Journal of Economic Literature
Vol. XXXVIII (March 2000) pp. 45–76

The Economic Theory of Public Enforcement of Law

A. MITCHELL POLINSKY AND STEVEN SHAVELL[1]

1. *Introduction*

PUBLIC ENFORCEMENT of law—the use of public agents (inspectors, tax auditors, police, prosecutors) to detect and to sanction violators of legal rules—is a subject of obvious importance. Enforcement policy affects, for example, the amount of pollution that firms generate, the extent of compliance with the income tax code, and the incidence of theft, robbery, and other crimes.

The earliest economically oriented writing on the subject of law enforcement dates from the eighteenth century contributions of Montesquieu (1748), Cesare Beccaria (1767), and, especially, Jeremy Bentham (1789), whose analysis of deterrence was sophisticated and expansive. Curiously, after Bentham, the subject of enforcement lay essentially dormant in economic scholarship until the late 1960s, when Gary S. Becker (1968) published a highly influential article. Since then, well over two hundred articles have been written on the economics of enforcement.[2]

The main purpose of our article is to present the economic theory of public enforcement of law in a systematic and comprehensive way. The theoretical core of our analysis (Sections 2 through 4) answers the following basic questions: How much of society's resources should be devoted to apprehending violators? If a violator is caught, should the rule of liability be strict or fault-based? Should the form of the sanction be a fine, an imprisonment term, or a combination of the two? At what level should sanctions be set?[3]

We then consider (Sections 5 through 16) a number of additional questions in deterrence theory, including: Should

[1] Stanford Law School and National Bureau of Economic Research; and Harvard Law School and National Bureau of Economic Research. Research on this article was supported by The Lynde and Harry Bradley Foundation, the John M. Olin Foundation, the John Simon Guggenheim Memorial Foundation, the Center for Advanced Study in the Behavioral Sciences (through National Science Foundation grant #SBR–960123), and the law and economics programs at Stanford Law School and Harvard Law School. We received able research assistance from Alexander Gould, Janine Scancarelli, and Timothy Stanley, and helpful comments from Daniel Klerman, participants at various seminars, and two anonymous referees. We also want to thank John Pencavel and John McMillan in their capacities as editors of the *JEL*. We are at work on a book-length treatment of the public enforcement of law.

[2] See, for example, the references cited in Boudewijn Bouckaert and Gerrit De Geest (1992, pp. 504–26), Nuno Garoupa (1997), and Dilip Mookherjee (1997).

[3] Our analysis of these questions is applicable in many important respects to firms and other private organizations that attempt to enforce their own internal rules using their own enforcement agents. One significant difference, however, is that private organizations generally cannot employ nonmonetary sanctions, such as imprisonment.

sanctions be increased to reflect society's cost of punishing violators? How should a system of enforcement be modified to account for mistakes made in the enforcement process? How should sanctions be structured so as to discourage a person from committing a more harmful rather than a less harmful act? What are the implications for enforcement theory, if any, if the wrongdoer is a corporation or organization as opposed to an individual? How does the settlement process (including plea bargaining) relate to the enforcement system? Should sanctions be reduced if violators report their own infractions to the enforcement authority before the authority catches them?

Finally, we summarize our main points (Section 17), consider public enforcement in practice in the light of the theory of optimal enforcement (Section 18), and discuss several topics that we think merit further inquiry (Section 19).[4]

Before proceeding, we should comment on the rationale for public, as opposed to private, law enforcement. An important element of the rationale concerns information about the identity of violators. When victims of harm naturally possess knowledge of who injured them, allowing private suits for harm will motivate victims to initiate legal action and thus will harness the information they have for purposes of law enforcement. This may help to explain why, for example, the enforcement of contract law and tort law is primarily private in nature. When victims do not know who injured them, however, and when identifying (or apprehending) violators is difficult, it may be desirable for public enforcement to be employed. For public enforcement to be preferred in these circumstances, one still needs to explain why society cannot rely on inducements to private parties (rewards of some type) to supply information and otherwise help in detecting violators. A difficulty with reliance on private enforcement is that if a reward is available to everyone, there might be wasteful effort devoted to finding violators (akin to excessive effort to catch fish from a common pool). Another problem is that private parties may find it hard to capture fully the benefits of developing expensive, but socially worthwhile, information systems (for instance, computerized databases of fingerprint records); such enforcement technologies may constitute natural monopolies. An additional obstacle to private enforcement is that force may be needed to gather information, capture violators, and prevent reprisal, yet the state frequently, if not usually, will not want to permit private parties to use force. For the preceding reasons, public enforcement often will be favored when effort is required to identify and apprehend violators.[5]

2. *The Basic Framework*

In this section we describe individual behavior, social welfare, and the enforcement authority's problem.

[4] Our treatment of the theory of public enforcement overlaps in some significant respects with recent surveys by Garoupa (1997) and Mookherjee (1997). However, their emphases are different: Garoupa focuses on reasons why optimal sanctions are not maximal, and Mookherjee concentrates on the control of tax evasion.

[5] The social justifications for public versus private enforcement have been discussed by Gary S. Becker and George J. Stigler (1974), William M. Landes and Richard A. Posner (1975), and Polinsky (1980a); see also David D. Friedman (1995) and Shavell (1993). In our analysis, we assume for simplicity that public enforcement is the exclusive means of enforcement, even though in practice private parties sometimes play a complementary role by supplying information to enforcement authorities and by bringing private suits. We also abstract from private parties' efforts to protect themselves from harm (and how such efforts might relate to public enforcement); on that topic, see comment (c) in Section 19 below.

2.1 *Individual Behavior*

Consider an individual who would obtain a gain from committing a harmful act. If he does commit it, he will be caught with some probability and then possibly have to pay a fine or go to jail, or both.[6] In general, he will commit the act if and only if his expected utility from doing so, taking into account his gain and the chance of his being caught and sanctioned, exceeds his utility if he does not commit the act.[7]

For simplicity, we focus on the assumption that individuals are risk neutral in fines and in imprisonment. The interpretation of risk neutrality with respect to fines is familiar; its interpretation with respect to imprisonment is that individuals' disutility from imprisonment rises proportionally with the length of the term. We also comment, however, on how our results differ if individuals are risk averse in fines or in imprisonment, or risk preferring in imprisonment.[8] Risk aversion in imprisonment (preference for a certain jail term over a risky term of the same expected length) results if the disutility of jail rises more than proportionally with the length of the term—say an individual can tolerate reasonably well a month in jail but not a year. Conversely, risk preference in imprisonment (preference for a risky jail term over a certain term equal to the risky term's expected value) arises if the disutility of jail increases less than proportionally with the length of the term—for example, the disutility from the stigma of being in jail might be substantial from having spent any time there at all, but not increase much with the length of imprisonment.[9] Individuals' discounting of the future disutility of imprisonment also makes earlier years of imprisonment more important than later ones.

Whether an injurer who has been caught will be sanctioned depends on the rule for imposing liability. Under strict liability, a sanction is imposed on the injurer regardless of his behavior. For example, a firm that spills oil might be held liable for the harm caused by the spill even if it took reasonable precautions to prevent the spill. Under fault-based liability, a sanction is imposed only if the injurer's act is determined to be socially undesirable. Automobile drivers, for instance, are sanctioned if they exceed the speed limit, but not if they drive within the limit.

To state matters formally, let

g = gain a party obtains from engaging in the harm-creating activity;
p = probability of detection;
f = fine;
t = length of the imprisonment term; and
λ = disutility borne by a prisoner per unit of the imprisonment term.

Then, under strict liability, a risk-neutral individual will commit the harmful act if and only if his gain from doing so exceeds the sum of the expected fine

[6] A jail term is just one example of a nonmonetary sanction. It will be clear that most of what we have to say about imprisonment—both here and in subsequent sections—would carry over, with only slight modification, to other forms of nonmonetary sanctions, such as probation, electronic monitoring, community service, and, in the extreme, the death penalty. See, for example, note 21 below.

[7] This statement about individual behavior of course reflects the notion of deterrence. For general discussion and analysis of deterrence, see, for example, Johannes Andenaes (1966), James Q. Wilson and Richard J. Herrnstein (1985), and Franklin E. Zimring and Gordon J. Hawkins (1973); and for a recent survey of empirical work on deterrence, see Erling Eide (1997). See also comment (d) in Section 19 below.

[8] We do not consider the assumption that individuals are risk preferring in fines—equivalently, that their marginal utility of wealth is increasing in wealth—but it would be straightforward to do so.

[9] Also, the first years of imprisonment may create special disutility due to brutalization of the prisoner.

and the expected disutility of the imprisonment term:[10]

$$g > p(f + \lambda t). \quad (1)$$

If the individual is risk averse in fines and/or imprisonment, his gain would have to be higher than that indicated in (1) before he would commit the harmful act; and if he is risk preferring in imprisonment, the requisite gain would tend to be lower.

Under fault-based liability, as noted above, an individual who causes harm will be held liable only if his act is determined to be socially undesirable. In our framework, this means that he will be held liable if he committed the harmful act when his gain was relatively low. We will refer to this critical level of gain as the *fault standard:*

$\hat{g}$ = fault standard.

Thus, if an individual commits the harmful act when his gain is less than $\hat{g}$, he will be said to be at fault and will be found liable; otherwise he will not be liable.[11] When considering fault-based liability, we assume that the enforcement authority can accurately observe the individual's gain, so that it can determine whether he was at fault. Obviously, if an individual's gain equals or exceeds $\hat{g}$, he will engage in the harmful activity because he will not be found at fault. If his gain is less than $\hat{g}$, then because he will be found at fault and made liable if he is caught, he will commit the harmful act if and only if (1) holds.

2.2 *Social Welfare*

Social welfare generally is presumed to equal the sum of individuals' expected utilities. An individual's expected utility depends on whether he commits a harmful act, on whether he is sanctioned, on whether he is a victim of someone else's harmful act, and on his tax payment, which will reflect the costs of law enforcement, less any fine revenue collected.[12] If individuals are risk neutral, social welfare can be expressed simply as the gains individuals obtain from committing their acts, less the harms caused, and less the costs of law enforcement. We assume, as is conventional, that fines are socially costless because they are mere transfers of money,[13] while imprisonment involves positive social costs because of the expense associated with the operation of prisons and the disutility due to imprisonment (which is not naturally balanced by gains to others).[14]

To state social welfare more precisely,

[10] We assume without loss of generality that he does not commit the harmful act if he is indifferent between committing it and not committing it.

[11] In practice, fault is often found if an individual did not take reasonable precautions to prevent harm, where a reasonable precaution is one whose cost is less than the harm that it prevents. Our characterization of fault-based liability is consistent with this practice if gain is interpreted as the savings an individual obtains from not taking a precaution.

[12] Note that because an individual's utility rises if he obtains a gain from committing a harmful act, his gain counts in social welfare. Some writers have questioned whether such gains should be credited in social welfare, especially when the gains derive from criminal acts. See, for example, Stigler (1970, p. 527) and Jeff L. Lewin and William N. Trumbull (1990). We include all gains in social welfare, as is conventional in the literature on enforcement. We also adopt the usual practice of not considering the disutility that individuals might suffer as a result of knowing about harms to others (such as the disutility from learning that someone was raped). If the gains from some type of harmful conduct were excluded from social welfare, or the disutility borne by non-victims were included, the main consequence for our analysis would be that, for this type of conduct, society would want to achieve greater, possibly complete, deterrence. That, in turn, would tend to make a higher sanction and a higher probability of detection desirable.

[13] In practice, of course, some costs are incurred in collecting fines, such as the cost of identifying and confiscating the injurer's assets if the injurer resists paying the fine. We discuss the implications of such costs in Section 6.

[14] It may be that victims (and perhaps individuals generally) obtain some utility from seeing that injurers are punished. See Posner (1980b). Were we to include this possible component of utility, the social cost of punishment would be lower than we presume, making the imposition of sanctions, including imprisonment, more desirable. See generally Polinsky and Shavell (1998b).

assume that individuals differ in the gain they obtain from committing the harmful act. There will be a *critical gain* above which individuals will commit the harmful act and below which they will be deterred. The critical gain is determined by the probability of detection, the level of sanctions, and the standard for imposing liability, as will be seen below.

Let

$z(g)$ = density of gains among individuals;
$Z(g)$ = cumulative distribution of $z(.)$;
$\tilde{g}$ = critical gain;
h = harm caused by an individual if he commits the harmful act;
α = cost to the public per unit of the imprisonment term;
e = enforcement expenditures by the government; and
$p(e)$ = probability of detection given e ($p' > 0, p'' < 0$).

The population is normalized to equal unity and the harm is assumed to be monetary.

Under strict liability, if individuals are risk neutral, social welfare can be expressed as:

$$\int_{\tilde{g}}^{\infty} gz(g)dg - [1 - Z(\tilde{g})](h + pt(\lambda + \alpha)) - e, \quad (2)$$

where

$$\tilde{g} = p(e)(f + \lambda t). \quad (3)$$

The first term in (2) is the aggregate gain obtained by those who commit the harmful act. The second term is the aggregate harm caused by such individuals, plus the disutility suffered by the subset of them who are caught and put in jail, plus the cost to the public of keeping them in jail. The last term is the public's enforcement costs.

Under fault-based liability, recall that if an individual's gain equals or exceeds $\hat{g}$, he will engage in the harmful activity because he will not be found at fault, while if his gain is less than $\hat{g}$, he will engage in the activity if and only if (1) holds. In other words, under fault-based liability the critical gain is the lesser of $\hat{g}$ and the right-hand side of (1). Hence, social welfare under fault-based liability when individuals are risk neutral is:

$$\int_{\tilde{g}}^{\infty} gz(g)dg - [1 - Z(\tilde{g})]h - [Z(\hat{g}) - Z(\tilde{g})]pt(\lambda + \alpha) - e, \quad (4)$$

where

$$\tilde{g} = \min[\hat{g}, p(e)(f + \lambda t)]. \quad (5)$$

Note that the third term in (4) reflects the private and public costs associated with imprisonment only for individuals who are caught and found to be at fault.

2.3 *The Enforcement Authority's Problem*

The enforcement authority's problem is to maximize social welfare by choosing enforcement expenditures, e (or, equivalently, the probability of detection p), the level of the fine, f, the length of the imprisonment term, t, and the standard for imposing liability.[15] If the authority chooses fault-based liability, it also must choose the fault standard, $\hat{g}$. We use an asterisk to denote the optimal values of these variables.[16]

[15] In practice, it may not be possible to set the probability of imposing sanctions independently of their level. This is because high sanctions may lead juries to be less likely to convict defendants, or may induce injurers to engage in greater efforts to avoid detection; on these points, see James Andreoni (1991) and Arun S. Malik (1990), respectively.

[16] The framework for studying public enforcement presented in this section derives, in many respects, from Bentham (1789). Becker (1968) first stated the enforcement problem in formal economic terms and added the choice of the probability of detection to Bentham's expression of the enforcement problem. For some examples of the analysis of public enforcement that use essentially the framework presented here, see Polinsky and Shavell (1984) and Mookherjee and I. P. L. Png (1994).

3. *Optimal Enforcement with a Fixed Probability of Detection*

In this section, we consider optimal enforcement given the assumption that enforcement expenditures are fixed, say at $\bar{e}$, and thus that the probability of detection is fixed at $p(\bar{e})$. For convenience, we refer to this probability simply as p. (The probability will be treated as a policy instrument in the next section.)

3.1 *Strict Liability*

Fines alone are considered first, then imprisonment alone, and finally the two sanctions together.

3.1.1 *Fines*

Social welfare in the present case is given by (2) with the imprisonment term t equal to zero and $e = \bar{e}$. Thus, the critical gain $\tilde{g}$ equals pf; that is, individuals commit the harmful act if and only if their gain exceeds the expected fine. Note that the level of the fine does not affect social welfare directly because we are assuming that fines are socially costless to impose; the fine does affect social welfare indirectly, however, by determining which individuals commit the harmful act—that is, by determining $\tilde{g}$.

Taking the derivative of social welfare with respect to the fine f, setting the result equal to zero, and solving for f gives the optimal fine when individuals are risk neutral:

$$f^* = h/p. \tag{6}$$

Therefore, $pf^* = h$, the expected fine equals the harm. Hence, individuals commit the harmful act if and only if their gain exceeds the harm, which is first-best behavior.[17] (It is important to note, however, that there will be underdeterrence if the injurer's wealth is less than h/p, in which case the optimal fine equals the injurer's wealth.)

If individuals are risk averse, the optimal fine tends to be lower than in the risk-neutral case for two reasons. First, lowering the fine reduces the bearing of risk by individuals who commit the harmful act. Second, because risk-averse individuals are more easily deterred than risk-neutral individuals, the fine does not need to be as high as before to achieve any desired degree of deterrence.[18]

3.1.2 *Imprisonment*

Social welfare now is given by (2) with $f = 0$ and $e = \bar{e}$, so that $\tilde{g} = p\lambda t$ — individuals commit the harmful act if and only if their gain exceeds the expected disutility of the imprisonment sanction. The difference from the case of fines is that now the disutility of those who are sanctioned, as well as the public's cost of imposing the sanction, reduce social welfare (by the amount $[1 - Z(\tilde{g})]pt(\lambda + \alpha)$).

There is not a simple formula for determining the optimal imprisonment term. The optimal term might be zero because imprisonment is socially costly to impose. If the optimal term is positive, it could be such that the critical gain $\tilde{g}$ is less than or greater than harm h—there is underdeterrence or overdeterrence relative to socially ideal behavior.

To explain, suppose initially that the imprisonment term is set so that $\tilde{g} = h$. Then if t is lowered or raised slightly, there is no first-order effect on social welfare in terms of the net effect of gain and harm, because the individuals whose behavior is influenced by the

[17] The general formula (6), or its equivalent, apparently was first put forward by Bentham (1789, p. 173), was emphasized by Becker (1968), and has been noted by many others since then.

[18] It is possible, however, that the optimal fine is higher in the risk-averse case than in the risk-neutral case, for the following reason. A way to reduce the bearing of risk is to deter more individuals from committing the harmful act, for then fewer individuals will be subject to the risk of the fine. See Louis Kaplow (1992).

change in t have gains just equal to the harm. But lowering t might reduce the aggregate disutility of imprisonment sanctions, as well as the public cost of the prison system, because individuals are imprisoned for a shorter duration (although more individuals will commit the harmful act and be sanctioned). Alternatively, raising t might reduce the aggregate disutility and public cost of imprisonment because fewer individuals will commit the harmful act and be sanctioned (although those who are sanctioned will serve longer sentences). Either effect could raise social welfare, meaning that the optimal imprisonment term could be such that $\tilde{g}$ is less than or greater than h.[19] (The possible optimality of overdeterrence strikes us as more theoretical than real, however.)

If individuals are risk averse in imprisonment, a given level of deterrence can be achieved with a lower sanction than in the risk-neutral case. This makes an imprisonment sanction more desirable because the cost to society of accomplishing deterrence is lower. Conversely, if individuals are risk preferring in imprisonment, an imprisonment sanction is less desirable than in the risk-neutral case.[20]

3.1.3 *Fines and Imprisonment*

When fines and imprisonment can be imposed together, social welfare is given by (2), and our principal point is that fines should be employed to the maximum extent feasible before resort is made to imprisonment. In other words, it is not optimal to impose a positive imprisonment term unless the fine is maximal. The rationale for this conclusion is that fines are socially costless to impose, whereas imprisonment is socially costly, so deterrence should be achieved through the cheaper form of sanction first.[21]

To elaborate, let

f_m = maximum possible fine.

The fine might be bounded for any number of reasons—the limited wealth of individuals, considerations of fairness, and so forth. It is easily seen that if f is less than f_m and t is positive, social welfare can be increased by raising the fine. Specifically, raise f and lower t so as to keep $f+\lambda t$ constant. Since $\tilde{g}=p(f+\lambda t)$ is not affected, the first term in (2) remains the same. The second term declines, however, because t has fallen. (The third term is fixed at $\bar{e}$, given present assumptions). Hence, social welfare rises. This type of argument applies quite generally: with slight variation it can be used to show that imprisonment should not be imposed unless fines are maximal regardless of individuals' risk preferences with respect to wealth or to imprisonment.[22]

When the optimal fine is maximal, it may or may not be desirable to impose

[19] The point that underdeterrence or overdeterrence may result when the imprisonment term is optimally chosen was made by Polinsky and Shavell (1984, p. 94). See also Friedman (1981, pp. 187–88, 192) and Kaplow (1990b).

[20] This discussion of the optimal imprisonment term implicitly presumes that the only function of imprisonment is deterrence. In practice, of course, imprisonment also serves the important function of incapacitating individuals who, were they not in jail, would be likely to commit additional crimes. Because the enforcement literature focuses almost exclusively on the deterrence rationale for imposing sanctions, we defer detailed discussion of the incapacitation rationale for imprisonment to Section 16.

[21] This basic point is made by Bentham (1789, p. 183), Becker (1968, p. 193), Posner (1980a), and Polinsky and Shavell (1984, pp. 95, 98). More generally, different types of sanctions should be employed in the order of their costs (per unit of deterrence). Thus, for example, electronic monitoring should be used before imprisonment is used (assuming the latter is more expensive per unit of deterrence).

[22] The argument is so general because, regardless of individuals' attitude toward risk in either fines or imprisonment, if f is less than f_m, f can be raised and t lowered so as to keep deterrence constant. This will always have the effect of lowering the costs of imprisonment.

an imprisonment sanction. Whether such a sanction should be imposed depends on whether the benefit of the additional deterrence thereby achieved is worth the additional cost of imprisonment.

3.2 *Fault-Based Liability*

3.2.1 *Fines*

Social welfare now is equal to (4) with $t = 0$ and $e = \bar{e}$, so that $\tilde{g} = \min[\hat{g}, pf]$.

It is easily seen that an optimal policy is to set the fault standard equal to the harm and the fine at a level that achieves compliance with the standard: $\hat{g}^* = h$ and $f^* \geq h/p$. Given this combination, it is clear that $\tilde{g} = h$, meaning that individuals commit the harmful act if and only if their gain exceeds the harm, the first-best outcome.[23] In the risk-neutral case, therefore, fault-based liability and strict liability are equally desirable. Moreover, the same fine that is optimal under strict liability—namely, h/p—will lead to compliance with the fault standard when the standard is optimally selected (any higher fine will too).

If individuals are risk averse, they are more easily deterred than if they are risk neutral, so the fine does not need to be as high to induce compliance with the fault standard. Furthermore, no one actually is sanctioned if compliance occurs, because no one is found at fault (assuming, as we do here, that there are no mistakes[24]).

Fault-based liability therefore may be preferable to strict liability when individuals are risk averse: Fault-based liability can accomplish desired deterrence of harm-creating conduct without imposing risk on risk-averse individuals, whereas under strict liability, individuals who commit the harmful act bear the risk of being fined.[25]

However, to choose properly between fault-based liability and strict liability also requires consideration of several other factors. First, fault-based liability is more difficult to administer. Under strict liability, the authority need only determine harm, whereas under fault-based liability, it must also be able to calculate optimal behavior and to ascertain actual behavior. Second, for reasons we discuss in Section 7, strict liability encourages better decisions by injurers regarding their level of participation in harm-creating activities. Third, fault-based liability will result in fewer enforcement actions than strict liability and thereby save enforcement costs: injurers who clearly were not at fault presumably would not be prosecuted under fault-based liability, but they would be under strict liability.

3.2.2 *Imprisonment*

Social welfare now is equal to (4) with $f = 0$ and $e = \bar{e}$; thus, $\tilde{g} = \min[\hat{g}, p\lambda t]$.

The optimal policy is to set the fault standard equal to the harm and the imprisonment term at a level that achieves compliance with the standard: $\hat{g}^* = h$ and $t^* \geq h/p\lambda$. Given this combination, individuals commit the harmful act if and only if their gain exceeds the harm (since $\tilde{g} = h$), which is the behavior desired in the first-best outcome, and the imprisonment sanction never is imposed because injurers are in compliance with the fault standard (since $\tilde{g} = \hat{g}$).

Note that the imprisonment term

[23] The point that injurers can be induced to behave optimally under fault-based liability is essentially due to John Prather Brown (1973), who showed this for privately-enforced liability rules.

[24] In Section 8 we briefly discuss the implications of errors under fault-based liability. See also note 25 below.

[25] The observation that fault-based liability does not impose risk on injurers, whereas strict liability does, is made in Shavell (1982). Clearly, this advantage of fault-based liability is lessened to the extent that mistakes in the operation of the fault system result in the imposition of fines.

needed to assure compliance with the standard can be shorter if individuals are risk averse in imprisonment, and must be longer if individuals are risk preferring in imprisonment.

Fault-based liability is again preferable to strict liability. Now this is because fault-based liability can accomplish desired deterrence without incurring the private and public costs of imprisonment.[26]

3.2.3 *Fines and Imprisonment*

If a fine and an imprisonment sanction are used together and chosen so that individuals are induced to comply with the fault standard, the optimal fault standard is the same as that discussed previously. The actual mix of the fine and the imprisonment term then is irrelevant because sanctions are not actually imposed. In particular, it is not advantageous for society to employ maximal fines before resorting to imprisonment. (Of course, if sanctions are imposed sometimes by mistake, it is better to use fines to the maximum extent possible before employing a more costly imprisonment sanction.)

4. *Optimal Enforcement with a Variable Probability of Detection*

This section considers the optimal system of enforcement when expenditures on enforcement e—and hence the probability of detection $p(e)$—are allowed to vary.[27]

4.1 *Strict Liability*

4.1.1 *Fines*

If individuals are risk neutral, we first want to establish that the optimal fine is maximal. To demonstrate this, suppose that f is less than f_m. Then f can be raised and e lowered so as to keep $p(e)f$ (that is, $\tilde{g}$) constant—so deterrence is constant. Because the behavior of individuals is unaffected but enforcement expenditures fall, social welfare rises (the first two terms in (2) do not change but e is lower). Hence, the optimal f cannot be less than f_m. In other words, because any particular level of deterrence can be achieved with different combinations of the fine and the probability of detection, society should employ the highest possible fine and a correspondingly low probability of detection in order to economize on enforcement expenditures.[28]

Second, we want to show that the optimal probability of detection is such that the expected fine is less than harm, $p(e^*)f_m < h$—that is, some degree of underdeterrence is desirable. Observe that the derivative of social welfare with respect to enforcement expenditures e is

$$-1 + (h - \tilde{g})[dZ(\tilde{g})/de], \qquad (7)$$

where $\tilde{g} = p(e)f_m$. The first term in (7) is the marginal cost of greater spending on enforcement. The second term is the deterrent effect of a higher probability of

[26] This point was originally emphasized in Shavell (1987b).

[27] Consideration of this issue originated with Becker (1968); the early writers on enforcement (including Bentham 1789) did not examine the issue of the choice of enforcement effort.

[28] The general point of this paragraph—that a low probability–high fine combination conserves enforcement costs—originates with Becker (1968) and is his principal contribution beyond what Bentham (1789) had written. However, Becker did not formally consider bounds on fines (and much of his analysis implicitly presumes that the optimal fine is not maximal). R. A. Carr-Hill and N. H. Stern (1979, pp. 300–304) and Polinsky and Shavell (1979, pp. 883–84) observed that Becker's argument implies that the optimal fine is equal to its upper bound. Many scholars have noted the unrealism of this result and have proposed additional considerations that argue for less-than-maximal fines. We will discuss several important factors of this type, notably risk aversion, general enforcement, and marginal deterrence. See also Andreoni (1991), Lucian Arye Bebchuk and Kaplow (1992, 1993), Malik (1990), and Polinsky and Shavell (1991, 1998b) for discussion of other such considerations.

detection, equal to the number of individuals who are deterred, $dZ(\tilde{g})/de$, multiplied by their net effect on social welfare, $h - \tilde{g}$. When e^* is positive, it follows from (7) that $h - \tilde{g} = h - p(e^*)f_m > 0$ (otherwise the derivative of social welfare with respect to enforcement expenditures would be negative at e^*, contradicting the optimality of e^*). In other words, $p(e^*)f_m < h$. To understand this result, suppose that p were such that $pf_m = h$. Then there would be no first-order loss of social welfare from lowering p because the individuals who would be induced to engage in the harmful activity would obtain gains equal to harm. But enforcement costs would be saved, making it desirable to lower the probability. How much p should be lowered depends on the resulting savings in enforcement expenses compared to the net social costs of underdeterrence.[29]

If individuals are risk averse, the optimal fine generally is not at its maximum. This is because the use of a very high fine would impose a substantial risk-bearing cost on individuals who commit the harmful act. More precisely, reconsider the argument that we used in the risk-neutral case. If f is less than f_m, it still is true that f can be raised and e lowered so as to keep deterrence—that is, $\tilde{g}$—constant. But because of risk aversion, the resulting probability of detection must fall more than proportionally, implying that the expected fine, and therefore fine revenue, falls. This reduction in fine revenue reflects the disutility caused by imposing greater risk on risk-averse individuals. If individuals are sufficiently risk averse, the decline in fine revenue associated with greater risk bearing could more than offset the savings in enforcement expenditures from reducing the probability of detection, implying that social welfare would be lower.[30]

In effect, when individuals are risk averse, fines become a socially costly sanction rather than a mere transfer of wealth. The more risk averse that individuals are, the better it is to control their behavior by using a lower fine and a higher probability of detection, even though this raises enforcement costs.

As in the risk-neutral case, there is a reason when individuals are risk averse to reduce enforcement costs by setting the probability such that the critical gain is less than the harm—resulting in some underdeterrence.[31]

4.1.2 *Imprisonment*

If individuals are risk neutral in imprisonment, the optimal imprisonment term is at its maximum.[32] The reasoning behind this result parallels that used to show that the optimal fine is maximal when individuals are risk neutral in fines. Specifically, if the imprisonment term is raised and the probability of detection lowered so as to keep the expected sanction constant, neither individual behavior nor the costs of imposing imprisonment are affected (by construction, the expected prison term is the same), but enforcement expenditures fall.

If individuals are risk averse in imprisonment, the argument for setting

[29] The point of this paragraph—that some underdeterrence is optimal—was first made by Polinsky and Shavell (1984).

[30] The point that the optimal fine may be less than maximal when individuals are risk averse was made initially by Polinsky and Shavell (1979). See also Kaplow (1992).

[31] It also is possible, however, that overdeterrence would be optimal. The reason is that the imposition of risk can be reduced by discouraging individuals from engaging in the harmful activity. See Kaplow (1992).

[32] It does not matter for our purposes what accounts for the bound on the imprisonment term. The bound could, for example, derive from the limited lifetimes of individuals or considerations of fairness.

the imprisonment sanction maximally is stronger than when individuals are risk neutral. This is because when the imprisonment term is raised, the probability of detection can be lowered even more than in the risk-neutral case without reducing deterrence. Thus, not only are there greater savings in enforcement expenditures, but also the public costs of imposing imprisonment sanctions now decline because the expected prison term falls.

If individuals are risk preferring in imprisonment, however, the optimal sanction may be less than maximal. In particular, the type of argument used above does not necessarily apply. Now, when the sanction is raised, the probability that maintains deterrence cannot be lowered proportionally, implying that the expected prison term rises. Because the resulting increased cost to the public of imposing imprisonment sanctions might exceed the savings in enforcement expenditures from lowering the probability, the optimal prison term might not be maximal.[33]

As in the case of fines, when the probability of detection is set optimally, together with the sanction, underdeterrence may result. An advantage of lowering the probability of detection from the level such that first-best behavior is induced is that this saves enforcement expenditures, while the decline in deterrence involves no first-order effect on social welfare in terms of gains and harm. Lowering the probability also tends to reduce the costs of imposing imprisonment sanctions because a lower fraction of injurers are caught.[34]

4.1.3 *Fines and Imprisonment*

It was demonstrated in Section 3 that, given the probability of detection, it is not optimal to use imprisonment without first using the maximum possible fine. Accordingly, our focus here is on the determination of the optimal imprisonment term when the fine is maximal and the probability of detection is chosen simultaneously.

Unlike when imprisonment is used alone, the optimal imprisonment term now may not be maximal even if individuals are risk neutral or risk averse in imprisonment. Suppose that individuals are risk neutral in imprisonment and fines. Then if the imprisonment term is raised and the probability of detection is lowered so as to keep the expected imprisonment term constant, deterrence declines because the expected fine falls (due to the reduction in the probability).[35] Hence, to maintain deterrence, the probability cannot fall proportionally. But this implies that the expected prison term, and the costs of imposing imprisonment, are higher than previously. Only if the savings in enforcement costs are sufficiently large,

[33] The preceding results in this subsection were first systematically presented by Polinsky and Shavell (1999a) (although Shavell 1991b notes the result in the case of risk neutrality). They also observed that if individuals discount the future disutility of imprisonment, individuals will act as if they are risk preferring in imprisonment, and hence the optimal imprisonment term may be less than maximal for this reason as well.

[34] However, weighed against these benefits is the added private and public cost of imprisonment due to the individuals who are induced to commit the harmful act as a result of the lowering of the probability of detection; some fraction of these individuals are detected and sanctioned. This effect could dominate the effects discussed in the text and cause the optimal expenditure on enforcement to be such that some overdeterrence is desirable.

[35] For example, suppose that the maximum fine is $10,000, that the imprisonment term initially is five years, and that the probability of detection initially is 20 percent. If the prison term is doubled to ten years and the probability is reduced by half to 10 percent, the expected prison term will remain at 1 year. The expected fine, however, will have declined from $2,000 to $1,000.

therefore, is it socially desirable to raise the imprisonment sanction.[36]

While the argument for raising the imprisonment term and lowering the probability is stronger if individuals are risk averse in imprisonment, it still might not be strong enough to make raising the prison term to its maximum desirable. Thus, when fines and imprisonment are used together and the probability is chosen optimally, the optimal imprisonment sanction may well not be maximal.

4.2 *Fault-Based Liability*

The least expensive way to accomplish compliance with the fault standard is to employ the highest possible sanction and, given this sanction, the lowest probability of detection that deters individuals who would be at fault. The reason is that if all individuals who would be at fault are deterred, the only cost incurred is associated with the setting of the probability; this cost is minimized by using the maximal sanction and a correspondingly low probability. (Note that this is true regardless of whether the sanction is a fine or imprisonment and regardless of individuals' attitudes toward the risk of fines or of imprisonment.)

Assuming that compliance with the standard is induced, the question remains what the standard should be. It is generally optimal for the standard to be lower than that associated with first-best behavior. The reason is that there is a savings in enforcement costs from reducing the standard (a lower standard does not require as high a probability of detection to induce compliance), and the first-order net social loss from more individuals committing the harmful act is zero (starting at a standard corresponding to first-best behavior).

[36] The point in this paragraph is made in Shavell (1991b).

As we previously emphasized, fault-based liability possesses the advantage over strict liability that costly sanctions are not actually imposed (in the absence of mistakes). We note here that, when the probability of detection is a policy instrument, this advantage of fault-based liability generates a secondary advantage: it may result in lower enforcement expenditures than under strict liability. Specifically, because sanctions are not imposed under fault-based liability, it becomes desirable to use high sanctions, which allows a relatively low probability of detection to be employed.[37] However, as previously mentioned, the choice between fault-based liability and strict liability is complicated by several other important factors.[38]

This concludes the presentation of the basic theory of public enforcement of law. We now turn to various extensions and refinements of the basic analysis.

5. *Accidental Harms*

In our basic analysis, we implicitly assumed that the acts that individuals commit result in harm with certainty. In many circumstances, however, acts result in harm only with a probability. For instance, if a driver speeds, he only creates a likelihood of a collision; or if a firm stores toxic chemicals in a substandard tank, the firm only creates the probability of a harmful spill.

Essentially all that we said in the

[37] For example, suppose that the sanction is a fine and that injurers are risk averse. Then, as we observed earlier, the optimal fine under strict liability generally is not maximal, due to risk bearing by injurers. This implies that the probability of detection needed to achieve any given level of deterrence is higher than if the fine were maximal. Under fault-based liability, however, the optimal fine *is* maximal despite the risk aversion of injurers, because the fine is not actually imposed. Hence, a lower probability can be employed.

[38] See the discussion at the end of subsection 3.2.1.

basic analysis applies in a straightforward way when harms are accidental. If individuals are risk neutral, sanctions are monetary, and the expected sanction equals harm whenever harm occurs, then induced behavior will be socially optimal; further, the optimal magnitude of sanctions is maximal if individuals are risk neutral because this allows enforcement costs to be saved, but is not necessarily maximal if they are risk averse, and so forth. Our general conclusions in the basic analysis can thus be interpreted to apply both when harms occur for sure, and when harms occur accidentally.

There is, however, an additional issue that arises when harm is uncertain: a sanction can be imposed either on the basis of the commission of a dangerous *act* that increases the chance of harm—storing chemicals in a substandard tank—or on the basis of the *actual occurrence of harm*—only if the tank ruptures and results in a spill. In principle, either approach can achieve optimal deterrence. To illustrate, suppose that the substandard tank has a 10 percent chance of rupturing, in which case the harm would be $10 million; the expected harm from using the tank therefore is $1 million. If injurers are risk neutral and sanctions are imposed only when harm occurs, deterrence will be optimal if, as usual, the expected sanction equals the harm of $10 million. Alternatively, if sanctions are imposed on the basis of the dangerous act of using the substandard tank, deterrence will be optimal if the owner of the tank faces an expected sanction equal to the expected harm due to his use of the substandard tank, $1 million.

Several factors are relevant to the choice between act-based and harm-based sanctions. First, act-based sanctions need not be as high to accomplish a given level of deterrence, and thus offer an underlying advantage over harm-based sanctions because of limitations in parties' assets. In the example in the preceding paragraph, the owner of the storage tank might be able to pay the $1 million required if sanctions are act-based (assuming for simplicity that injurers are always found liable) but not the $10 million required if sanctions are harm-based. Second, and closely related, because act-based sanctions need not be as high to accomplish deterrence, they offer an advantage over harm-based sanctions when parties are risk averse. Third, act-based sanctions and harm-based sanctions may differ in the ease with which they can be applied. In some circumstances, act-based sanctions may be simpler to impose (it might be less difficult to determine whether an oil shipper properly maintains its vessels' holding tanks than to detect whether one of the vessels leaked oil into the ocean); in other circumstances, harm-based sanctions may be more readily applied (a driver who causes harm might be caught without difficulty, but not one who speeds). Fourth, it may be hard to calculate the expected harm due to an act, but relatively easy to ascertain the actual harm if it eventuates; if so, this constitutes an advantage of harm-based liability.[39]

6. *Costs of Imposing Fines*

We inquire in this section about the implications of costs borne by enforcement authorities in imposing fines.[40] Our principal observation is that such costs should raise the level of the fine.

To elaborate, suppose that the probability of detection is fixed at p, that

[39] Act-based versus harm-based enforcement is discussed in Shavell (1993).

[40] We have already discussed the cost of imposing imprisonment sanctions—specifically, the cost to the public per unit of the imprisonment term, α.

liability is strict, and that individuals are risk neutral. If fines are costless to impose, the optimal fine is h/p, the harm divided by the probability of detection (see (6)). Now suppose that the enforcement authority bears a cost each time a fine is imposed; let

k = cost of imposing the fine.

It is easy to verify that the optimal fine then is

$$f^* = h/p + k; \tag{8}$$

the cost k should be added to the fine that would otherwise be desirable. The explanation is that if an individual commits a harmful act, he causes society to bear not only the immediate harm h, but also, with probability p, the cost k of imposing the fine—that is, his act results in an expected total social cost of $h + pk$. If the fine is set according to (8), the individual's expected fine is $h + pk$, which leads him to engage in the harmful act if and only if his gain exceeds the expected total social cost of the act.

There may be other costs associated with the imposition of fines. In particular, suppose that detection is followed by a costly second stage during which the state investigates and prosecutes an individual, and at the end of which a fine is imposed only with a probability. Let

s = cost of the investigation–prosecution stage; and
q = probability of a fine being imposed after the investigation–prosecution stage.

Hence, the probability that an individual will have to pay a fine is pq and the expected costs of imposing a fine, including the expected investigation–prosecution cost, become $ps + pqk$. It is readily shown that the optimal fine now is

$$f^* = h/pq + s/q + k. \tag{9}$$

This formula illustrates a general principle: the optimal fine equals the costs incurred by society as a result of the harmful act divided by the probability—at the time that each component of cost is incurred—that the injurer will have to pay the fine. Thus, h is divided by pq because, when the harm occurs, the probability of having to pay the fine is pq; and s is divided by q because, when the investigation–prosecution costs are incurred, the probability of having to pay the fine is q. If the fine is computed according to this principle, the expected fine will equal the expected social costs due to an individual committing a harmful act, including the harm caused and the expected sanctioning costs—that is, $h + ps + pqk$.

Note that under fault-based liability, it is less important to take explicit account of the costs of imposing fines. This is because, if individuals comply with the fault standard, they do not bear sanctions, in which case there are no costs associated with imposing sanctions. However, when individuals are found at fault (perhaps because of errors), the fines imposed on them also should reflect the costs of imposing fines.

Additionally, observe that, not only does the state incur costs when fines are imposed, so do individuals who pay the fines (such as legal defense expenses). The costs borne by individuals, however, do not affect the formula for the optimal fine. Individuals properly take these costs into account, because they bear them.[41]

7. *Level of Activity*

We have been assuming that the sole decision that an individual makes is whether to act in a way that causes

[41] The points developed in this section were first presented in Polinsky and Shavell (1992), although early writers on enforcement theory—including Becker (1968, p. 192) and Stigler (1970, p. 533)—recognized that sanctions should reflect enforcement costs.

harm when engaging in some activity. In many contexts, however, an individual also makes a choice about his *activity level*—that is, not only does he choose whether to commit a harmful act while engaging in an activity, he also chooses whether to engage in that activity, or, more generally, at what level to do so. For example, in addition to deciding how to behave when driving (whether to speed, whether to exercise care in changing lanes), an individual also chooses how many miles to drive; the number of miles driven is the individual's level of activity. Similarly, not only does a firm decide how to conduct its operations during production (for example, whether to pollute), it also chooses its level of production; the output of the firm is its level of activity.

The socially optimal activity level is such that the individual's marginal utility from the activity just equals the marginal expected harm caused by the activity. Thus, the optimal number of miles driven is the level at which the marginal utility of driving an extra mile just equals the marginal expected harm per mile driven. The determination of the optimal level of activity presumes that individuals act optimally when engaging in the activity—for example, that they drive with appropriate care.[42]

Will parties' choices about their activity levels be socially correct under the two major forms of liability? The answer is that under strict liability, their choices about activity levels will be correct, but under fault-based liability, they will participate in activities to a socially excessive extent. Under strict liability, parties will choose the optimal level of activity because they will pay for all harm done. They will choose the optimal number of miles to drive because they will pay for all harm per mile driven. Under fault-based liability, however, parties generally do not pay for the harm they cause because, as we have discussed, they will usually behave so as not to be found at fault. Consequently, when deciding on their level of activity, they will choose an excessive level. They will not take into account the harm that each additional mile of driving causes, and therefore they will drive too much.

The interpretation of the preceding points in relation to firms is that under strict liability, the product price will reflect the expected harm caused by production, so that the price will include the full social cost of production. Hence, the amount purchased, and thus the level of production, will tend to be socially optimal. However, under fault-based liability, the product price will not reflect harm, but only the cost of precautions; thus, the amount sold, and the level of production, will be excessive.[43]

The tendency of parties to choose an excessive level of activity under fault-based liability, but not under strict liability, constitutes a fundamental advantage of strict liability. This advantage, note, is stronger the greater is the harm engendered by engaging in the activity (given that behavior is optimal when engaging

[42] We can sketch how the basic model we discussed in Sections 2–4 can be modified to take into account the activity level. Let $u(x)$ be the utility a person obtains from engaging in an activity at level x (such as driving to x different destinations). Assume also that when a person engages in the activity, he can commit a harmful act (exceed the speed limit by 10 mph) which generates a gain to him of g (time saved in reaching each destination) and increases harm (per destination) from h_1 to h_2. If it is optimal for the person not to commit the harmful act when engaging in the activity (that is, if $g < h_2 - h_1$), the optimal x is determined by $u'(x) = h_1$. If, however, it is optimal for the person to commit the harmful act when engaging in the activity (because $g > h_2 - h_1$), then the optimal x is determined by $u'(x) + g = h_2$.

[43] Our discussion here about activity-level considerations in the context of public enforcement closely parallels the analysis of activity-level issues in the context of tort liability. See generally Shavell (1980) and Polinsky (1980b).

in the activity). Thus, for activities for which expected harm is likely to be substantial, the disadvantage of fault-based liability will be significant.

We conclude with two observations about the interpretation of the point of this section. First, safety regulations and other regulatory requirements are often framed as standards of care that have to be met, but which, if met, free the regulated party from penalties. Hence, regulations of this character are subject to the criticism that they lead to excessive levels of the regulated activity. Making parties strictly liable for harm would be superior to safety regulation with respect to inducing socially correct activity levels.

Second, the advantage of strict liability over fault-based liability applies to any dimension of behavior which affects expected harm but which is not included in the definition of fault. For example, suppose that pollution damage depends both on whether a scrubber is installed as well as on the degree of care with which it is cleaned. Because the existence of a scrubber is easy to verify, but whether it is properly cleaned may not be, a fault-based system of liability might, of necessity, reflect only the first dimension of behavior. Consequently, the injurer will have a socially inadequate incentive to clean the scrubber under fault-based liability. This problem does not arise under strict liability because the injurer has to pay for harm regardless of its cause.[44]

8. *Errors*

Errors of the two classic types can occur in public enforcement of law. First, an individual who should be found liable might mistakenly not be found liable—a Type I error. Second, an individual who should not be found liable might mistakenly be found liable—a Type II error. For an individual who has been detected, let:

ε_1 = the probability that an individual who should be liable is mistakenly found not liable (a Type I error); and

ε_2 = the probability that an individual who should not be liable is mistakenly found liable (a Type II error).

For example, suppose police randomly monitor drivers by stopping them and administering a blood alcohol test. The test might understate the amount of alcohol in the driver's blood and result in a Type I error, or might overstate the amount and lead to a Type II error.

We initially consider the effect of mistake assuming that liability is strict, the sanction is a fine, and individuals are risk neutral. Given the probability of detection p and the chances of Type I and Type II errors, an individual will commit the wrongful act if and only if his gain net of his expected fine if he does commit it exceeds what he bears if he does not commit it:

$$g - p(1 - \varepsilon_1)f > -p\varepsilon_2 f, \qquad (10)$$

or, equivalently, if and only if[45]

$$g > (1 - \varepsilon_1 - \varepsilon_2)pf. \qquad (11)$$

Note initially that *both* types of error reduce deterrence: the right-hand side of (11) is declining in both ε_1 and ε_2. The first type of error diminishes deterrence because it lowers the expected fine if an individual violates the law. The second type of error, mistaken liability, also

[44] The advantage of strict liability over fault-based liability with respect to activity-level considerations is an example of the point of this paragraph. An injurer's activity level is a dimension of his behavior that affects expected harm but is not included in the definition of fault.

[45] We assume that $1 - \varepsilon_1 - \varepsilon_2 > 0$, so that the probability that a guilty person will be found liable, $1 - \varepsilon_1$, exceeds the probability that an innocent person will be found liable, ε_2.

lowers deterrence because it reduces the difference between the expected fine from violating the law and not violating it. In other words, the greater is ε_2, the smaller the increase in the expected fine if one violates the law, making a violation less costly to the individual.[46]

Because mistakes dilute deterrence, they tend to reduce social welfare. Specifically, to achieve any level of deterrence, it may be necessary to raise the probability of detection or a costly sanction to offset the effect of errors. It should also be noted that, were one to take into account an individual's decision whether to engage in an activity (like driving), Type II errors have the additional effect of discouraging socially desirable participation in the activity.

Now consider the optimal choice of the fine. If the probability of detection is fixed, the dilution in deterrence caused by errors requires a higher fine to restore deterrence, so the optimal fine is higher.[47] If both the probability and the fine are policy instruments, the optimal fine remains maximal despite mistakes. The explanation is essentially that used previously: If the fine f were less than maximal, then f could be raised and the probability p lowered so as to keep deterrence constant, but saving enforcement costs.

If individuals are risk averse, however, the possibility of mistakes does affect the optimal fine. As we emphasized in Section 4, the optimal fine generally is less than maximal when individuals are risk averse—lowering the fine reduces the bearing of risk. Introducing the possibility of mistakes may increase the desirability of lowering the fine because, due to Type II errors, individuals who do not violate the law are subject to the risk of having to pay a fine. Indeed, because the number of persons who do not violate the law often would far exceed the number who do, the desire to avoid imposing risk on the former group can lead to a substantial reduction in the optimal fine.[48]

The possibility of mistakes generally affects the optimal probability of detection. On one hand, the deterrence-diluting effects of mistakes means, as we noted, that a higher probability of detection may be needed to achieve any given level of deterrence, tending to raise the optimal expenditure on enforcement. On the other hand, mistakes effectively reduce the productivity of enforcement expenditures (by a factor of $1 - \varepsilon_1 - \varepsilon_2$), thereby making enforcement more costly and tending to reduce the optimal expenditure on enforcement. Either of these effects could dominate and lead to an optimal probability of detection that is higher or lower than in the absence of mistakes.[49]

Next, consider imprisonment and mistake. As in the case of fines, mistakes of both type dilute the deterrent effects of imprisonment. Additionally, the optimal imprisonment term is maximal if individuals are risk neutral or risk averse in imprisonment, but is generally not maximal if they are risk preferring in imprisonment.[50]

[46] This point was first emphasized by Png (1986).

[47] Specifically, to achieve first-best behavior, it must be that $(1 - \varepsilon_1 - \varepsilon_2)pf = h$, which implies that f must be higher the greater are either of the errors, ε_1 or ε_2.

[48] Building on Polinsky and Shavell (1979), Michael K. Block and Joseph Gregory Sidak (1980, pp. 1135–39) emphasize the desirability of lowering sanctions when there are mistakes and injurers are risk averse.

[49] Regardless of how the optimal probability is affected, it can be demonstrated that mistakes reduce the optimal level of deterrence (that is, lower $\tilde{g}$). The explanation is that, because the optimal fine is maximal, the only way to alter deterrence is by changing the probability of detection. And because mistakes reduce the productivity of enforcement expenditures, it is not worth accomplishing as much deterrence.

[50] That the optimal term remains maximal if individuals are risk neutral or risk averse might seem

We have not yet commented on fault-based liability. Here, an important implication of mistake is that some individuals will bear sanctions even if they comply with the fault standard. Hence, the results regarding the optimal probability of detection and the sanction under strict liability apply to some extent under fault-based liability as well. Moreover, individuals will often have a motive to take excessive precautions in order to reduce the chance of erroneously being found at fault.[51]

Finally, observe that, although we have treated the probabilities of error as fixed, they can be influenced by policy choices. For example, prosecutorial resources can be increased in order to reduce the probability of a Type I error, or the standard of proof can be raised to reduce the chance of a Type II error (although this presumably increases Type I errors). Because the reduction of both types of error increases deterrence, expenditures made to reduce errors may be socially beneficial.[52]

9. *General Enforcement*

In many settings, enforcement may be said to be *general* in the sense that several different types of violations may be detected by an enforcement agent's activity. For example, a police officer waiting at the roadside may notice a driver who litters as well as a driver who goes through a red light or speeds; or a tax auditor may detect a variety of infractions when he examines a tax return. To investigate such situations, suppose that a single probability of detection applies to all harmful acts, regardless of the magnitude of the harm.[53] (The contrasting assumption is that enforcement is *specific*, meaning that the probability is chosen independently for each type of harmful act.)

The main point that we want to make is that in contexts in which enforcement is general, the optimal sanction rises with the severity of the harm and is maximal only for relatively high harms. To see this, assume that liability is strict, the sanction is a fine, and injurers are risk neutral. Let $f(h)$ be the fine given harm h. Then, for any general probability of detection p, the optimal fine schedule is

$$f^*(h) = h/p, \tag{12}$$

provided that h/p does not exceed the maximal fine f_m; if h/p is not feasible, the optimal fine is maximal. This schedule is obviously optimal given p because it implies that the expected fine equals harm, thereby inducing first-best behavior, whenever that is possible.

The question remains whether it would be desirable to lower p and raise fines to the maximal level for the low-harm acts for which $f^*(h)$ is less than maximal. The answer is that if p is reduced for the relatively low-harm acts (and the fine raised for them), then p—being general—is *also* reduced for the high-harm acts for which the fine is already maximal, resulting in lower deterrence of these acts. The decline in deterrence of high-harm acts may cause a greater social loss than the savings in

surprising because one might expect that the presence of Type II errors would result in a lower optimal term. But the usual argument still applies: If the term were not maximal, it could be raised and the probability of detection could be lowered at least proportionally without sacrificing deterrence. Hence, the aggregate amount of jail time served by individuals who do not commit the harmful act would remain the same or fall, and enforcement expenditures would fall for sure.

[51] This point was first emphasized by Richard Craswell and John E. Calfee (1986).

[52] On the value of accuracy in adjudication, see Kaplow and Shavell (1994a).

[53] It will be clear that the main point developed in this section does not depend on the assumption that the same probability applies to all acts. The only requirement is that the probabilities for different acts are linked, each a function of the same enforcement expenditure.

enforcement costs from lowering p. To express this point differently, p must be sufficiently high to avoid significant underdeterrence of high-harm acts (for which fines are maximal). But since this p also applies to less harmful acts, the fines for them do not need to be maximal in order to deter them appropriately.[54]

The result that, when enforcement is general, sanctions should rise with the severity of harm up to a maximum also holds if the sanction is imprisonment and if liability is fault-based. The underlying reasoning is the same as that given above.[55]

10. *Marginal Deterrence*

In many circumstances, an individual may consider which of *several* harmful acts to commit, for example, whether to release only a small amount of a pollutant into a river or a large amount, or whether only to kidnap a person or also to kill him. In such contexts, the threat of sanctions plays a role in addition to the usual one of deterring individuals from committing harmful acts: for individuals who are not deterred, expected sanctions influence which harmful acts individuals choose to commit. Notably, such individuals will have a reason to commit less harmful rather than more harmful acts if expected sanctions rise with harm. Deterrence of a more harmful act because its expected sanction exceeds that for a less harmful act is sometimes referred to as *marginal deterrence*.[56]

Other things being equal, it is socially desirable that enforcement policy creates marginal deterrence, so that those who are not deterred from committing harmful acts have a reason to moderate the amount of harm that they cause. This suggests that sanctions should rise with the magnitude of harm and, therefore, that most sanctions should be less than maximal. However, fostering marginal deterrence may conflict with achieving deterrence generally: for the schedule of sanctions to rise steeply enough to accomplish marginal deterrence, sanctions for less harmful acts may have to be so low that individuals are not deterred from committing some harmful act.[57]

Two additional observations should be made about marginal deterrence. First, marginal deterrence can be promoted by increasing the probability of detection as well as the magnitude of sanctions. For example, kidnappers can be deterred more from killing their victims if greater police resources are devoted to apprehending kidnappers who murder their victims than to those who do not. (Note, though, that in circumstances in which enforcement is general, the probability of detection cannot be independently altered for acts that cause different degrees of harm.) Second, marginal deterrence is naturally accomplished if the expected sanction equals harm for all levels of harm; for if a person is paying for harm done, he will have to pay appropriately more if he does greater harm. Thus, for instance, if a polluter's expected fine would rise from $100 to $500 if he dumps five gallons instead of one gallon of waste

[54] Note that if p could be varied *independently* for a low-harm act and for a high-harm act—that is, if enforcement is specific rather than general—then it would be desirable to lower p and raise the fine for a low-harm act if the fine for it were less than maximal.

[55] The basic point of this section was first made in Shavell (1991b); see also Mookherjee and Png (1992) for a closely related analysis.

[56] The notion of marginal deterrence was remarked upon in some of the earliest writing on enforcement; see Beccaria (1767, p. 32) and Bentham (1789, p. 171). The term "marginal deterrence" apparently was first used by Stigler (1970).

[57] For formal treatments of marginal deterrence, see Shavell (1992), Louis Wilde (1992), and Mookherjee and Png (1994).

into a lake, where each gallon causes $100 of harm, his marginal incentives to pollute will be correct.[58]

11. *Principal-Agent Relationship*

Although we have assumed that an injurer is a single actor, in fact the injurer is often a principal and the principal's agent. For example, the principal could be a firm and the agent an employee; or the principal could be a contractor and the agent a subcontractor.

When harm is caused by the behavior of principals and agents, many of our conclusions carry over to the sanctioning of principals. For example, given the probability of detection p, it is optimal for a risk-neutral principal to face a fine of h/p. Then the expected fine is equal to harm done, so the principal will in effect be in the same position vis-a-vis his agent as society is vis-a-vis a single potential violator of law. Consequently, the principal will behave socially optimally in controlling his agents, and in particular will contract with them and monitor them in ways that will give the agents socially appropriate incentives to reduce harm.[59]

A question about enforcement that arises when there are principals and agents is how to allocate financial sanctions between them. First observe that the particular allocation of sanctions may not matter when, as would be the natural presumption, the principal and the agent can reallocate sanctions through their own contract. For example, if the agent finds that he faces a large fine but is more risk averse than the principal, the principal can assume it; conversely, if the fine would be imposed on the principal, he can bear that risk and not impose an internal sanction on the agent. Thus, the post-contract penalties that the agent suffers may not be affected by the particular division of sanctions initially selected by the enforcement authority.

The allocation of monetary sanctions between principals and agents would matter, however, if some allocations allow the pair to reduce their total burden. An important example is when a fine is imposed only on the agent and he is unable to pay it because his assets are less than the fine.[60] Then he and the principal (who often would have higher assets) would jointly escape part of the fine, diluting deterrence. The fine therefore should be imposed on the principal rather than on the agent (or at least the part of the fine that the agent cannot pay).

A closely related point is that the imposition of imprisonment sanctions on agents may be desirable when their assets are less than the optimal fine, even if the principal's assets are sufficient to pay the fine. The fact that an agent's assets are limited means that the principal may be unable to control him adequately through use of contractually-determined penalties, which can only be monetary. For example, a firm may not

[58] As we emphasized in Section 4, however, it often is desirable for society to tolerate some underdeterrence in order to save enforcement costs, in which case expected sanctions will be less than harm. Then, consideration of marginal deterrence alters the structure of sanctions that would otherwise be best.

[59] There is relatively little literature on the question of optimal enforcement when injurers are principals and agents. Harry A. Newman and David W. Wright (1990) study the optimal monetary sanction to impose on a risk-neutral principal when liability is strict and is imposed for sure; they show that it equals harm. Jennifer A. Arlen (1994) examines the effect of sanctions on corporations' incentives to monitor their employees, and she emphasizes the possibility that corporations may have perverse incentives not to monitor if they would become liable as a result of their discovering and reporting employee violations. Also, Shavell (1997b) finds that optimal sanctions on corporations could be above or below harm when employee assets are less than harm.

[60] See Alan O. Sykes (1981) and Lewis A. Kornhauser (1982).

be able, despite the threat of salary reduction or dismissal, to induce its employees never to rig bids. In such circumstances, it may be socially valuable to use the threat of personal criminal liability and a jail sentence to better control agents' misconduct.[61]

12. *Settlements*

We have thus far assumed that when an injurer who should be found liable is discovered, he will be sanctioned in some automatic fashion. In practice, however, an injurer must be found liable in a trial, and before this occurs, it is common for an injurer to settle in lieu of trial. (In the criminal context, the settlement usually takes the form of a *plea bargain*, an agreement in which the injurer pleads guilty to a reduced charge.) Given the prevalence of settlements, it is important to consider how they affect deterrence and the optimal system of public enforcement, and whether settlements are socially desirable.

There are two general reasons why parties might prefer an out-of-court settlement to a trial. First, a trial is costly to the parties in terms of time and/or money. Second, settlements eliminate the risks inherent in the trial outcome, a benefit to parties who are averse to such risks.[62] These advantages of settlement to the parties suggest that settlement is socially valuable, but the effect of settlement on deterrence is a complicating factor.

Specifically, settlements dilute deterrence: for if injurers desire to settle, it must be because the expected disutility of sanctions is lowered for them. However, because settlements reflect the sanctions that would be imposed at trial, the state may be able to offset this settlement-related reduction in deterrence by increasing the level of sanctions. If so, settlements need not compromise the overall level of deterrence.[63]

Settlements may have other socially undesirable consequences. First, they may result in sanctions that are not as well tailored to harmful acts as would be true of court-determined sanctions. For example, if injurers have private information about the harm that they have caused, they may be able to obtain settlements that are less than is appropriate, whereas trial outcomes may better approximate the correct sanctions. Thus, the distribution of sanctions effected through settlements may not be as good in properly deterring injurers. Second, settlements hinder the amplification and development of the law through the setting of precedents, a factor of occasional relevance. Third, settlements also sometimes allow defendants to keep aspects of their behavior secret, which can reduce deterrence. Fourth, settlements for prison terms can result in increases in public expenditures on jail if defendants are risk averse in imprisonment.[64]

[61] This point is discussed by Kathleen Segerson and Tom Tietenberg (1992) and emphasized by Polinsky and Shavell (1993).

[62] These benefits of settlement are well-recognized in the economic literature on civil litigation; see the survey by Robert D. Cooter and Daniel L. Rubinfeld (1989). For early discussions of settlement in the context of public enforcement, see William M. Landes (1971) and Gene M. Grossman and Michael L. Katz (1983), and more recently, see, for example, Jennifer F. Reinganum (1988), Bruce H. Kobayashi and John R. Lott, Jr. (1992), and Thomas J. Miceli (1996).

[63] The deterrence-diluting effects of settlement and other aspects of the social desirability of settlement have been discussed in the private litigation context by Polinsky and Rubinfeld (1988), Shavell (1997a), and Kathryn E. Spier (1997). A related discussion in the public enforcement context appears in Polinsky and Rubinfeld (1989).

[64] For example, suppose a defendant faces a 50 percent chance of a five-year sentence and a 50 percent chance of a fifteen-year sentence, with an expected sentence of ten years. If he is risk averse, he will strictly prefer a certain sentence of ten years. This implies that if prosecutors want to

13. *Self-Reporting*

We have assumed above that individuals are subject to sanctions only if they are detected by an enforcement agent, but in fact parties sometimes disclose their own violations to enforcement authorities. For example, firms often report violations of environmental and safety regulations, individuals usually notify police of their involvement in traffic accidents, and even criminals occasionally turn themselves in.

We explain here why it is generally socially desirable for the structure of enforcement to be such as to encourage self-reporting. Self-reporting can be induced by lowering the sanction for individuals who disclose their own infractions. Moreover, the reward for self-reporting can be made small enough that deterrence is only negligibly reduced.

To amplify, assume for simplicity that individuals are risk neutral, and suppose that if an individual commits a violation and does not self-report, his expected fine is pf. Set the fine if a violator does self-report just below pf. Specifically, let

$$f' = \text{fine if a violator self-reports,}$$

and set

$$f' = pf - \varepsilon, \tag{13}$$

where $\varepsilon > 0$ is arbitrarily small. A violator will therefore want to self-report because f' is less than pf, but the deterrent effect of the sanction will be (approximately) the same as if he did not self-report. For example, suppose that the fine if an individual does not self report is \$1,000 and that the probability of detection is 10 percent, so the expected fine is \$100. If the fine for self-reporting is slightly below \$100, individuals will self-report but deterrence will barely be reduced.

Given that self-reporting can be induced essentially without compromising deterrence, why is self-reporting socially advantageous? There are several reasons. First, self-reporting lowers enforcement costs because, when it occurs, the enforcement authority does not have to identify and prove who the violator was. Environmental enforcers do not need to spend as much effort trying to detect pollution and establishing its source if firms that pollute report that fact.[65] Second, self-reporting reduces risk, and thus is advantageous if injurers are risk averse. Drivers bear less risk because they know that if they cause an accident, they will be led to report this to the police and suffer a lower and certain sanction, rather than face a substantially higher sanction (for hit and run driving) imposed only with some probability. Third, self-reporting may allow harm to be mitigated. Early identification of a toxic leak will facilitate its containment and clean-up.[66]

14. *Repeat Offenders*

In practice, the law often sanctions repeat offenders more severely than first-time offenders. For example, under the U.S. Sentencing Commission's sentencing guidelines for Federal

maintain deterrence, they must demand a settlement of more than ten years, say twelve years, which increases the cost of imprisonment.

[65] In some contexts, however, self-reporting will not save enforcement costs. For example, suppose that a police officer waits by the roadside to spot speeders. Then, were a driver to report that he had sped, this would not reduce policing costs, presuming that the officer still needs to be stationed at the roadside to watch for other speeders. Usually, though, there would be some cost savings as a result of self-reporting (for example, the police officer would not have to chase as many speeders).

[66] The basic theory of self-reporting in public enforcement is developed in Kaplow and Shavell (1994b); see also Malik (1993). Related literature concerns the reporting of income by individuals to tax authorities and the reporting of costs by regulated firms to regulatory authorities. See, for example, Andreoni, Brian Erard, and Jonathan Feinstein (1998) and Jean-Jacques Laffont and Jean Tirole (1993).

crimes, both imprisonment terms and criminal fines are enhanced if a defendant has a prior record of offenses. Civil money penalties also sometimes depend on whether the defendant has a prior record. We explain here why such policies may be socially desirable.

Note first that sanctioning repeat offenders more severely cannot be socially advantageous if deterrence always induces first-best behavior. If the sanction for polluting and causing a $1,000 harm is $1,000, then any person who pollutes and pays $1,000 is a person whose gain from polluting (say the savings from not installing pollution control equipment) must have exceeded $1,000. Social welfare therefore is higher as a result of his polluting. If such an individual polluted and was sanctioned in the past, that only means that it was socially desirable for him to have polluted previously. Raising the sanction because of his having a record of sanctions would overdeter him now.

Accordingly, only if deterrence is inadequate is it possibly desirable to condition sanctions on offense history to increase deterrence. But deterrence often will be inadequate because, as we emphasized in Section 4, it will usually be worthwhile for the state to tolerate some underdeterrence in order to reduce enforcement expenses.

Given that there is underdeterrence, making sanctions depend on offense history may be beneficial for two reasons. First, the use of offense history may create an additional incentive not to violate the law: if detection of a violation implies not only an immediate sanction, but also a higher sanction for a future violation, an individual will be deterred more from committing a violation presently.[67] Second, making sanctions depend on offense history allows society to take advantage of information about the dangerousness of individuals and the need to deter them: individuals with offense histories may be more likely than average to commit future violations, which might make it desirable for purposes of deterrence to impose higher sanctions on them.[68]

There is also an incapacitation-based reason for making sanctions depend on offense history. Repeat offenders are more likely to have higher propensities to commit violations in the future and thus more likely to be worth incapacitating by imprisonment.[69]

15. *Imperfect Knowledge about the Probability and Magnitude of Sanctions*

Although we have made the simplifying assumption that injurers know the probability and magnitude of sanctions, it is obvious that individuals frequently have only imperfect knowledge of these variables. They generally possess only estimates, or more likely, subjective probability distributions, of the probability of a sanction and its magnitude.

---

[67] There is a subtlety in demonstrating the optimality of punishing repeat offenses more severely. Namely, if there is a problem of underdeterrence, one might wonder why it would not be optimal to raise the sanction to the maximum level for every offense (meaning that repeat offenses could not be punished more severely). It must be shown that punishing all offenses maximally is inferior to punishing first offenses less than maximally and punishing repeat offenses more severely. See Polinsky and Shavell (1998a) on the possible optimality of making sanctions depend on offense history because of the additional deterrence that such a policy creates.

[68] Note that this reason for making sanctions depend on offense history is different from the first reason: the second reason involves the assumption that offenders are different and that the optimal sanction for some offenders is higher than for others; the first reason applies even if individuals are identical. On the second, information-based reason for making sanctions depend on offense history, see C. Y. Cyrus Chu, Sheng-cheng Hu, and Ting-yuan Huang (1997), Polinsky and Rubinfeld (1991), and Ariel Rubinstein (1979).

[69] We elaborate on this point in our discussion in Section 16 of the incapacitation rationale for use of imprisonment sanctions.

They might not know the true probability of a sanction for several reasons: because the enforcement authority refrains from publishing information about the probability (perhaps hoping that individuals will believe it to be higher than it is); because the probability depends on factors that individuals do not fully understand (the probability of a tax audit, for example, is influenced by a large number of considerations); and because probabilities seem to be difficult for individuals to assess.[70] Also, individuals may have incomplete knowledge of the true magnitudes of sanctions, particularly if sanctions are not fixed by law, but are to some degree discretionary.[71]

The implications of injurers' imperfect knowledge are straightforward. First, to predict how individuals behave, what is relevant, of course, is not the actual probability and magnitude of a sanction, but the perceived levels or distributions of these variables.

Second, to determine the optimal probability and magnitude of sanctions, account must be taken of the relationship between the actual and the perceived values. For example, suppose that there is a delay of at least a year before individuals fully comprehend a change in the probability of enforcement. Then if enforcement resources are increased so as to make the probability, say, 15 percent rather than 10 percent, there might not be a significant increase in deterrence for some time, making such an investment less worthwhile.[72] Or, for instance, suppose that the sanction for some act, such as robbery, can vary (say from one month of jail time to ten years), and that individuals' perceptions are quite rough, not based on true averages, but rather on the possible range. Then increasing the average sentence might have very little effect on deterrence, in contrast to increasing the probability of apprehension. The processes through which individuals formulate probabilities of sanctions and their magnitudes are important, therefore, to determining how deterrence functions and to optimal policy.[73]

16. *Incapacitation*

Our discussion of public enforcement has presumed that the threat of sanctions reduces harm by discouraging individuals from causing harm—that is, by deterring them. However, a different way for society to reduce harm is by imposing sanctions that remove parties from positions in which they are able to cause harm—that is, by *incapacitating* them. Imprisonment is the primary incapacitative sanction, although there are other examples: individuals can lose their drivers licenses, preventing them from doing harm while driving; businesses can lose their right to operate in certain domains, and the like. We focus

[70] On difficulties that individuals have in evaluating and using probabilities, see Daniel Kahneman, Paul Slovic, and Amos Tversky (1982).

[71] In addition, individuals could have imperfect information about the prevailing standard of liability, not being sure whether it is strict or fault-based. This type of mistake, about a discrete issue, seems less likely to be significant than errors in assessing the probability and magnitude of sanctions.

[72] Similarly, suppose that individuals treat all probabilities of enforcement that are low, say below 1 percent, as if they were probabilities of 1 percent, because it is not possible for individuals to make discriminations finer than 1 percent. Then if the actual probability is 1/2 percent, spending more on enforcement to make the probability 1 percent would not be beneficial because deterrence would not increase.

[73] Bebchuk and Kaplow (1992) consider imperfect information about the probability of sanctions and emphasize that maximal sanctions may not be socially desirable. See also Kaplow (1990a), which takes into account learning about whether acts are subject to sanctions, and Raaj K. Sah (1991), which focuses on the process by which individuals form perceptions of the probability of detection.

here on imprisonment, but what we say applies to incapacitative sanctions generally.

To better understand public enforcement when sanctions are incapacitative, suppose that the sole function of sanctions is to incapacitate; that is, assume for simplicity that sanctions do not deter. (For instance, deterrence might not occur if, given the relevant range of the probabilities and magnitudes of the sanctions, individuals' gains from harmful acts exceed the expected sanctions.) We assume that the social goal is as before, to maximize gains from acts less harm, and less the costs of enforcement and sanctions, including the costs of keeping individuals in prison.

The optimal sanction to impose on an individual who is apprehended is determined by comparing the expected harm, net of gains, he would cause if not in prison to the private and public costs of imprisonment. If the expected net harm exceeds the costs of imprisonment, he should be put in prison and kept there as long as this condition holds. Thus, the optimal sanction as a function of expected net harm is zero up to a threshold—the point at which expected net harm equals the costs of imprisonment—and then rises discontinuously to the length of time during which the person's net expected harm exceeds imprisonment costs. Jail should only be used to incapacitate individuals whose net harm is relatively high.

Two points about the incapacitative rationale are important to note. First, there is evidence suggesting that the expected harm caused by individuals declines with their age.[74] Thus, from the incapacitative standpoint, it often will be desirable to release older prisoners from jail. Second, as a matter of logic, the incapacitative rationale might imply that a person should be put in jail even if he has not committed a crime—if his danger to society makes incapacitating him worthwhile. This would be true, for example, if there were some accurate way to predict a person's dangerousness independently of his actual behavior. In practice, however, the fact that a person has committed a harmful act may be the best basis for predicting his future behavior, in which case the incapacitation rationale would imply that a jail term should be imposed only if the individual has committed an especially harmful act.

The optimal probability of detection is determined by a straightforward tradeoff. The higher the probability, the greater the number of individuals who will be incapacitated, resulting in social gains equal to the difference between the individuals' expected net harm and the cost of their incapacitation. But the higher the probability, the higher are enforcement costs. At some point, it is optimal to stop raising the probability, when the marginal social gains just equal the marginal cost of raising the probability.

Last, we briefly comment on the relationship between the nature of optimal enforcement when incapacitation is the goal versus when deterrence is the goal. First, when incapacitation is the goal, the optimal magnitude of the sanction is independent of the probability of apprehension. In contrast, when deterrence is the goal, the optimal sanction depends on the probability—the sanction generally is higher the lower the probability. Second, when incapacitation is the goal, the probability and magnitude of sanctions are independent of the ability to deter. Thus, for example, if this ability is limited (as, for instance, with the enraged), a low expected sanction may be optimal under

[74] See, for example, Wilson and Herrnstein (1985, pp. 126–47) and U.S. Department of Justice (1997a, p. 371, Table 4.4; pp. 378–79, Table 4.7).

the deterrence rationale, but a high expected sanction still might be called for to incapacitate.[75]

17. *Summary*

In this section we summarize the main points of our article:

(a) When the probability of detection of a harmful act is taken as fixed, the optimal fine is the harm divided by the probability of detection, for this results in an expected fine equal to the harm. However, costs incurred by the public in collecting the fine should be added to this benchmark level of the fine, and risk aversion of injurers should usually lower the level of the fine.

(b) When the probability of detection can be varied for a harmful act, high sanctions may be optimal, for this allows a relatively low probability to be employed and thereby saves enforcement costs. Indeed, the optimal fine is maximal for this reason if individuals are risk neutral in wealth, and the optimal imprisonment term is maximal if individuals are risk neutral or risk averse in imprisonment. More realistically, optimal sanctions generally are not maximal when individuals are risk averse in wealth or risk preferring in imprisonment (as well as for other reasons[76]), although the motive to set sanctions at relatively high levels in order to reduce enforcement costs still applies.

(c) Optimal enforcement tends to be characterized by some degree of underdeterrence relative to first-best behavior, because allowing some underdeterrence conserves enforcement resources. More precisely, by lowering the probability of detection from a level that would lead to first-best behavior, the state reduces enforcement costs, and although more individuals commit the harmful act, these individuals do not cause social welfare to decline substantially because their gains are approximately equal to the harm.

(d) The use of fines should be exhausted before resort is made to the costlier sanction of imprisonment.

(e) An advantage of fault-based liability over strict liability is that sanctions that are costly to impose—imprisonment, and fines when individuals are risk averse—are imposed less often. Under fault-based liability, injurers generally are induced (in the absence of mistakes) to obey fault standards, and therefore ordinarily do not bear sanctions. Under strict liability, however, injurers are sanctioned whenever they are caught.

(f) An advantage of strict liability over fault-based liability is that the former is easier to apply. Another advantage is that injurers' activity-level decisions generally will be better. Under strict liability, injurers' activity levels will tend to be optimal, because injurers will pay for the harm that they cause. But under fault-based liability, their activity levels will tend to be excessive because they generally will not pay for the harm that they cause (due to their being led to behave without fault).

18. *Theory versus Practice*

Having reviewed the economic theory of public enforcement of law, we briefly comment on the relationship between optimal enforcement and enforcement in practice.

[75] See Shavell (1987a) for a theoretical examination of optimal incapacitation policy; Isaac Ehrlich (1981, pp. 315–16, 319–21) for a model used to estimate the relative importance of incapacitation and deterrence; and Steven D. Levitt (1998) for an empirical study of incapacitation and deterrence. Economists have paid much less attention to incapacitation than to deterrence, despite the significance of the incapacitation rationale in criminal law enforcement.

[76] For example, as seen in Section 9, if enforcement is general rather than specific, optimal sanctions are not maximal when harm is relatively low.

First observe that important features of actual public enforcement are congruent, at least in a broad sense, with what is theoretically desirable. Obviously, and significantly, public enforcement is often characterized by low probabilities of detection. This is true for many criminal acts, and also is frequently the case in other spheres of public enforcement, such as traffic control and tax collection.[77] That probabilities of detection are low undoubtedly reflects the cost of raising the probability, a central factor in our discussion.

Corresponding to the low probabilities of detection are relatively high sanctions, often exceeding harm. For example, it seems that the sentence for theft typically outweighs the harm from that act, that the penalty for double parking frequently surpasses the resulting congestion costs, and that the sanction for tax evasion tends to exceed the social losses thereby created. Sanctions that are substantially in excess of harm are needed for proper deterrence when the probabilities of enforcement are significantly less than one, as they are in these examples.

Additionally, the magnitudes of sanctions tend to increase with the severity of harms. This is so in criminal law, where, for example, punishment for theft is less than that for rape or for murder. In safety and health regulation, sanctions generally rise with the actual or the expected level of harm, and similarly in other areas of enforcement. For various reasons, including general enforcement and marginal deterrence, this basic relationship between sanctions and harm makes sense.

It also seems that the theory of optimal enforcement helps to explain why society uses the sanction of imprisonment when it does—for the category of harmful acts labeled criminal, notably, for theft, robbery, rape, murder, and so forth.[78] Because such acts often are detected with a low probability, frequently yield significant benefits to those who commit them, and also cause substantial harm, the magnitudes of penalties that would be desirable are high. If these penalties were solely monetary, they often would exceed the assets of individuals who might commit the acts. This point is strongly reinforced by the observation that individuals who commit crimes tend to have very low assets.[79] Imprisonment sanctions, therefore, usually will be required to maintain a tolerable level of deterrence of acts classified as criminal. The use of imprisonment sanctions also makes sense in view of their incapacitative function: crimes cause substantial harm and are difficult to deter (for the reason we just emphasized, as well as others), so that it often will be desirable to incapacitate individuals who have committed them.

Note, too, that the standard of liability when imprisonment sanctions are imposed is fault-based—imprisonment is premised on an evaluation of the character of wrongdoing, not merely because harm is done. This is socially desirable because, as we stressed, fault-based

[77] U.S. Department of Justice (1997b, p. 205, Table 25) indicates, for example, that the likelihood of arrest was 13.8 percent for burglary, 14.0 percent for automobile theft, and 16.5 percent for arson. Donald S. Kenkel (1993, p. 145) estimates that the probability of arrest for drunk driving is "only about .003." And according to Andreoni, Erard, and Feinstein (1998, p. 820), the audit rate for individual tax returns was 1.7 percent in 1995.

[78] See generally Posner (1985, pp. 1201–05) and Shavell (1985, pp. 1236–41).

[79] For example, in U.S. Department of Justice (1988, p. 35) it is reported that "the average inmate was at the poverty level before entering jail" and in U.S. Department of Justice (1998, p. 4) it is stated that almost half of jail inmates reported incomes of less than $600 a month in the month before their most recent arrest.

liability reduces the use of socially costly sanctions.[80]

Although actual public enforcement is consistent in many respects with the theory of optimal enforcement, actual enforcement also appears to deviate in various ways from what is theoretically desirable. We note two discrepancies of general importance. First, substantial enforcement costs could be saved without sacrificing deterrence by reducing enforcement effort and simultaneously raising fines. This is possible in many enforcement contexts because fines are presently very low relative to the assets of violators. For example, the fines for most parking violations are less than $50, penalties for underpayment of income taxes are typically on the order of 20 percent of the amount not paid, and fines for corporate violations of health and safety regulations are frequently minuscule in relation to corporate assets. In such areas of enforcement, therefore, fines could readily be, say, doubled and enforcement costs reduced significantly, while maintaining deterrence at present levels.

Not only can present levels of deterrence be achieved more cheaply, it also appears that these levels are often too low. This is a reasonable supposition given the limited use of fines that we just noted and the low probabilities of their application. For example, the probability of a tax audit averages only 1.7 percent; when combined with the modest penalties for underpayment, one would predict substantial tax avoidance.[81] Evidence also suggests that the expected fine for driving while intoxicated is on the order of one-quarter of the expected harm caused by such behavior,[82] and that total monetary sanctions imposed on corporations equal on average only 33 percent of the harms caused.[83] Given the ample opportunities that exist for augmenting penalties, as well as the possible desirability of increasing enforcement effort, society probably should raise levels of deterrence in many areas of enforcement.

19. *Future Research*

Although a significant body of research already has accumulated concerning public enforcement of law, there are several lines of inquiry that we feel merit further development, and in conclusion we comment on some of them here.

(a) The behavior and compensation of enforcement agents have not been examined in this article, but this topic is important and should be studied for two reasons. First, the incentives of

[80] There are many other public enforcement practices that are consistent with enforcement theory. For example, repeat offenders are often sanctioned more severely than first-time offenders, and individuals who report their own violations are often given a reduced penalty.

[81] In 1995 the audit rate for individual returns was 1.7 percent, as noted above, and the civil penalty for underpayment of taxes ordinarily is calculated as 20 percent of the underpayment that results from wrongful conduct (such as substantially misstating a valuation). See Andreoni, Erard, and Feinstein (1998, p. 820). Thus, for every dollar of underpayment, the expected payment, including the underpayment and the civil penalty, is only $0.0204 (= .017 × $1.20).

[82] See Kenkel (1993, p. 145). The expected fine is $12.82 and the expected harm is $47.77 (both in 1986 dollars). While the latter number may seem low, keep in mind that it is the product of the probability that a harm will occur as a result of drunken driving, and the level of harm if harm does occur. (To properly determine whether dangerous driving is underdeterred, one also would have to take into account the threat of liability from private suits brought by accident victims. But the deterrent effect of such suits will be dulled to the extent that drivers do not have sufficient assets to pay for the harms suffered by accident victims, or have liability insurance and therefore only partially bear the financial consequences of a lawsuit.)

[83] See Mark A. Cohen (1989, pp. 617–18, 658). Cohen notes, however, that he did not take into account other sanctions imposed on corporate criminals, including restitution, civil penalties, and private tort suits.

enforcement agents to discover violations is affected by the structure of their payments (notably, whether they are rewarded for finding more violations). Second, enforcement agents may be corrupted: they may accept bribes, or demand payments, in exchange for not reporting violations. Corruption tends to reduce deterrence, and therefore its presence obviously will affect the theory of optimal law enforcement.[84]

(b) We also have not discussed social norms as a general alternative to law enforcement in channeling individuals' behavior. By a social norm, we mean a rule of behavior (for example, that people should not litter or should not discriminate on the basis of race) whose violation may have the following consequences: the violator may experience an internal sanction (guilt, remorse); others may impose on the violator external, extra-legal social sanctions (gossip, ostracism); and others may experience utility or disutility from punishment of the violator. There is an emerging literature on social norms that seems worth amplifying because of the influence that social norms have on behavior, because of their role as a substitute for and supplement to formal laws, and also because of the possibility that laws themselves might influence social norms.[85]

(c) The actions of private parties to prevent crime (for example, the use of locks, the carrying of weapons, the hiring of private police) have not been considered above but obviously are important and need to be integrated with the theory of public enforcement. Notably, private effort to reduce crime can serve, at least to some extent, as a substitute for public effort; moreover, sometimes private effort is more efficient than public effort (citizens may know better where to put locks) and other times is less efficient (public authorities may know better how to assign police). The optimal coordination of private and public efforts to reduce crime needs to be examined.[86]

(d) Empirical work on law enforcement is strongly needed to better measure the deterrent effects of sanctions, especially to separate the influence of the magnitude of sanctions from their probability of application. Additionally, when the sanction is imprisonment, the effects of deterrence need to be distinguished from those of incapacitation. It also is of some interest to investigate public policy alternatives to law enforcement to reduce crime, such as job training programs and related social investment.[87]

[84] Some research has been undertaken on enforcers' incentives and the problem of corruption. See, for example, Becker and Stigler (1974), Mookherjee and Png (1995), Roger Bowles and Nuno Garoupa (1997), and Polinsky and Shavell (1999b).

[85] See, for example, *University of Pennsylvania Law Review* (1996), Posner (1997), and Richard H. McAdams (1997, pp. 343–54).

[86] There is some literature that discusses the issues raised in this paragraph; see, for example, Charles T. Clotfelter (1977; 1978) and Shavell (1991a).

[87] For empirical literature on law enforcement, focusing on crime, see the survey by Eide (1997); and see also, for example, John J. DiIulio, Jr. and Anne Morrison Piehl (1991), Daniel P. Kessler and Levitt (1999), and Levitt (1996, 1998). On the relative value of enforcement and social investment programs for controlling crime, see John J. Donohue III and Peter Siegelman (1998). A useful survey of empirical research concerning environmental enforcement is contained in Cohen (1999).

REFERENCES

Andenaes, Johannes. 1966. "The General Preventive Effects of Punishment," *U. Pennsylvania Law Rev.*, 114, pp. 949–83.

Andreoni, James. 1991. "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" *RAND J. Econ.*, 22, pp. 385–95.

Andreoni, James; Brian Erard, and Jonathan Feinstein. 1998. "Tax Compliance," *J. Econ. Lit.*, 36:2, pp. 818–60.

Arlen, Jennifer A. 1994. "The Potentially Perverse Effects of Corporate Criminal Liability," *J. Legal Stud.*, 23, pp. 833–67.

Bebchuk, Lucian Arye and Louis Kaplow. 1992. "Optimal Sanctions When Individuals Are Imperfectly Informed About the Probability of Apprehension," *J. Legal Stud.*, 21, pp. 365–70.

———. 1993. "Optimal Sanctions and Differences in Individuals' Likelihood of Avoiding Detection," *Int. Rev. Law Econ.*, 13, pp. 217–24.

Beccaria, Cesare. 1767. *On Crimes and Punishments, and Other Writings*. Richard Bellamy, ed. Translated by Richard Davies et al. NY: Cambridge U. Press, 1995 (originally published 1767).

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach," *J. Polit. Econ.*, 76, pp. 169–217.

Becker, Gary S. and George J. Stigler. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers," *J. Legal Stud.*, 3, pp. 1–18.

Bentham, Jeremy. 1789. *An Introduction to the Principles of Morals and Legislation*. In *The Utilitarians*. Rept. Garden City, NY: Anchor Books, 1973 (originally published 1789).

Block, Michael K. and Joseph Gregory Sidak. 1980. "The Cost of Antitrust Deterrence: Why Not Hang a Price Fixer Now and Then?" *Georgetown Law J.*, 68, pp. 1131–39.

Bouckaert, Boudewijn and Gerrit De Geest, eds. 1992. *Bibliography of Law and Economics*. Dordrecht: Kluwer Academic Publishers.

Bowles, Roger and Nuno Garoupa. 1997. "Casual Police Corruption and the Economics of Crime," *Int. Rev. Law Econ.*, 17, pp. 75–87.

Brown, John Prather. 1973. "Toward an Economic Theory of Liability," *J. Legal Stud.*, 2, pp. 323–49.

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

Chu, C. Y. Cyrus; Sheng-cheng Hu, and Ting-yuan Huang. 1997. "Punishing Repeat Offenders More Severely!" Unpublished manuscript.

Clotfelter, Charles T. 1977. "Public Services, Private Substitutes, and the Demand for Protection Against Crime," *Amer. Econ. Rev.*, 67, pp. 867–77.

———. 1978. "Private Security and the Public Safety," *J. Urban Econ.*, 5, pp. 388–402.

Cohen, Mark A. 1989. "Corporate Crime and Punishment: A Study of Social Harm and Sentencing Practice in the Federal Courts, 1984–1987," *Amer. Criminal Law Rev.*, 26, pp. 605–60.

———. 1999. "Monitoring and Enforcement of Environmental Policy," in *The International Yearbook of Environmental and Resource Economics 1999/2000*. Henk Folmer and Tom Tietenberg, eds. Cheltenham, UK: Edward Elgar, forthcoming.

Cooter, Robert D. and Daniel L. Rubinfeld. 1989. "Economic Analysis of Legal Disputes and Their Resolution," *J. Econ. Lit.*, 27, pp. 1067–97.

Craswell, Richard and John E. Calfee. 1986. "Deterrence and Uncertain Legal Standards," *J. Law Econ. Organ.*, 2, pp. 279–303.

DiIulio, Jr., John J. and Anne Morrison Piehl. 1991. "Does Prison Pay? The Stormy National Debate over the Cost-effectiveness of Imprisonment," *Brookings Rev.*, 9, pp. 28–35.

Donohue III, John J. and Peter Siegelman. 1998. "Allocating Resources among Prisons and Social Programs in the Battle against Crime," *J. Legal Stud.*, 27, pp. 1–43.

Eide, Erling. 1997. "Economics of Criminal Behavior: Survey and Bibliography." U. Oslo, Institute for Private Law, Law and Econ. Working Paper C 5.

Ehrlich, Isaac. 1981. "On the Usefulness of Controlling Individuals: An Economic Analysis of Rehabilitation, Incapacitation and Deterrence," *Amer. Econ. Rev.*, 71, pp. 307–32.

Friedman, David D. 1981. "Reflections on Optimal Punishment, or: Should the Rich Pay Higher Fines?" *Res. Law Econ.*, 3, pp. 185–205.

———. 1995. "Making Sense of English Law Enforcement in the Eighteenth Century," *U. Chicago Law School Roundtable*, 2, pp. 475–505.

Garoupa, Nuno. 1997. "The Theory of Optimal Law Enforcement," *J. Econ. Surveys*, 11, pp. 267–95.

Grossman, Gene M. and Michael L. Katz. 1983. "Plea Bargaining and Social Welfare," *Amer. Econ. Rev.*, 73, pp. 749–57.

Kahneman, Daniel; Paul Slovic, and Amos Tversky, eds. 1982. *Judgment under Uncertainty: Heuristics and Biases*, Cambridge and NY: Cambridge U. Press.

Kaplow, Louis. 1990a. "Optimal Deterrence, Uninformed Individuals, and Acquiring Information About Whether Acts are Subject to Sanctions," *J. Law Econ. Organ.*, 6, pp. 93–128.

———. 1990b. "A Note on the Optimal Use of Nonmonetary Sanctions," *J. Public Econ.*, 42, pp. 245–47.

———. 1992. "The Optimal Probability and Magnitude of Fines for Acts That Definitely Are Undesirable," *Int. Rev. Law Econ.*, 12, pp. 3–11.

Kaplow, Louis and Steven Shavell. 1994a. "Accuracy in the Determination of Liability," *J. Law Econ.*, 37, pp. 1–15.

———. 1994b. "Optimal Law Enforcement with Self-Reporting of Behavior," *J. Polit. Econ.*, 102, pp. 583–606.

Kenkel, Donald S. 1993. "Do Drunk Drivers Pay Their Way? A Note on Optimal Penalties for Drunk Driving," *J. Health Econ.*, 12, pp. 137–49.

Kessler, Daniel P. and Steven D. Levitt. 1999. "Using Sentence Enhancements to Distinguish between Deterrence and Incapacitation," *J. Law Econ.*, forthcoming.

Kobayashi, Bruce H. and John R. Lott, Jr. 1992. "Low-Probability-High-Penalty Enforcement Strategies and the Efficient Operation of the Plea-Bargaining System," *Int. Rev. Law Econ.*, 12, pp. 69–77.

Kornhauser, Lewis A. 1982. "An Economic Analysis of the Choice between Enterprise and Personal Liability for Accidents," *California Law Rev.*, 70, pp. 1345–92.

Laffont, Jean-Jacques and Jean Tirole. 1993. *The Theory of Incentives in Procurement and Regulation.* Cambridge, MA: MIT Press.

Landes, William M. 1971. "An Economic Analysis of the Courts," *J. Law Econ.*, 4, pp. 61–107.

Landes, William M. and Richard A. Posner. 1975. "The Private Enforcement of Law," *J. Legal Stud.*, 4, pp. 1–46.

Levitt, Steven D. 1996. "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation," *Quart. J. Econ.*, 111, pp. 319–51.

———. 1998. "Why Do Increased Arrest Rates Appear to Reduce Crime: Deterrence, Incapacitation, or Measurement Error?" *Econ. Inq.*, 36, pp. 353–72.

Lewin, Jeff L. and William N. Trumbull. 1990. "The Social Value of Crime?" *Int. Rev. Law Econ.*, 10, pp. 271–84.

McAdams, Richard H. 1997. "The Origin, Development, and Regulation of Norms," *Michigan Law Rev.*, 96, pp. 338–433.

Malik, Arun S. 1990. "Avoidance, Screening and Optimum Enforcement," *RAND J. Econ.*, 21, pp. 341–53.

———. 1993. "Self-Reporting and the Design of Policies for Regulating Stochastic Pollution," *J. Environmental Econ. Manage.*, 24, pp. 241–57.

Miceli, Thomas J. 1996. "Plea Bargaining and Deterrence: An Institutional Approach," *European J. Law Econ.*, 3, pp. 249–64.

Montesquieu. 1748. *The Spirit of the Laws.* Rept. ed. Berkeley: U. California Press, 1977 (originally published 1748).

Mookherjee, Dilip. 1997. "The Economics of Enforcement," in *Issues in Economic Theory and Policy, Essays in Honor of Tapas Majumdar.* Amitava Bose, Mihir Rakshit, and Anup Sinha, eds. New Delhi: Oxford U. Press, pp. 202–49.

Mookherjee, Dilip and I. P. L. Png. 1992. "Monitoring vis-à-vis Investigation in Enforcement of Law," *Amer. Econ. Rev.*, 82, pp. 556–65.

———. 1994. "Marginal Deterrence in Enforcement of Law," *J. Polit. Econ.*, 102, pp. 1039–66.

———. 1995. "Corruptible Law Enforcers: How Should They Be Compensated?" *Econ. J.*, 105, pp. 145–59.

Newman, Harry A. and David W. Wright. 1990. "Strict Liability in a Principal-Agent Model," *Int. Rev. Law Econ.*, 10, pp. 219–31.

Png, I. P. L. 1986. "Optimal Subsidies and Damages in the Presence of Judicial Error," *Int. Rev. Law Econ.*, 6, pp. 101–05.

Polinsky, A. Mitchell. 1980a. "Private versus Public Enforcement of Fines," *J. Legal Stud.*, 9, pp. 105–27.

———. 1980b. "Strict Liability vs. Negligence in a Market Setting," *Amer. Econ. Rev.: Papers and Proceedings*, 70, pp. 363–70.

Polinsky, A. Mitchell and Daniel L. Rubinfeld. 1988. "The Deterrent Effects of Settlements and Trials," *Int. Rev. Law Econ.*, 8, pp. 109–16.

———. 1989. "A Note on Optimal Public Enforcement with Settlements and Litigation Costs," *Res. Law Econ.*, 12, pp. 1–8.

———. 1991. "A Model of Optimal Fines for Repeat Offenders," *J. Public Econ.*, 46, pp. 291–306.

Polinsky, A. Mitchell and Steven Shavell. 1979. "The Optimal Tradeoff between the Probability and Magnitude of Fines," *Amer. Econ. Rev.*, 69, pp. 880–91.

———. 1984. "The Optimal Use of Fines and Imprisonment," *J. Public Econ.*, 24, pp. 89–99.

———. 1991. "A Note on Optimal Fines When Wealth Varies Among Individuals," *Amer. Econ. Rev.*, 81, pp. 618–21.

———. 1992. "Enforcement Costs and the Optimal Magnitude and Probability of Fines," *J. Law Econ.*, 35, pp. 133–48.

———. 1993. "Should Employees Be Subject to Fines and Imprisonment Given the Existence of Corporate Liability?" *Int. Rev. Law Econ.*, 13, pp. 239–57.

———. 1998a. "On Offense History and the Theory of Deterrence," *Int. Rev. Law Econ.*, 18, pp. 305–24.

———. 1998b. "The Fairness of Sanctions: Some Implications for Optimal Enforcement Policy," Working Paper 167, John M. Olin Program in Law and Econ., Stanford Law School. *Amer. Law Econ. Rev.* forthcoming.

———. 1999a. "On the Disutility and Discounting of Imprisonment and the Theory of Deterrence," *J. Legal Stud.*, 28, pp. 1–16.

———. 1999b. "Corruption and Optimal Law Enforcement." Working Paper 171, John M. Olin Program in Law and Econ., Stanford Law School.

Posner, Richard A. 1980a. "Optimal Sentences for White-Collar Criminals," *Amer. Criminal Law Rev.*, 17, pp. 409–18.

———. 1980b. "Retribution and Related Concepts of Punishment," *J. Legal Stud.*, 9, pp. 71–92.

———. 1985. "An Economic Theory of the Criminal Law," *Columbia Law Rev.*, 85, pp. 1193–231.

———. 1997. "Social Norms and the Law: An Economic Approach," *Amer. Econ. Rev.: Papers and Proceedings*, 87, pp. 365–69.

Reinganum, Jennifer F. 1988. "Plea Bargaining and Prosecutorial Discretion," *Amer. Econ. Rev.*, 78, pp. 713–28.

Rubinstein, Ariel. 1979. "An Optimal Conviction Policy for Offenses that May Have Been Committed by Accident," in *Applied Game Theory.* S. J. Brams, A. Schotter, and G. Schwödiauer, eds. Wurzburg: Physica-Verlag, pp. 406–13.

Sah, Raaj K. 1991. "Social Osmosis and Patterns of Crime," *J. Polit. Econ.*, 99, pp. 1272–95.

Segerson, Kathleen and Tom Tietenberg. 1992. "The Structure of Penalties in Environmental Enforcement: An Economic Analysis," *J. Environmental Econ. Manage.*, 23, pp. 179–200.

Shavell, Steven. 1980. "Strict Liability versus Negligence," *J. Legal Stud.*, 9, pp. 1–25.

———. 1982. "On Liability and Insurance," *Bell J. Econ.*, 13, pp. 120–32.

———. 1985. "Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent," *Columbia Law Rev.*, 85, pp. 1232–62.

———. 1987a. "A Model of Optimal Incapacitation," *Amer. Econ. Rev.: Papers and Proceedings*, 77, pp. 107–10.

———. 1987b. "The Optimal Use of Nonmonetary Sanctions as a Deterrent," *Amer. Econ. Rev.*, 77, pp. 584–92.

———. 1991a. "Individual Precautions to Prevent Theft: Private versus Socially Optimal Behavior," *Int. Rev. Law Econ.*, 11, pp. 123–32.

———. 1991b. "Specific versus General Enforcement of Law," *J. Polit Econ.*, 99, pp. 1088–108.

———. 1992. "A Note on Marginal Deterrence," *Int. Rev. Law Econ.*, 12, pp. 345–55.

———. 1993. "The Optimal Structure of Law Enforcement," *J. Law Econ.*, 36, pp. 255–87.

———. 1997a. "The Fundamental Divergence Between the Private and the Social Motive to Use the Legal System," *J. Legal Stud.*, 26, pp. 575–612.

———. 1997b. "The Optimal Level of Corporate Liability Given the Limited Ability of Corporations to Penalize Their Employees," *Int. Rev. Law Econ.*, 17, pp. 203–13.

Spier, Kathryn E. 1997. "A Note on the Divergence Between the Private and the Social Motive to Settle Under a Negligence Rule," *J. Legal Stud.*, 26, pp. 613–21.

Stigler, George J. 1970. "The Optimum Enforcement of Laws," *J. Polit. Econ.*, 78, pp. 526– 36.

Sykes, Alan O. 1981. "An Efficiency Analysis of Vicarious Liability Under the Law of Agency," *Yale Law J.*, 91, pp. 168–206.

University of Pennsylvania Law Review. 1996. "Symposium: Law, Economics, and Norms," *U. Pennsylvania Law Rev.*, 144, pp. 1643–2339.

US Department of Justice. 1988. *Report to the Nation on Crime and Justice, Technical Appendix.* Washington, DC: US Dept. of Justice, Bureau of Justice Statistics, 2nd ed., NCJ–112011.

———. 1997a. *Sourcebook of Criminal Justice Statistics—1996.* Washington, DC: US Dept. of Justice, Office of Justice Statistics, Bureau of Justice Statistics, NCJ–165361.

———. 1997b. *Uniform Crime Reports for the United States, 1996.* Washington, DC: US Dept. of Justice, FBI.

———. 1998. *Profile of Jail Inmates 1996.* Washington, DC: US Dept. of Justice, Office of Justice Programs, NCJ–164620.

Wilde, Louis L. 1992. "Criminal Choice, Nonmonetary Sanctions, and Marginal Deterrence: A Normative Analysis," *Int. Rev. Law Econ.*, 12, pp. 333–44.

Wilson, James Q. and Richard J. Herrnstein. 1985. *Crime and Human Nature.* NY: Simon and Schuster.

Zimring, Franklin E. and Gordon J. Hawkins. 1973. *Deterrence.* Chicago: U. Chicago Press.