the firm's cost of relying on negotiation to set an acceptable w_r may be quite high, the cost of appealing the level of w_s to the central agency is typically much higher. Conversely, with a tax regime the firm no longer needs to negotiate with the local agency over w_r but can report a level of its own choosing. However, the consequent elimination of negotiation costs and gain in flexibility for the firm is accompanied by a reduction in flexibility for the local agency. The agency loses bargaining leverage with the firm, since the difference between w_r and w_s is now irrelevant, and is forced to replace its negotiated settlements with even-handed, rigid treatment. As a result, the agency must not only sacrifice its opportunity for persuasion and mutual accommodation, but also shoulder the full costs of its detection errors.

The shift to a tax regime is likely to affect the agency's optimal decision rule in several ways. First, the a priori odds of a violation are more likely to be close to even, largely because the agency is in a less secure position in predicting the incidence of violations. Without information on the firms' control technology or on their plans for production changes, the behavior of violators may become more erratic from the agency's perspective. Secondly, although one might expect the value of a "hit" to increase for the agency given the importance of accurate fee assessments, this is probably not the crucial change. Unless the central agency can discover underassessments and sanction the local agency for them, "misses" will not be nearly as important as "false alarms." Once we eliminate cooperative relations, an overassessment due to a "false alarm" is likely to lock the agency and the alleged violator in a costly proceeding to resolve the dispute. Of course, the firm may refuse to pay altogether, thereby undermining the credibility of the agency as assessor.

If, as we argue, the "false alarm" generates a higher expected cost than a "miss," the optimal criterion will assume a higher value under a tax regime, all other things being equal. This will shift x_c to the right in Fig. 1 and result in a relatively lower "hit" rate and a higher frequency of undetected violations. In the aggregate, these violations may reverse progress toward mandated standards. However, efforts to artificially lower the "false alarm" costs to the agency may induce firms to systematically underreport emissions as a compensatory measure against overassessment. Nevertheless, in some jurisdictions, norms of cooperation and informal modes of conflict resolution may restrain the costs of "false alarms" and nurture a "hit" rate that approaches the one under direct controls. While some agencies may do almost as well at detecting violations under a tax regime as they did under direct controls, many will likely do much worse.

Under our model of agency behavior, the only way to raise the "hit" rate for a given rate of "false alarms" is to increase the quality of discrimination between violations and nonviolations through more sensitive monitoring devices. There is no need to assume that monitoring emissions is any more or less difficult under one regime than under the other; nor is there a need to impose differences in the required frequency of inspections between the two regimes. Monitoring under a tax regime will require a more sophisticated, and thus a more expensive, detection system to achieve at least the same level of detection as under direct controls. This is not necessarily because detection becomes a greater burden, but because of the local agency's response to higher error costs.

The following section of the paper presents a simple model of a profit maximizing firm under the two alternative regimes of direct controls and emission taxes. Firm behavior is analyzed with respect to levels of output, waste emission, and concealment efforts for a range of potential changes in the parameters controlled by the government.

THE FIRM'S RESPONSE TO ENFORCEMENT

Consider a firm producing a good x from which it receives revenue $R(x)$; it can be operating in any type of market structure. The costs of producing good x, $C(x, w, u)$, depends on three critical elements: x, the amount of the good produced; w, the amount of wastes emitted; and u, the amount of concealment effort that the firm undertakes to avoid the detection of any waste emissions exceeding the level permitted by the government. Marginal revenue is R_x and the marginal cost of output is C_x, where partial derivatives are indicated by subscripts.

In the absence of any external constraints, the firm finds it profitable to increase its wastes to w_0, where $C_w = 0$. For the full range of waste emission levels from no emissions to w_0, the marginal costs of polluting are negative (i.e., benefits). At waste levels greater than w_0, however, C_w becomes positive and the costs of production will rise as the level of waste emission increases. Clearly, the firm will always avoid levels of waste that place it on this portion of the cost curve. Consistent with this description, we assume that the marginal costs of waste emission are increasing $C_{ww} > 0$ and that the marginal cost of output decreases with increases of waste emissions $C_{xw} \leqslant 0$ as long as the level of emissions does not exceed w_0. We further assume that the marginal costs of concealment efforts are positive $C_u > 0$ and increasing $C_{uu} > 0$. And as the level of concealment escalates, the marginal cost of waste emissions decreases $C_{wu} < 0$. Unlike changes in the level of waste emissions, however, increases in concealment are likely to raise rather than lower the marginal costs of output $C_{xu} \geqslant 0$.

Because a specific abatement technology is required of firms under direct controls but not under a tax regime, we can expect the firm's cost functions under the two regimes to differ. With direct controls, technology requirements imposed industry-wide without regard for differences in marginal abatement costs will typically generate an efficiency loss for the firms who choose to comply. Rather than attempting to specify either the size of the efficiency loss or the lump sum and operating costs involved, we simply assume that the firm's cost function for a direct control regime $C'(x, w, u)$ will not be less than its minimum cost function $C(x, w, u)$ at every value of output, waste, and concealment.

Direct Controls

The central agency determines several parameters that affect the firm. The first is a technology-based control standard w_s which places a limit on the amount of waste the firm is allowed to emit. This limit is gradually tightened to bring industry emissions to a level w^* sufficient to guarantee a desired degree of environmental quality. Given the stringency of w^* necessary to meet air and water quality goals, we assume that the control standard w_s in any given period is invariably below the level of waste emission the firm would choose in an unconstrained situation w_0 and frequently below the level chosen in practice w'. In order to ensure compliance with w_s the central agency introduces a set of civil and criminal fines intended to sanction firms whose waste levels exceed w_s. A fine is levied at each of three enforcement stages: a fine follows the initial detection of a violation; at the litigation stage an effective fine results from the expense of mounting a credible legal defense; and a fine is imposed upon conviction. These three fines can be represented as a column vector $\mathbf{F} = \{F(D), F(L), F(C)\}$. But a fine alone is not sufficient to deter potential violations; there must be some recognized likelihood that these fines will be imposed. Accordingly, the central agency also determines the chances that a given violation will be fined and litigated. The probability accompanying each stage of fines can be represented as a row vector $\mathbf{P} = \{P(D|V), P(L|D), P(C|L)\}$ whose elements are the conditional probability of detection given a violation, the probability of litigation following detection, and the probability of a conviction resulting from litigation.

The firm's decision whether to comply with the central agency's control standard is based, in part, upon the expected costs of noncompliance. The major component of the expected costs is the sum of possible fines weighted by the risk of their imposition. This sum is known as the *expected penalty* from noncompliance, assuming that firms are neither risk adverse nor risk acceptant, and is typically represented as the product of the probability and fine vectors set by the central agency, $\mathbf{P} \cdot \mathbf{F}$. However, in a decentralized system, the local enforcer, through its surveillance and inspection activities, is responsible for the actual likelihood of detection. Because of resource limitations and political pressures constraining local efforts, the chances of a local enforcer detecting a violation are generally smaller than those desired by the central agency.

In effect, the local enforcer revises the central agency's probability of detection $P(D|V)$ downward by an amount p resulting in a new probability $p(D|V)$. Changing the probability of detection alters the expected fines at each stage of enforcement and produces a new expected penalty. We can represent the local enforcer's adjustment as a row vector $\mathbf{p}$ with the amount p in the first row and zeros elsewhere. The *adjusted penalty* G is given by $G = [\mathbf{P} - \mathbf{p}] \cdot \mathbf{F}$.

In our earlier discussion of the local agency, we suggested that the firm could influence its chances of being detected $p(D|V)$ indirectly by changing the local agency's error costs. However, the firm can also reduce $p(D|V)$ more directly by actively avoiding detection through various measures designed to conceal the extent of its violation. Concealment measures include: attempts to change operations or employ idle control technology on a temporary basis in order to "pass" prearranged onsite inspections, reliance on unconventional disposal practices, falsifying reports, and image-building designed to erode the credibility of critics and divert attention from noncompliance. Depending upon the magnitude of the violation involved, the firm's concealment efforts can be quite costly. Nonetheless, even if such efforts are only partially successful, the firm will be able to reduce its chances of being caught by local enforcers. And in reducing $p(D|V)$ through concealment, the firm is altering the expected penalty attached to noncompliance.

More specifically, the firm's concealment efforts reduce $p(D|V)$ by an amount u, resulting in a new probability of detection $u(D|V)$. The impact of concealment on the expected penalty can be represented by a row vector $\mathbf{u}$ with the amount u in the first row and zeros elsewhere which, when subtracted from the probability component of the adjusted penalty, yields a *firm-adjusted* penalty $H = [\mathbf{P} - \mathbf{p} - \mathbf{u}] \cdot \mathbf{F}$.

Aside from indirect pressure on the local enforcer and concealment measures, the most effective means available to the firm for mitigating the expected costs of noncompliance is simply to limit the size of its violation. Since both the chances of detection and any subsequent fines increase with the seriousness of the violation, adjusting the size of v can have a direct impact on the size of the noncompliance penalty and upon the costliness of effective concealment. We can represent the impact of v on the noncompliance penalties by expressing the adjusted penalty G as $G(v)$ and the firm-adjusted penalty H as $H(v, u)$. The functions $G(v)$ and $H(v, u)$ are assumed to be continuous and twice differentiable. Similarly, the local enforcer's probability of detection $p(D|V)$ and the firm's revision $u(D|V)$ can be expressed as $p(v)$ and $u(v)$, respectively. Finally, the marginal costs of concealment vary as a positive function of the size of the firm's violation $u_v > 0$.

In order to complete our picture of the firm under direct controls, we must incorporate the costs to the firm of meeting compliance requirements. For most industries, the control standard w_s set by the central agency is accompanied by specific technology requirements. Unfortunately, the abatement technology designated for industry use need not be the least costly method of reaching w_s and may force some firms to operate inefficiently. The profit maximizing firm will consider these costs relative to the expected costs of noncompliance and concealment opportunities and choose an optimal waste level w' somewhat below w_0. The compliance cost is indicated by the area labeled A in Fig. 3. The size of the resulting violation is equal to $w' - w_s$. However, once this violation is detected, the firm will be forced to lower its w' to w_s in order to achieve full compliance. The expense of reducing w' represents an additional compliance cost equal to the area in Fig. 3 labeled B.

Since the additional costs of coming into full compliance are contingent upon detection, they are best understood as expected costs supplementing the firm-adjusted penalty for noncompliance. The expected cost to the firm of the additional move toward compliance is given by

$$[p(v) - u][C'(x, w_s, u) - C'(x, w', u)].$$

We can reexpress $C'(x, w_s, u)$ as $C^s(x)$ since the firm must view w_s as a constant, and u goes to zero once w_s is reached. Given our previous definitions, the firm's expected profit is equal to revenue minus costs minus the firm-adjusted penalty minus the expected costs of additional compliance, or

$$R(x) - C'(x, w, u) - H(v, u) - (p(v) - u)[C^s(x) - C'(x, w, u)].$$

The firm chooses an output x, a level of waste w, and a level of concealment u, so as to maximize expected profit.

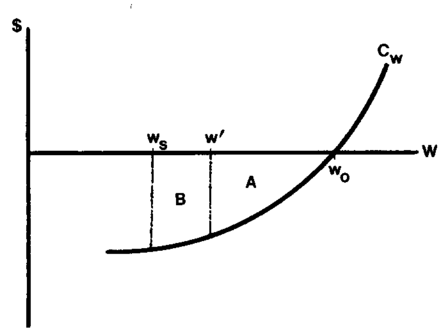

FIG. 3. The expected additional costs of compliance.

The first order conditions necessary for a solution are

$$\text{(a)} \qquad R_x - C_x' - (p(v) - u)(C_x^s - C_x') = 0$$

$$\text{(b)} \qquad -C_w' - H_w - \left[(p_w - u_w)(C^s - C') + (p(v) - u)(-C_w')\right] = 0$$

$$\text{(c)} \qquad -C_u' - H_u + (p(v) - u)(C_u^s - C_u') + (C^s - C') = 0.$$

Condition (a) states that marginal revenue should equal the marginal costs of output plus the expected marginal costs of output arising from additional compliance efforts. Note that if the cost function is separable in output and wastes ($C_{xw}' = 0$), this last term is zero and we get the conventional notion that marginal revenue should equal the marginal cost of output. If $C_{xw}' < 0$, we can expect the term $C_x^s - C_x'$ to be positive, and marginal revenue will exceed the marginal cost of output. Thus, the threat of detection, together with the additional burden to achieve full compliance, effectively reduces the firm's level of output.

Condition (b) tells us that the firm chooses a level of waste such that the marginal cost (benefit) of one more unit of waste C_w' is equal to the marginal firm-adjusted penalty plus the marginal expected costs of additional compliance. The marginal expected cost of additional compliance is positive; therefore, the firm emits wastes at a level lower than if a penalty alone were imposed. While the equality in (b) might fail because of a discontinuity at $w = w_s$ where the violation disappears, given our modelling of agency behavior the firm still faces some positive probability of mistaken detection. Another corner solution occurs if the firm ignores enforcement altogether and sets $w = w_0$. This condition is relevant whenever the cost savings from the marginal effect of concealment exceed the cost increases resulting from the marginal effect of the probability of detection on both the penalty and the costs of additional compliance. Finally, if the probability of detection, the penalty set by the central agency, and the firm's costs of concealment are fixed and independent of the size of the violation, the firm will make no effort to comply (i.e., the firm chooses $w = w_0$).

Condition (c) implies that the firm chooses the optimal level of concealment activity by setting the marginal cost of concealment (C_u') equal to its marginal benefit. Concealment's benefits derive from the marginal reduction it produces in both the expected penalty and the expected costs of additional compliance. Under these circumstances, the firm will undertake more concealment effort than it would if only a penalty were at risk. Note that the imposition of the additional costs of compliance contingent upon detection causes the firm to reduce profit maximizing waste levels *and* to increase socially unproductive concealment activities.

Under the second order conditions the cross partials of the cost and expected penalty functions should not be too large in comparison with the repeated partials in x, w, and u. Satisfaction of these conditions ensures a global maximum with an interior solution as long as there is no discontinuity at $v = 0$ and the cost of concealment is not too low relative to the expected penalty. While our derivation of the comparative statics for the firm's response to the agency parameters $p(v)$, $\mathbf{F}$, and w_s fails to indicate the appropriate signs for a number of elements, we will offer tentative predictions by focusing largely upon responses occurring after the first round of changes.

Consider an increase in the probability of detection $p(v)$. If $p(v)$ goes up, the expected penalty *and* the expected costs of additional compliance will rise, necessitating a redress of all three conditions. A change in output will start to redress the balance in condition (a). First, as output drops, the marginal cost of production drops and the marginal revenue rises. Second, if $C'_{wx} < 0$ and $p(v)$ goes up, the optimal level of waste emission drops causing the term $(C_x^s - C_x')$ to decrease and reducing further the expected costs of additional compliance. Finally, the drop in x and w is likely to decrease the marginal cost of concealment (through C'_{ux} and C'_{uw}) while, at the same time, the expected cost of additional compliance is falling. This leaves the balancing of concealment u uncertain. The first round effects of an increase in the probability of detection may raise or lower u, but in either event appear to diminish both output and waste emissions.

Now consider an increase in the level of fines. If F goes up, the expected penalty goes up, but the expected costs of additional compliance do not. Thus, there is no first round effect on the level of output the firm chooses. As the expected penalty goes up, creating an inequality in condition (b), the optimal level of emitted wastes will drop. This drop in w' will reduce the size of the violation and lead to second round effects on x, w and u. The rise in the fine will also cause the marginal benefit from concealing the violation to rise. This will increase concealment efforts necessary to redress the balance in condition (c). Because the fine works by itself while the probability operates on both the expected penalty and the expected costs of additional compliance, raising the level of the fine will have less of an impact on firm behavior than will increasing the likelihood of detection.

Finally, consider the effect on the firm of the agency's setting a stricter standard (reducing w_s). By augmenting the size of exiting violations, a more stringent w_s increases both the firm's chances of being detected and its concealment efforts. But it is uncertain whether the firm is also given an incentive to lower its waste emissions; this depends largely on how the expected costs of noncompliance are affected. If increasing the violation by lowering w_s has a larger marginal effect on the probability of detection than on the expected penalty and costs of additional compliance, the expected costs of noncompliance will increase and the firm's optimal levels of output and waste will drop. If, on the other hand, the marginal effect on the expected penalty and costs of additional compliance is greater, the costs of noncompliance will fall, stimulating an increase in waste and output levels. It is likely that the second and third round effects will be sizable in this particular set of changes, contributing to the ambiguity regarding the precise effect that changing standards will have on firm behavior.

Changes under a Tax Regime

The firm's behavior under direct control is likely to differ substantially from its response to a tax assessed on the level of emitted wastes. Evasion of an emissions tax occurs when the firm sets reported wastes w_r below the level of emitted wastes. Again we assume that the firm can reduce the probability of detection $p(v)$ by an amount u through efforts to conceal the extent of its underreporting (i.e., violation). Concealment activities include those listed in the previous section and perhaps the additional expense of satisfying more extensive and detailed reporting requirements to establish tax liability. The firm's costs C are a function of output x, level of wastes w, and concealment effort u, as in the previous section. All of the signs of the partial derivative in the cost function remain the same.

A unit tax t is applied to reported wastes w_r, and the total pollution tax equals $t \cdot w_r$. In order to prevent the firm from completely avoiding the tax by setting $w_r = 0$, the agency exhibits a set of fines and probabilities of detection for violations $v = w' - w_r$. Again we will use $H(u, v)$ as the firm-adjusted penalty. The expected cost of additional compliance is no longer relevant under a tax regime. Instead, the firm found in violation for underreporting its actual wastes must pay both the penalty given by $H(u, v)$ and the outstanding tax on the unreported wastes, $t(w - w_r) = tv$. The expected tax liability following detection equals the probability of detection, adjusted by effective concealment, times the amount of additional tax, or $(p(v) - u)(vt)$.

Using $R(x)$ to denote total revenue as a function of output, the expected profit of the firm may be written as

$$R(x) - C(x,w,u) - w_r t - H(u,v) - (p(v) - u)vt.$$

The firm now maximizes expected profit by setting *four* variables; x, the level of output; w, the level of emitted wastes; w_r, the level of reported wastes; and u, the level of concealment effort. Accordingly, the four first order conditions are

$$\text{(a}'\text{)} \qquad R_x - C_x = 0$$

$$\text{(b}'\text{)} \qquad -C_w - H_w - [(p_w - u_w)vt + (p(v) - u)v_w t] = 0$$

$$\text{(c}'\text{)} \qquad -C_u - H_u + vt = 0$$

$$\text{(d}'\text{)} \qquad -t - H_{wr} - [(p_{wr} - u_{wr})vt + (p(v) - u)v_{wr} t] = 0.$$

Condition (a′) simply states that the firm chooses its output level by setting marginal revenue equal to the marginal costs of output. Note that, in contrast to the standards regime, the expected costs of additional tax liability do *not* affect the firm's choice of output level.

Condition (b′) tells us that the firm chooses a level of waste such that the marginal cost (benefit) of one more unit of waste C_w is equal to the marginal expected costs of both the penalty H_w and additional tax liability. Since, the marginal cost of additional tax liability is positive, the firm will emit wastes at a level lower than if only a penalty were imposed. While the equality in (b′) might fail because of a discontinuity at $w = w_r$, where the violation disappears, given our modeling of agency behavior, the firm will continue to face some positive probability of mistaken detection.

Condition (c′) states that the firm chooses the optimal level of concealment by setting its marginal cost C_u equal to the marginal benefit. The marginal benefit of concealment is composed of the marginal reduction in both the expected penalty and the expected costs of additional tax liability. Again, the imposition of the additional compliance requirements following detection causes the firm to reduce the amount of emitted wastes and to step up its concealment activity.

Condition (d′) shows that at maximum profits, the firm reports wastes only to the point where the acknowledged tax liability equals the marginal expected penalty H_{wr} plus the marginal additional tax liability from changing w_r. If conditions (a′) through (d′) hold, it is obvious that $-C_w = t$; in other words, the marginal benefit of waste emission will be set equal to the marginal unit tax on the reported emissions.

Once again, the second order conditions for an interior maximal solution are assumed to hold, and because of problems with the comparative statics, we will limit our focus to the effects following the first round of changes. Consider an increase in the probability of detection $p(v)$ *or* in the level of the fine structure **F**. If $p(v)$ or **F** rise, the expected penalty and the expected costs of additional tax liability will rise, necessitating a redress of the four, first order conditions. The output x will not change with this increase until the second round of effects. Any increases in the expected penalty and in the additional tax liability will force an initial drop in the level of emitted wastes and will also tend to reduce the level of reported wastes. This decrease in w, in turn, reduces the marginal cost of concealment (through C_{uw}), raising its optimal level.

Now consider the impact of a change in the tax rate t. Since the first order conditions require that $-C_w = t$, an increase in the tax rate will cause an increase in the marginal cost (reduction in the marginal benefit) of emitting wastes. This will reduce the level of emitted wastes directly. Also note that the increase in t will decrease the level of reported wastes w_r as shown in condition (d′). Raising the tax rate can also raise the size of the violation by lowering w_r faster than it lowers w. Any increase in the size of the violation v will escalate the level of concealment in a compensatory fashion to balance the resulting increase in the expected penalty.

CONCLUDING REMARKS

The firm's behavior under the two alternative regimes sheds light on the ability of the central or local agency to encourage compliance with any control scheme. Under both regimes, the local agency's likelihood of detecting noncompliance has a greater capacity to affect the firm's behavior than does the level of the fine because it affects both the expected penalty and the expected cost of additional compliance/tax liability. For any given percentage change in the expected penalty due to either **P** or **F**, a p-induced change will produce a greater percentage change in wastes emitted w. Moreover, a tax regime is likely to lead to a relatively higher level of emitted wastes since the firm can optimally set both the level of emitted wastes *and* the level of reported wastes. This allows the firm to trade off the gains from emitting wastes against the costs in order to select the optimal size of its violation. Conversely, under a direct control regime, the firm chooses only the level of w relative to a w_s which is fixed from the firm's point of view.

A tax regime offers the firm considerable flexibility in choosing w_r and w, without the burden of negotiating with the local agency. From the perspective of the local agency, however, a tax regime rules out discretion in the treatment of firms and consequently removes the opportunity for the informal resolution of detection mistakes. Restricting resolutions to formal modes, such as adjudication, raises the cost of "false alarms," but probably has little impact on the cost of "misses." Under these circumstances, the local agency's optimal response is to lower its frequency of "false alarms" at the expense of aggressive detection. And in spite of the sensitivity of the firm's compliance incentives to changes in the probability of detection, unless the detection system is upgraded to compensate for the conservatism of many local enforcers, noncompliance can be expected to increase relative to levels under direct controls. Thus, damage costs are likely to escalate under a tax regime with increased enforcement costs as the only remedy.

The probability of detection is not only subject to changes that might occur in the local agency's values and expectations, but is also vulnerable to mitigation by the firm's concealment efforts. Reducing the impact of the probability of detection is also likely to weaken the other components of the penalty structure, since all of the expected penalties are contingent on detection. The fine structure, which enables the central agency to set the level of expected penalty, induces the firm to comply with mandated standards or reporting requirements through its contribution to raising the costs of noncompliance. However, a low probability of detection lowers the expected costs from noncompliance, and thus raises the firm's optimal level of waste and violation. The size of this increase will depend upon whether the firm found a need to engage in concealment, and whether the cost of these efforts remained smaller than the potential gains from higher waste levels. Any effort by the central agency to reverse the increase in waste levels by raising the expected penalty—this includes lowering w_s, raising t, or raising the size of the fines—will be discounted by the probability of detection.

It would appear then that raising the probability of detection is the most crucial step toward securing compliance under either regime. However, under a decentralized administration, any changes in the probability of detection can occur only if the local agency has both adequate resources and proper incentive. In short, regardless of the stringency of the controls and penalties promulgated by the central authorities, unless violations are first detected by the local agency, other measures will have little practical significance for the profit-maximizing firm.

By examining regulation within a decentralized system, we have not only relaxed the assumption of full compliance by firms, but also rejected the notion of uniform, perfect detection on the part of enforcement authorities. Clearly, there is widespread diversity in enforcement practices across different local areas and often a divergence between the objectives of central and local authorities. If the realities of decentralization are ignored, efforts to enhance compliance can evolve into futile attempts to gain control over enforcement. In these instances, the central authorities will revert to rigid operating procedures designed to remove the discretion of local agencies. Nevertheless, detection will remain ideosyncratic, a product of how the local agency responds both to the requirements imposed by several higher authorities and to the compliance behavior of the firms under its jurisdiction.

How then can the enforcement performance of local agencies be enhanced? It makes little sense to ignore the diversity of enforcement situations or to try to negate it through the imposition of inflexible formulas. According to our model, decentralized enforcement has the advantage of supporting a cooperative relationship between firms and local agencies. The local agency sets a threshold for its detection decisions depending, in part, upon knowledge of firms' past behavior and contact with the firm's management. In other words, the rational local agency relies on prior odds to discriminate among firms, rather than using a rigid rule promulgated by an unfamiliar agency from a distant location. Better technology and more resources for surveillance would reduce the error in these odds.

More importantly, the rational local agency is highly sensitive to the potential, political costs and benefits of its decisions. The aggressiveness of local detection efforts—the trade-off between likely error and success rates—can be influenced by changing the local agency's incentives. By increasing the value of some outcomes while reducing others, local agencies can be made either more lenient or more severe in their detection decisions, altering the probability of detection in the aggregate.

Achieving a higher level of compliance by firms will be aided, not by demanding inflexible compliance by local authorities, but by: (1) acknowledging the separation of enforcement powers common to a decentralized system, and (2) altering either the detection problem or the incentives faced by the local agency.

Our analysis has been limited in the sense that we have taken a comparative statics approach to delineate the basic features of the enforcement relation, rather than attempting to develop a complex but more powerful dynamic model of the enforcement situation. The next stage of research should move beyond our rudimentary systems perspective, with its emphasis upon the distinctive motivations of the local enforcer and the industry enforcee in a decentralized enforcement environment, to an interactive framework allowing for explicit endogeneity in the participants' objective functions. In this way, we can model each actor's strategic responses to changing parameters, while identifying optimal adjustments for better avoiding detection or ensuring greater compliance.

REFERENCES

1. G. Becker, Crime and punishment: An economic approach, *J. Pol. Econ.* **76**, 169–180 (1968).
2. C. Coombs, R. Dawes, and A. Tversky, "Mathematical Psychology," Prentice-Hall, Englewood Cliffs, N.J. (1970).
3. P. Downing and W. Watson, Jr., The economics of enforcing air pollution controls, *J. Environ. Econ. Manag.* **1**, 219–236 (1974).
4. J. Harford, Firm behavior under imperfectly enforceable pollution standards, *J. Environ. Econ. Manag.* **5**, 26–43 (1978).
5. D. Laming, "Mathematical Psychology," Academic Press, New York (1973).
6. B. Lindgren, "Statistical Theory," 3rd ed., Macmillan Co., New York (1976).
7. R. McKean, Enforcement costs in environmental and safety regulation, *Policy Anal.* **6**, 269–289 (1980).
8. A. Polinsky, Private versus public enforcement of fines, *J. Legal Studies* **8**, 105–127 (1979).
9. R. Posner, The behavior of administrative agencies, *J. Legal Studies* **1**, 305–344 (1972).
10. G. Stigler, The optimum enforcement of laws, *J. Pol. Econ.* **78**, 526–537 (1970).
11. W. Viscusi and R. Zeckhauser, Optimal standards with incomplete enforcement, *Public Policy* **27**, 437–456 (1979).