



## Efficient Institutions for the Private Enforcement of Law

David Friedman

*The Journal of Legal Studies*, Vol. 13, No. 2. (Jun., 1984), pp. 379-397.

Stable URL:

<http://links.jstor.org/sici?sici=0047-2530%28198406%2913%3A2%3C379%3AEIFTPE%3E2.0.CO%3B2-A>

*The Journal of Legal Studies* is currently published by The University of Chicago Press.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# EFFICIENT INSTITUTIONS FOR THE PRIVATE ENFORCEMENT OF LAW

DAVID FRIEDMAN\*

## I. INTRODUCTION

OUR present legal institutions combine elements of private and public enforcement of law. If someone breaks your arm you call a policeman; if he breaks a window or a contract, you call a lawyer. Becker and Stigler have suggested that it would be advantageous to extend private enforcement into the area where law is now enforced publicly.<sup>1</sup> Their central argument is that the public system has perverse incentives. Suppose a policeman has evidence that will convict me of an offense, the punishment for which is equivalent (to me) to a twenty thousand dollar fine. If the cost to the policeman of "losing" the evidence is anything less than twenty thousand dollars, an opportunity exists for a mutually beneficial transaction. Preventing such transactions is costly. The solution proposed by Becker and Stigler is for the policeman's salary to consist of the fines produced by his activity. The only bribe he would then be willing to take would be one at least as large as the fine, in which case the "bribe" is simply a way of collecting the fine while avoiding the cost of a trial. In such a system the "policeman" is essentially a private agent. Becker and Stigler envisage a system of private enforcement firms that support themselves by the fines they collect from the criminals they apprehend and convict. If the criminal is judgment proof, the state would provide a reward equal to the fine the criminal would have paid if payment could have been enforced.

Landes and Posner argued in response that the private system has essential flaws that make it inferior to an ideal public system except for

\* Assistant Professor, UCLA, and Visiting Associate Professor, Tulane University.

<sup>1</sup> Gary S. Becker & George J. Stigler, *Law Enforcement, Malfeasance, and Compensation of Enforcers*, 3 *J. Legal Stud.* 1 (1974).

[*Journal of Legal Studies*, vol. XIII (June 1984)]

© 1984 by The University of Chicago. All rights reserved. 0047-2530/84/1302-0006\$01.50

offenses that can be detected and punished at near zero cost.<sup>2</sup> They concede that the private system might still be preferable to the less than ideal public system that we observe. However they argue that the prevalence of private enforcement for offenses that are easily detected (most civil offenses) and its rarity for offenses that are difficult to detect (most criminal offenses) suggest that our legal system is, at least in broad outline, efficient, using in each case the most efficient system of enforcement.

The purpose of this paper is to show that the inefficiency Landes and Posner have demonstrated in the particular private enforcement institutions they describe can be eliminated by minor changes in the institutions. The result is a system of private enforcement that is equivalent to an ideal public system and hence superior to any likely public system.

Section II explains the institutions for private enforcement described by Landes and Posner and sketches their demonstration that those institutions produce an inefficient outcome. Section III presents the argument in terms of an explicit model of optimal punishment developed in an earlier paper,<sup>3</sup> showing how the institutions of private enforcement (in the context of optimal behavior by the governmental court system) can be designed to produce the efficient outcome. Section IV shows that the institutions I describe can also produce an optimal level of defensive expenditure by potential victims. Section V describes a new problem introduced by the proposed institutions, and Section VI offers a possible solution. Section VII summarizes the results.

## II. WHY PRIVATE ENFORCEMENT MUST BE INEFFICIENT

Posner and Landes start by assuming that there is a single kind of crime and that each offender commits one offense. Competing private firms apprehend and convict offenders. The number of offenders apprehended (and convicted),  $A$ , is an increasing function of the number of offenses,  $O$ , and of the quantity of resources  $R$  (available in any quantity at a constant price per unit of  $r$ ) spent by the industry.  $A(O, R)$ , the industry production function, is assumed to exhibit constant returns to scale;  $A(aO, aR) = aA(O, R)$ . The penalty for an offense is a fine  $f$ , paid by the offender and received by the firm that apprehends him. The number of offenses is a decreasing function of  $f$  and of  $p = A/O$ , the probability that an offender will be apprehended (and convicted). Any firm can investigate any of-

<sup>2</sup> William M. Landes & Richard A. Posner, *The Private Enforcement of Law*, 4 *J. Legal Stud.* 1 (1975).

<sup>3</sup> David D. Friedman, *Reflections on Optimal Punishment, or: Should the Rich Pay Higher Fines*, 3 *Research Law & Econ.* 185 (1981).

fense, so the "supply of offenses," which Landes and Posner view as an input to the production of  $A$ , is treated as a common pool resource.

Landes and Posner discuss a number of problems with private enforcement, including the common-pool nature of this "input,"<sup>4</sup> but their central argument for the inefficiency of private enforcement does not depend on whether offenses are private, public, or common property nor on whether the industry is monopolistic or competitive, nor on the particular assumption about punishment costs that the authors use to simplify their exposition. It may be stated as follows.

The output of crimes depends on the punishment imposed and its probability; in the simple case of a fine imposed on risk-neutral criminals it is a function of  $fp$ , the expected punishment. One can imagine solving the optimization problem faced by a public enforcement agency in two stages. For simplicity I assume that criminals are risk neutral; the generalization is straightforward. For any given expected punishment find the optimal combination of  $f$  and  $p$ ; having done so, find the optimal expected punishment. In solving the first problem, costs associated with the crime rate  $O$  may be ignored, since all combinations of  $p$  and  $f$  which produce the same expected punishment will result in the same crime rate. Increases in  $p$  (for a given crime rate) require increases in the (costly) resources spent on apprehending and convicting criminals. Increases in  $f$  increase the fraction of the criminals who are unable to pay the fine and must be punished in other ways. A fine is a "costless" punishment since the criminal's loss is someone else's gain.<sup>5</sup> A punishment such as flogging or execution has a cost roughly equal to the amount of the punishment, since the criminal's loss is nobody's gain (unless, as was often the case in the past, punishment is made a public spectacle). Imprisonment has a cost greater than the amount of the punishment, since the cost to the state of the imprisonment must be added to the cost to the criminal. Hence the higher  $f$ , the higher the net cost of punishment.<sup>6</sup> An efficient system will, for any given

<sup>4</sup> Landes and Posner's claim that treating the supply of offenses as a common pool resource leads to inefficiencies is discussed in a longer version of this article available from the author.

<sup>5</sup> If the cost of the trial does not depend on the punishment, it may be included with enforcement costs; if it does, it must be considered a punishment cost. I am assuming the former for purposes of expository convenience.

<sup>6</sup> A very severe punishment (execution) might be less costly than a lesser punishment (imprisonment), in which case the net cost would decrease as the level of punishment increased. In such a case, however, imprisonment need never be used. There is some probability of execution that the criminal would regard as equivalent to imprisonment. If execution were imposed with that probability, net punishment cost would be less and all other costs would remain the same. Hence "probabilistic execution" dominates certain

level of expected punishment, choose  $p$  and  $f$  to minimize the sum of enforcement and punishment costs.<sup>7</sup> Having done so, it will then choose the level of expected punishment that minimizes total cost.

The problem with a system of private enforcement (combined with a state run court system that specifies the fine) is that the state has available only one control variable,  $f$ , with which to do both maximizations. Since  $f$  is the price that enforcers are paid for apprehensions and the price that convicted criminals are charged for crimes, it will simultaneously determine  $A$  and  $O$  (and their ratio  $p$ ) from the industry supply curve for apprehensions and the criminals' supply curve for offenses. If the state chooses the value of  $f$  that generates the optimal expected punishment (in the sense of the preceding paragraph), it has no way of adjusting the combination of  $f$  and  $p$  in order to produce that expected punishment in the least costly way. By adjusting  $f$  the state is, in effect, moving along a line in a plane whose axes are  $f$  and  $p$ ; there is no reason save chance to expect that line to intersect the optimal combination  $f^*, p^*$ .

The graphical form of the argument is made explicit in Figures 1 and 2; Figures 1*a* and 1*b* show the case assumed by Landes and Posner, where there is a maximum fine  $f_m$  that can be collected costlessly, and where no more severe punishment is possible. Figures 2*a* and 2*b* show the more general case in which punishment cost increases continuously with the amount of the punishment. The details of the figures are explained in the Appendix.

The curve  $PR$  in Figure 1*a* is the set of possible combinations of  $p$  and  $f$  under private enforcement. The point  $p^*, f^*$  is the optimal combination of  $p$  and  $f$ , the combination that would be chosen by a wise and benevolent public enforcement system with access to the same production function for apprehensions as the private industry.  $PR$  depends only on the form of  $A(O, R)$  and the price  $r$  of enforcement resources; it is independent of the supply function for offenses  $O$  and of the damage function that describes the cost to victims of an offense rate  $O$ . Independence of  $O$  follows from the assumption of constant returns to scale; since in equilibrium  $f$  equals the average cost of the industry, an exogenous doubling (say) of  $O$  with  $f$

---

imprisonment. It follows that as long as we limit ourselves to the least costly among alternative punishments, net cost of punishment is a nondecreasing function of amount of punishment.

<sup>7</sup> It follows from the argument of note 6 *supra* that in an efficient system we should never observe imprisonment if execution is less costly. The fact that we do observe imprisonment may be interpreted to mean either that our system is not an efficient one or that execution is more costly (relative to imprisonment) than might at first appear. The question is discussed at some length in Friedman, *supra* note 3.

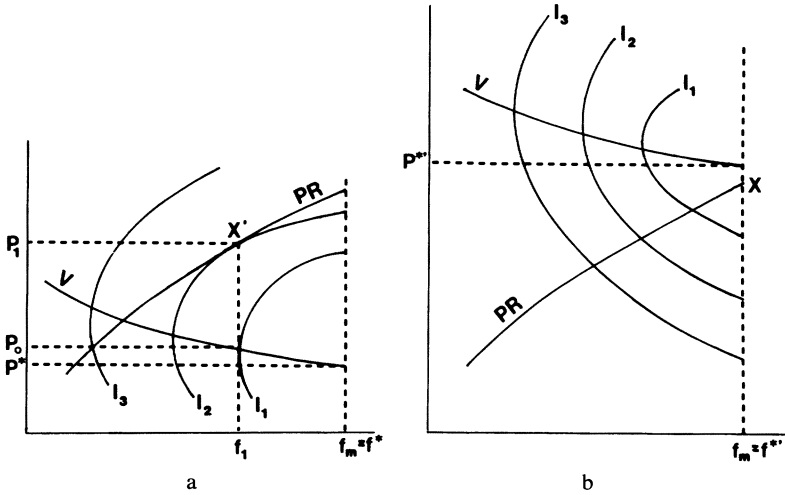


FIGURE 1

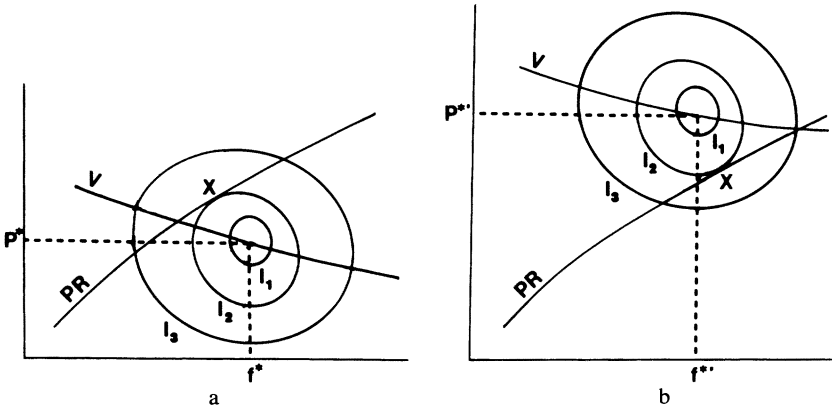


FIGURE 2

fixed will result in a doubling of  $R$  and  $A$ , giving the same value of  $p = A/O$  as before. The same result can be demonstrated for a monopoly enforcement industry.

But the optimal combination of punishment and probability—the combination that minimizes social loss—does depend on both the supply function for crimes and the damage function. The larger the decrease in the crime rate produced by a given increase in expected punishment, the

larger is the maximum price—in enforcement and punishment costs—that it is worth paying, if necessary, to get that increase. Similarly, the larger the cost imposed by a crime on its victims, the larger the maximum price it is worth paying to prevent the crime.<sup>8</sup> Hence if we keep  $r$  and the functional form of  $A(O, R)$  fixed while altering the form of  $O$  and/or the relation between  $O$  and damage,  $PR$  will remain fixed while  $p^*$ ,  $f^*$  will change, as shown by  $p^{*'}, f^{*'}$  in Figure 1*b*. In Figure 1*b*,  $f^{*'}$  is the same as  $f^*$  in Figure 1*a*, since under the Landes-Posner assumption of a costless maximum fine social loss is minimized by charging the maximum fine and choosing  $p^*$  to give the optimal expected punishment. In the more general case of Figures 2*a* and 2*b*, this is no longer true. In both cases, since the optimal point shifts while the private enforcement line stays fixed, it can only be by chance that the two intersect for any particular forms of the relevant functions.

Landes and Posner assert that not only does private enforcement lead to an inefficient result, it leads in general to overenforcement; the best combination  $(p, f)$  available on the trajectory  $PR$ ,  $(p_1, f_1)$ , has  $p^* < p_0 < p_1$  ( $p_0$  is the optimal probability given  $f_1$ ), save in the special case of corner solutions. In the Appendix, I show that this result is misleading: what the authors have shown is not that private enforcement tends to lead to overenforcement but only that when it leads to underenforcement it also, under their particular assumption about punishment cost, produces a corner solution at the maximum possible punishment. The result disappears in the more general case illustrated in Figures 2*a* and 2*b*.<sup>9</sup>

However we put the argument, and whether or not we assume a costless maximum fine, the important result is unchanged; private enforcement, under the institutions described by Landes and Posner, cannot lead to an efficient outcome save by chance. Hence it is inferior to an ideal system of public enforcement.

One solution to this problem is for the state to tax (or subsidize) the private enforcement firms, driving a wedge between the price charged criminals and the price paid firms.<sup>10</sup> But this eliminates the desirable

<sup>8</sup> The exact relationship is derived in Friedman, *supra* note 3; it may be summarized for risk-neutral criminals by the formula "expected punishment equals increase in costs to victims, enforcement costs, and punishment costs produced by a reduction in the expected punishment sufficient to increase the number of offenses by one."

<sup>9</sup> A proof that private enforcement must under some circumstances lead to underenforcement is provided in A. Mitchell Polinsky, *Private versus Public Enforcement of Fines*, 9 J. Legal Stud. 105 (1980), but does not contradict Landes and Posner's argument since it involves a corner solution at the maximum collectable fine.

<sup>10</sup> Polinsky, *supra* note 9, at 120, demonstrates that the appropriate bounty will lead to an efficient outcome in the case of a private competitive industry but not in the case of a private monopoly industry.

incentives that were the original argument for the private system; since the cost to a criminal of being convicted no longer equals the benefit to the enforcer of convicting him, there is again an incentive for the criminal to bribe the enforcer to let him go (in the case of a tax)—or for the enforcer to bribe people to confess to crimes they have not committed (in the case of a subsidy). It appears that a private system must, save by chance, be inferior to an optimal public system.

### III. INSTITUTIONS FOR EFFICIENT PRIVATE ENFORCEMENT

I begin by introducing an explicit model of optimal enforcement, borrowed with minor alterations from an earlier paper.<sup>11</sup> I define:

- $f$ : punishment imposed upon any criminal who is punished;<sup>12</sup>
- $E$ : the certainty equivalent to the criminal (in dollars) of a probability  $p$  of punishment  $f$ ;
- $F \equiv E/p$ : the amount of the punishment, as perceived by the criminal;
- $O \equiv O(E)$ : number of occurrences of the crime per year;
- $A$ : number of perpetrators per year apprehended and punished;
- $p \equiv A/O$ : probability that an occurrence of the crime will result in apprehension and punishment of the perpetrator;
- $C(p, O) \equiv rR$ : total cost of system of maintaining a probability  $p$  of punishment for a given  $O$ ;
- $F'$ : the amount received by the court system when it imposes punishment  $f$ ;
- $Z(p, f) \equiv (F - F')/F$ : the punishment inefficiency;
- $D(O)$ : the aggregate net damage resulting from  $O$ , defined as the loss to the victims minus the gain to the criminals;<sup>13</sup> and
- $L \equiv D + C + E O Z$ : the social loss function.

<sup>11</sup> Friedman, *supra* note 3.

<sup>12</sup> Since the punishment need not be a fine,  $f$  is not a number of dollars but a particular punishment—a fine of a certain amount, imprisonment for a certain length of time, a particular number of blows with a whip.

<sup>13</sup> Some would argue that benefits to the criminal should not be included in such calculations. I disagree, since I believe that in the economic analysis of law what is or is not a crime should be a conclusion of the theory, not an assumption; robbery should, in this context, be considered bad only because permitting it is inefficient. If benefits to the criminal were eliminated (or discounted) the result would be to change the optimal level of punishment but not the conclusions of Sections III and V of the paper; I am not certain of the effect on the conclusions of Section IV.



Note that social loss is the sum of the net cost of the crimes, the cost of enforcement, and the cost of punishment. In the special case of a fine  $f$  imposed on a risk-neutral criminal,  $F$ , the "amount" of the punishment, is equal to  $f$ . For a fine imposed on a risk-averse criminal,  $F > f$ ; for a risk preferring criminal,<sup>14</sup>  $F < f$ . The inefficiency of the punishment,  $Z$ , is the ratio of the cost of the punishment to the amount of the punishment; for a fine imposed on a risk-neutral criminal, it would be the ratio of the collection cost to the amount of the fine. More generally, it includes as one of the costs (or benefits) of punishment the cost (or benefit) of imposing a punishment lottery on a risk avoiding (or preferring) criminal.

For any particular  $p$ , consider different punishments  $f$  that produce the same  $F$ . Since they all represent equivalent lotteries from the standpoint of the criminal, the only term in  $L$  that depends on  $f$  is  $Z$ ; an efficient system will always choose  $f^*$ , the  $f$  for which  $Z$  is minimized. We may then consider  $Z$  as a function of  $F$ . It is a nondecreasing function; the lower the punishment, the more likely it is that the criminal can pay it as a fine.<sup>15</sup>

Since  $D$  depends on  $O$  and hence on  $E$ , but not separately on  $F$  and  $p$ , we can choose for each value of  $E$  a pair  $F^*, p^*$  that minimizes  $G = C + E OZ$ . We may then define

$$L(E) = D[O(E)] + C[p^*(E), O(E)] + E O(E) Z[F^*(E)].$$

Since this is only a function of one variable, we choose  $E$  to minimize it; we call this value  $E^*$ .

Now, following Landes and Posner, assume a private system with a supply curve  $A(O, f) = p(f) O(p, f)$ ;  $p$  depends only on  $f$  because of the assumption of constant returns in the enforcement industry. The form of  $A$  depends on the production function for apprehensions; the cost  $C$  is proportional to  $O$  and an increasing function of  $p$ , again by constant returns. Assuming that the court chooses the particular punishment  $f$  so as to minimize  $Z$  for a given  $F$ , the only control variable is  $F$ ; in the special case discussed by Landes and Posner the punishment is always a fine and  $Z = 0$  for  $f < f^*$ ,  $Z = \infty$  for  $f > f^*$ ; if criminals are risk neutral, the control

<sup>14</sup> But see David D. Friedman, Why There Are No Risk Preferrers, 89 J. Pol. Econ. 600 (1981), for the reasons criminals (and others) are unlikely to exhibit much risk preference. Gary S. Becker, Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968), concludes that in the equilibrium of an efficient system criminals must be risk preferrers, but I show in Friedman, *supra* note 3, at 202-3 n. 3, that the result depends either on omitting costs and benefits associated with risk preference from the social loss function and assuming  $Z$  to be independent of  $E$  or on an artificial and implausible assumption about the behavior of  $Z$ . The point is discussed at greater length in David D. Friedman, Are Criminals Risk Preferrers? A Belated Comment (manuscript) (available from author).

<sup>15</sup> The result follows in general from the argument given in note 6 *supra*.

variable is  $f = F \leq f^*$ . Once chosen,  $F$  determines  $p$  and  $O$  through the supply functions for apprehensions and crimes and hence  $E$ ; there is no reason save chance why the  $F$  that implies  $E^*$  should be  $F^*(E^*)$ .

Three changes in the institutions assumed by Landes and Posner eliminate this problem. First, assume that offenses belong to the victims and must be purchased before or immediately after they occur. Second, assume that the state, instead of imposing a fine  $f$  or a punishment amount  $F$ , imposes an expected punishment  $E$ . For simplicity in the discussion I shall assume that all criminals are risk neutral, that all punishments are fines, and that the ratio of fine collected (fine paid minus collection costs) to fine paid is a decreasing function of the size of the fines; the generalization is straightforward.<sup>16</sup> Finally, assume that the firm receives not  $F$  but  $F'$ , the fine collected rather than the fine paid.

The requirement that crimes be bought at the latest immediately after they occur (that is, before the criminal has been apprehended) is essential in order to make any sense out of the idea that the court system is to set the expected punishment rather than the actual punishment. Suppose, for example, that the expected punishment is set at one thousand dollars. A particular firm has purchased one hundred occurrences from the victims. If it succeeds in catching and convicting all one hundred perpetrators, it can fine them a thousand dollars each—a thousand dollar fine times a probability of unity is an expected punishment of one thousand dollars. If it catches and convicts only one criminal, it can fine him one hundred thousand dollars—again an expected punishment of a thousand dollars, this time in the form of one chance in a hundred of a punishment of a hundred thousand dollars.

If it were as easy to collect a fine of one hundred thousand dollars, or one million, or ten million, as a fine of one thousand dollars, then one firm would buy all offenses, catch one perpetrator (thus minimizing its enforcement costs), and collect the entire sum from him. But under those assumptions, as I have shown elsewhere,<sup>17</sup> and as the reader can easily prove to himself, the corner solution of an infinite punishment imposed with infinitesimal probability is optimal. More realistically, punishment inefficiency increases with the size of fine; the firm must weigh the cost of catching more criminals against the advantage of being able to collect a larger fraction of the fines they pay.

In choosing the firm's values  $p_i$ ,  $F_i$ ,  $p_i F_i = E$ , the firm is minimizing  $C_i(p_i, O_i) + p_i O_i Z(F_i)$  for given values of  $E$  and  $O_i$ . Assuming that there

<sup>16</sup> To make the argument general one replaces "expected value" with "certainty equivalent" and notes that  $Z$  includes the cost or benefit of imposing a punishment lottery on a criminal who is not risk neutral.

<sup>17</sup> Friedman, *supra* note 3.

are no costs external to the firm but internal to the industry, such as those associated with a common-pool resource (I shall return to this point later), we have  $C_i(p_i, O_i) = C(p_i, O)/N$  by our assumption of constant returns to scale, where  $N$  is the number of identical firms. Since  $O_i = O/N$ , the firm's problem is to minimize

$$G_i \equiv \frac{C(p_i, O) + p_i OZ(F_i)}{N} = \frac{G[p_i, F_i, O(E)]}{N}.$$

But this is done by minimizing  $G$ , hence the solution is  $p_i = p^*(E)$ ,  $F_i = F^*(E)$ . The first part of the optimization has been done by the firm as a consequence of its own profit maximization. It only remains for the court system to set  $E = E^*$  and the private system will produce the optimal result.

Landes and Posner have suggested that the common-pool nature of the supply of offenses may lead to inefficiencies under private enforcement. Under the institutions I have just described that problem disappears; particular offenses are private property, a possibility suggested by Landes and Posner, and the size of the total pool of offenses is not affected by the decisions of the enforcement firm since it is the court system, not the firm, that determines the expected punishment  $E$ , which in turn determines  $O$ , the output of offenses.

There is, however, another "commons" problem to be considered. A firm investigating one offense might come across evidence of another—hence it should regard all offenses as potential subjects of investigation. But under the institutions I have just suggested, different offenses belong to different firms, so a firm that comes across evidence relevant to someone else's offense cannot use it to produce an apprehension.

The obvious solution is for the firm that obtains the information to sell it to the firm that owns the relevant offense. The conventional arguments for imperfections in the market for information should not apply: the buyer has no incentive to resell since the information is of no use to anyone else, and the seller has an incentive to represent accurately what he is selling in order to maintain the reputation necessary for future sales. If the market functions perfectly, private property in offenses should not increase the cost of producing apprehensions.

But markets do not function perfectly; the transactions costs associated with transmitting such information among firms may be one of the costs of having many firms. If the advantages of replacing such transactions with equivalent transactions internal to the firm outweigh the diseconomies of larger scale operations, that will be a reason for firms to become larger; in the limiting case the result is a natural monopoly. The mechanism by which the firm produces  $E$  with the optimal mix of  $f$  and  $p$  works for a monopoly as well as a competitive industry, so an efficient system is still

possible. The organizational costs associated with a natural monopoly should also exist in a public monopoly enforcement agency; hence the private alternative is still as good as an ideal public system.

One point I have not so far considered is the value of  $P_v$ , the price paid to the victim to “buy the offense” (more precisely, to buy the victim’s claim against the criminal). To find  $P_v$  we use the zero profit condition and solve for  $P_v$ :  $\Pi = E \times O \times (1 - Z) - P_v O - rR = 0$

$$P_v = E(1 - Z) - (rR/O). \quad (1)$$

The second term on the right hand side of equation (1) is positive, hence  $Z > 1$  is a sufficient (but not necessary) condition for  $P_v$  to be negative. To put the same argument verbally, the price per offense is, in equilibrium, equal to the fine collected per offense minus the enforcement expenditure per offense. Since the fine collected may be negative and even if positive may be smaller than the expenditure, the price paid the victim may well be negative!

There appears to be a problem here; if  $P_v$  is negative, victims are paying firms to take their offenses. Why?

When I originally described the institutions, I stated that the offenses were purchased before or just after they occurred. If  $P_v$  is negative, they must be purchased before. An individual who “sells” future offenses against himself (for a negative price) is permitted to attach a medallion to his door warning burglars that if they rob him, they will be pursued by SureDeath Inc.<sup>18</sup> SureDeath is then obligated to impose the expected penalty  $E$  on those who rob their clients.<sup>19</sup>

<sup>18</sup> Such arrangements currently exist in several contexts, as discussed in David Friedman, *Private Creation and Enforcement of Law: A Historical Case*, 8 J. Legal Stud. 399 (1979), which also describes the workings of a real-world system of private enforcement in medieval Iceland. That system gave victims the property right in offenses; it imposed actual rather than expected punishments but also punished the criminal for actions which lowered  $p$ . “Murder,” for example, was defined as “secret killing” and punished more severely than “open killing.” The possibility of selling offenses for a negative price undercuts the proof of the inefficiency of private enforcement offered by Polinsky, *supra* note 9, at 113, which depended on the revenue from fines being insufficient to cover the costs of an optimal level of enforcement if the damage imposed by an offense were sufficiently high. If the damage is high, so is the amount potential victims are willing to pay firms to take charge of offenses against them. Such negative prices for offenses make sense, however, only for offenses where the criminal knows the identity of the victim (and hence whether the victim has sold offenses against him in advance) before he commits the crime. Polinsky’s argument is correct where that is not the case—reckless driving, for example. In such situations private enforcement leads to an efficient outcome only if the price of offenses implied by the zero-profit condition (for a competitive enforcement industry) is nonnegative.

<sup>19</sup> A completely private system (of courts, enforcement, and legislation) having some of the features of the system analyzed here is proposed in David D. Friedman, *The Machinery of Freedom* (1971). Terry L. Anderson & P. J. Hill, *An American Experiment in Anarcho-Capitalism: The Not So Wild, Wild West*, 3 J. Libertarian Stud. 9 (1979), discuss the American west in the nineteenth century as a historical example of such a system.

## IV. EFFICIENT LOCKS

I have so far ignored the possibility of controlling crime by defensive expenditures by potential victims—locks, burglar alarms, walls, etc. In considering such expenditures, an obvious question is whether the institutions I have described lead to an optimal level of defense as well as optimal levels of apprehension and punishment.

To answer this question, I must make some assumptions about the way in which defensive expenditures affect crime. The first question is whether the effect of such expenditures is to make apprehension easier or crime more expensive. In the former case enforcement firms will pay a higher price to victims who have made such expenditures, since they know that the cost of apprehension is lower; in equilibrium the increased price is equal to the reduction in cost, hence defensive expenditures will be made up to the point where a marginal dollar spent on defense reduces apprehension costs by a dollar. The case is analogous to fire insurance companies that give special rates to customers with sprinkler systems.

Consider next the case where defensive expenditures raise the cost to the criminal of committing the crime and so reduce  $O$ . In this case, one assumption necessary for a system in which the private interest of potential victims provides an optimal amount of defense is that the reduction in offenses produced by my defensive expenditures affects only offenses of which I am the victim. There are at least two reasons why this might not be the case. First, criminals might not be able to distinguish defended from undefended houses; if so the reduction in crime produced by my expenditure on defense is shared among all potential victims. In this case I have little incentive to make such expenditures.

The second case is the one in which the criminal, observing that my house is defended, robs your house instead. Here the externality is in the opposite direction; the reduction in offenses committed against me overstates the benefit produced by my defense; again it seems unlikely that I will purchase an optimal amount of defense.

Suppose, however, that criminals can costlessly tell which houses are defended (and how well) and further suppose that the resources used by criminals are available in unlimited supply at a constant price, just as the resources used in enforcement were assumed to be. In this case my defense defends only me and does not increase offenses against anyone else, so the reduction in offenses against me as a result of my defensive expenditure is equal to the reduction in total offenses.

I shall make two additional assumptions before comparing social and private optimal defense. The first assumption is that offenses are homogeneous so far as the damage done to the victim is concerned, with the harm

done by each offense equal to  $H$ . The second is that a marginal increase in defensive expenditure that reduces the number of offenses by 1 eliminates the same offense that would be eliminated by a marginal increase in  $E$  that reduced the number of offenses by 1. To put the assumption somewhat differently: I am assuming that  $D$ , the net damage done by  $O$  occurrences of the crime, is a function only of  $O$  and depends on  $E$  and the quantity of defensive expenditure  $W$  only through their effect on  $O$ . This makes sense if we imagine the criminal deciding to commit a particular offense according to whether the benefit to him is greater than the cost. An increase in cost, whether in the form of an increase in  $E$  or in  $W$ , eliminates the least attractive opportunities.

Having made our assumptions, we are now ready to consider private and public optimal defense. The cost of  $O_j$  occurrences of the crime to an individual victim is  $O_j(H - P_v)$ , since each occurrence inflicts damage  $H$  on him and can be sold for a price  $P_v$  (possibly negative). Hence he will increase his defensive expenditure  $W_j$  until  $d[W_j + O_j(H - P_v)]/dW_j = 0$ . Substituting in  $P_v$  from equation (1), we get

$$0 = 1 + \left( H + EZ + \frac{rR}{O} - E \right) \frac{dO_j}{dW_j}. \quad (2)$$

Note that  $rR/O$ , the enforcement expenditure per offense, does not depend on the number of offenses, since we have assumed that the industry production function exhibits constant returns to scale. The same argument implies that the optimal  $F$ , and hence the value of  $Z$ , does not change with  $W_j$ .

$E$  is the certainty equivalent of the punishment lottery faced by the criminal; it follows that  $E$  is also the value that the criminal expects to get by committing the marginal crime. Equation (2) tells us that the potential victim will increase  $W$  up to the point where the cost of a further increase is equal to the resulting decrease in the number of offenses times the sum of the cost of punishing an additional offense, plus the additional expenditure on apprehension necessary to apprehend the additional offense, with probability  $p$ , plus the harm done to the victim by an additional offense, minus the benefit to the criminal of the additional offense. But the cost of punishing an additional offense plus the cost of maintaining  $p$  when the number of offenses increases by one, plus the harm done by an additional offense, minus the benefit to the criminal of an additional offense is precisely the social cost of an additional offense; hence the potential victim is spending up to the point where the marginal cost of additional defense is equal to its marginal social benefit. It follows that the private system, under the assumptions I have made, produces the optimal level of defense!

## V. PROBLEM: NEW BRIBERY INCENTIVES

The original argument made by Becker and Stigler for a private enforcement system was that setting the compensation of the enforcer equal to the cost imposed on the criminal eliminated the incentive for bribes. Landes and Posner pointed out that if the state taxed the return to the enforcement firms, or if the state or victim charged the firms for the right to collect the fine, the result would be to reintroduce the incentive for bribery. In the system I have described the firm buys the offense before or just after it occurs, hence the purchase price is a sunk cost by the time the criminal is located. The amount the firm receives is less than the amount the criminal pays, but only by the collection cost. It seems as though bribery should be no problem.

Unfortunately, that is not the case. The system as so far described contains two incentives for bribery. The first depends on the criminal's being better at extracting funds from himself than the courts are and so being willing to offer a bribe lower than the cost of punishment to him but higher than the amount the firm would receive. In some ways this is desirable—it reduces the inefficiency of punishment. But it also reduces the expected punishment below the optimal level set by the courts.

The second incentive for bribery comes from the constraints on the firm. Suppose a particular firm has purchased a thousand offenses, each of which is to receive an expected punishment of a thousand dollars, and has collected sufficient evidence to apprehend and convict ten criminals. If it does so, it must impose a punishment of one hundred thousand dollars on each. Assume that punishment inefficiency is .5; the firm actually collects an average of fifty thousand dollars from each criminal, the rest of the punishment taking the form of flogging, imprisonment, and so forth.

Now suppose the firm, for a bribe of forty thousand dollars, destroys the evidence against one of the criminals. At first sight this seems foolish, since it would have received fifty thousand by convicting him. But since bribes are not reported to the court system as punishments, the firm can now impose punishments of \$111,000 on each of the other criminals— $p$  (as measured by the court system) has gone down, since nine instead of ten criminals have been caught, so  $F$  goes up in order to keep the expected punishment (as observed by the court system),  $E = pF$ , constant.

To make the argument more precise, let us assume away the first incentive for bribes by supposing that the inefficiency for bribes is the same as for punishments. Now consider a firm which accepts a bribe equal to  $F'$ , the amount it would have received by convicting the criminal. The only difference between accepting the bribe and convicting the criminal is that since the bribe is not reported to the courts the firm can increase the  $F$  it

imposes on other convicted criminals. If increasing  $F$  results in increasing  $F'$ , the firm gains; it will, if necessary, be willing to divide the gain with the criminal by accepting a bribe somewhat lower than the amount it would have collected by convicting him. The effect of its accepting such bribes is that expected punishment  $E$  (including the cost to criminals of both bribes and court administered punishments) is higher than the optimal level set by the court.

Consider the alternative case, where an increase in  $F$  decreases  $F'$ . Suppose, for example, the punishment consists of the largest fine the criminal can pay plus some imprisonment;  $F'$  is the fine minus the cost of keeping him in prison. Increasing  $F$  means lengthening the imprisonment, which decreases  $F'$ . In this case the firm would prefer to impose lower punishments but is forbidden to do so (unless it catches a larger fraction of the criminals), because that would lower  $E$ .

This situation also provides an opportunity for bribery. The firm passes the word in criminal circles that if a criminal who has committed one of the offenses the firm has bought turns himself in, the firm will pay his fine for him and give him a bonus as well. In the previous case, the firm was paying to make  $p$  (and hence  $E$ ) appear smaller than it was by concealing a "conviction"; in this case it is paying to make  $p$  appear larger than it is by manufacturing a "conviction."

In the cases discussed by Becker and Stigler and Landes and Posner, bribery was a way in which the enforcer got money which would otherwise have gone to the court system. In the cases I have just been describing, bribery is a way in which the enforcer relaxes the expected punishment constraint imposed by the court. If his revenue increases with increasing expected punishment, he, in effect, bribes criminals who have been convicted and punished to pretend they have not (by letting them pay a somewhat smaller punishment directly to the enforcer instead of through the court system); if his revenue increases with decreasing expected punishment, he bribes criminals who have not been punished (since their fine has been paid by the enforcer) to pretend they have been. In both cases the expected punishment imposed is different from the optimal level set by the court system.

## VI. SOLUTION: OBSERVING EXPECTED PUNISHMENT

I have described three sorts of bribery that might exist in the proposed system of private enforcement. The first depends on bribes' having lower collection costs than fines; it is undesirable only because of its effect on expected punishment. The second and third are ways in which the enforcement firm makes the expected punishment higher or lower than it appears to be. All three are ways in which the system breaks down be-



cause the court system is unable to observe, and hence unable to control, the variable (expected punishment) it uses to produce optimal enforcement. All that is needed to eliminate the problems associated with bribery under the proposed system is some way for the court system to observe the level of expected punishment imposed by each enforcement firm.

There is an easy and inexpensive way by which the court system can do so; watching the criminals. The purpose of punishment, after all, is to deter crime, and the amount of punishment can be measured by the amount of deterrence. Earlier, when discussing the possibility that offenses might sell for a negative price, I suggested that offenses could be sold before they occur under arrangements in which a potential victim puts a medallion on his door telling the criminal which firm he has sold future offenses to. If a firm is using bribes to impose a punishment higher (or lower) than that set by the court, the result will be a lower (or higher) than normal crime rate against its customers.<sup>20</sup> The court system need only observe the rate at which crimes occur against the customers of each firm. If the rate is consistently "too" low then the firm should be instructed to lower its expected punishment; if "too" high, to raise it. If, as assumed, the court system has the information necessary to set an optimal level of expected punishment, it will know whether  $F'$  is an increasing or decreasing function of  $F$  and hence which sort of cheating to watch out for.

With the court system measuring  $E$  directly by the behavior of the criminals instead of indirectly by observations of  $p$  and  $f$ ,<sup>21</sup> bribery intended to distort the value of  $p$  no longer serves any function. Bribery which exists because it is more efficient than fines remains, but since it does not affect expected punishment it is no longer undesirable. We are back at an efficient system.

## VII. SUMMARY

Landes and Posner described a private enforcement system, in which the court sets the punishment to be imposed on convicted criminals and the enforcement firm receives an amount equal to that punishment, and demonstrated that such a system is inefficient. The reason is that an

<sup>20</sup> This assumes that criminals know which firms accept bribes or fabricate convictions and hence have higher or lower than normal expected punishments, which seems plausible.

<sup>21</sup> I have assumed away statistical problems, nonhomogeneity of victims, and the like. Within the constraints of the model, where the court system is assumed to know the supply function of offenses, the production functions of the firms, and everything else necessary to calculate the optimal expected punishment, this is reasonable. In a more realistic setting the court would set  $E$ , enforce it in terms of permitted punishments given observed values of  $p$ , and monitor crime rates in order to spot the firms that ought to be investigated on suspicion of bribery.

optimal system must produce the optimal disincentive to the criminal in the form of an optimal combination of probability of being convicted and punishment if convicted. Under the private system they describe, the fine imposed by the court is both the incentive to the enforcer and part of the disincentive to the criminal; since the optimal level of the two is not in general the same, no fine can produce (save by chance) the optimal outcome.

The problem is eliminated by requiring the court to set the expected punishment rather than the actual punishment and making the reward to the firm the punishment net of collection costs—the fine collected, not the fine paid. Since the firm has no control over expected punishment its actions have no effect on the total output of offenses; that is decided by the court system when it sets the expected punishment. Since the two costs that enter into the choice of an optimal combination of punishment and probability (for a given expected punishment), enforcement cost and net punishment cost, are now internal to the firm, its own profit-maximizing behavior automatically generates the optimal combination.

In setting up the optimal private system, I abstracted away from many of the complications of the real world. I considered only a single crime, ignored problems associated with the conviction of innocent parties, and ignored costs borne by potential victims trying to protect themselves and criminals trying to avoid capture. I assumed that the public court system would have the information necessary to set an optimal level of expected punishment and an adequate incentive to do so. These are, however, the same simplifications employed in the Landes-Posner argument. In one respect I have made my model more realistic than theirs. They assume a maximum fine (which can be collected costlessly) with no higher punishments possible; I merely assume that collection cost increases as punishment increases. Their assumption is a special case of mine.

Within those assumptions a set of private enforcement institutions was described that would produce the same result as a perfect public system. Having demonstrated that, I went on to investigate the consequence of dropping one of the simplifying assumptions (shared by most writers in this area) by allowing for defensive expenditures by potential victims. The conclusion was that if expenditures by a potential victim affected only offenses against him and not offenses against others, the private system would generate the optimal amount of defense.<sup>22</sup> This result is especially encouraging since the possibility of defense was ignored in constructing

<sup>22</sup> I made some other assumptions designed to simplify the analysis rather than to eliminate obvious sources of inefficiency; I suspect but have not proved that the conclusion would survive the elimination of those assumptions, provided that the optimal public system to which the private system is compared is required to deal with the same complications.

the original model and the institutions designed to produce optimal private enforcement within that model. There seems no obvious reason why a more complicated private system could not deal with the other complications that were assumed away in the analysis.

While the system of private enforcement that was described eliminates a number of the problems discussed in the literature, it also introduces some new ones associated with the court system's need to observe the expected punishment imposed by each protection firm. One solution is for the courts to deduce the expected punishment imposed by each firm from observing the behavior of the criminals; levels of expected punishment higher or lower than the level set by the courts should result in lower- or higher-than-expected crime rates against the customers of the corresponding firm.

This suggests the possibility of some further modification of the institutions of private enforcement, in which the effect of expected punishment on crime rates would become part of the market incentive system within which the enforcement firms operate rather than a device used by the court system to detect cheating by the firms. The private enforcement system so far described substantially reduces the information requirements of the court system relative to the requirements under a system of public enforcement, since minimization of the sum of enforcement and collection costs is produced by the self-interest of the firms, and the resulting cost for any given level of expected punishment can be observed by the court system. It would be interesting to try to construct a system in which all the court had to determine was guilt or innocence, with the entire structure of optimal punishment determined by the rational behavior of the participants under an appropriate set of legal rules.

## APPENDIX

### THE OVERENFORCEMENT THEOREM

Landes and Posner assert that under the institutions they describe "the 'best' one can do under private enforcement is to set a fine equal to  $f_1$ , but at  $f_1$  one observes a greater probability of apprehension and conviction ( $p_1 > p_0$ ) and a greater social loss under private than optimal public enforcement."<sup>23</sup> They add in a footnote that "[t]his overenforcement theorem may not hold if the optimal fine under private enforcement is the corner solution  $f^*$ . It is conceivable that  $f^*$  may be sufficiently small relative to enforcement costs so that the positively sloped

<sup>23</sup> Landes & Posner, *supra* note 2, at 14–15 (footnote omitted).

$PR_1$  curve is below the  $\partial L/\partial p = 0$  curve at  $f^*$ . . . . We disregard these possibilities in the subsequent analysis."<sup>24</sup> They then go on to discuss an "intuitive explanation for the overenforcement theorem. . . ."

What Landes and Posner say is correct, but to describe it as an overenforcement theorem is misleading. The situation is illustrated in Figures 1a and 1b, which are similar to their Figure 3.<sup>25</sup> In Figure 1a,  $PR$  is the set of points  $(p, f)$  consistent with private enforcement; it is easily shown that along  $PR$  increasing  $f$  implies increasing  $p$ ;  $f_m$  is the maximum fine—under Landes and Posner's assumptions it can be collected costlessly, and no higher punishment is possible;  $(p^*, f^* = f_m)$  is the optimal combination of probability and fine.  $I_1, I_2,$  and  $I_3$  are social loss indifference curves (SLICs).  $V$  is a line through the vertical points of the SLICs; it is defined by the condition  $\partial L/\partial p = 0$ , where  $L$  is the social loss function. For reasons explained in Landes and Posner,<sup>26</sup> along  $V$  an increase in  $f$  implies a decrease in  $p$ .

At the point  $X$ ,  $PR$  is tangent to  $I_2$ ; since  $PR$  is the "opportunity set" available to the court that sets  $f$  under a private enforcement system and  $I_2$  is the highest SLIC it touches,  $X$  is the optimal  $(p, f)$  under private enforcement. Since  $PR$  slopes up and to the right, it must be tangent to the upper part of  $I_2$ ; since  $V$  slants up and to the left, this implies  $p_1 > p_0 > p^*$ . The probability of conviction under optimal private enforcement is greater than the probability that would be optimal for the same fine under public enforcement, and also greater than  $p^*$ , the optimal probability (with fine and probability both free to vary) under public enforcement. This is the proof of the "overenforcement theorem."

The argument shows there is overenforcement provided that  $PR$  is tangent to some SLIC. Figure 1b shows the case where it is not. The situation is exactly the same, except that the optimum point is now at  $(p^{*'}, f^{*'} = f_m)$ .<sup>27</sup> Since  $PR$  passes below  $(p^{*'}, f^{*'})$ , its optimum occurs when it hits the boundary  $f = f_m$ . This is the "corner solution" mentioned by Landes and Posner in a footnote.

The reason this is not an overenforcement theorem is that nothing in the argument implies that the situation shown in Figure 1a is any more likely than that shown in Figure 1b; hence we have no reason to believe that overenforcement is any more likely than underenforcement. All Landes and Posner have shown is that the situation that leads to underenforcement also leads to a corner solution.

Even this result disappears once we drop the assumption that there is a maximum punishment that can be imposed costlessly. Figures 2a and 2b show the more general case where there is no maximum punishment but the ratio of the cost of punishment to the size of the punishment is an increasing function of the latter. With the barrier at  $f = f_m$  eliminated, both underenforcement and overenforcement are consistent with interior solutions.

<sup>24</sup> *Id.* at 15, n. 32.

<sup>25</sup> *Id.*

<sup>26</sup> *Id.*

<sup>27</sup> Since the location of  $PR$  is not affected by either the supply of offenses function  $O$  or the damage function  $D$ , both of which affect  $(p^*, f^*)$ , the relative position of the two can easily be altered by altering  $O$  or  $D$ . Since the argument of Landes and Posner does not assume any particular form for  $O$  and  $D$  (except for the sign of their derivatives), Figures 1b and 2b are just as plausible as Figures 1a and 2a.