SIEPR Discussion Paper No. 05-16

PUBLIC ENFORCEMENT OF LAW

By
A. MITCHELL POLINSKY
and
STEVEN SHAVELL

May, 2006

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

To appear in: Lawrence Blume and Steven Durlauf (Editors), *The New Palgrave Dictionary of Economics* (Second Edition, forthcoming).

May 2006

Public Enforcement of Law

A. Mitchell Polinsky and Steven Shavell

Abstract: This entry for the forthcoming *The New Palgrave Dictionary of Economics* (Second Edition) surveys the economic analysis of public enforcement of law — the use of public agents (inspectors, tax auditors, police, prosecutors) to detect and to sanction violators of legal rules. We first discuss the basic elements of the theory: the probability of imposition of sanctions, the magnitude and form of sanctions (fines, imprisonment), and the rule of liability. We then examine a variety of extensions, including the costs of imposing fines, mistake, marginal deterrence, settlement, self-reporting, repeat offenses, and incapacitation.

Keywords: fines; imprisonment; probability of detection; sanctions; crime; enforcement; strict liability; fault-based liability; mistake; marginal deterrence; settlement; self-reporting; repeat offenses; incapacitation

JEL classification: H23; K14; K42

---

* Stanford Law School and Harvard Law School, respectively. Both authors also are affiliated with the National Bureau of Economic Research. Polinsky's research was supported by the John M. Olin Program in Law and Economics at Stanford Law School, and Shavell's research was supported by the John M. Olin Center for Law, Economics, and Business at Harvard Law School.

In this entry we consider the theory of public enforcement of law — the use of public agents (inspectors, tax auditors, police, prosecutors) to detect and to sanction violators of legal rules. After briefly discussing the rationale for public (as opposed to private) enforcement, we present the basic elements of the theory: the probability of imposition of sanctions, the magnitude and form of sanctions (fines, imprisonment), and the rule of liability. We then examine a variety of extensions of the central theory, including the costs of imposing fines, mistake, marginal deterrence, settlement, self-reporting, repeat offenses, and incapacitation. (For a fuller treatment of the material in this entry, see Polinsky and Shavell, 2007.)

Before proceeding, we note that economically-oriented analysis of public law enforcement dates primarily from the eighteenth century contribution of Bentham (1789), whose analysis of deterrence was sophisticated and expansive. After Bentham, the subject of enforcement lay essentially dormant in economic scholarship until Gary Becker (1968) published a highly influential article, which has led to a voluminous literature.

RATIONALE FOR PUBLIC ENFORCEMENT. A basic question is why there is a need for public enforcement of law (see generally Becker and Stigler, 1974; Landes and Posner, 1975; Polinsky, 1980a). In particular, why not rely solely on private suits brought by victims? The answer depends importantly on the locus of information about the identity of injurers. When victims of harm naturally possess knowledge of the identity of injurers, allowing private suits for damages will motivate victims to sue and thus harness the information they have for purposes of law enforcement. This may explain why the enforcement of contractual obligations and of accident law is primarily private. When victims do not know who caused harm, however, or when finding injurers is difficult, society may need to rely instead on public investigation and

prosecution; this is broadly true of crimes and of many violations of environmental and safety regulations.

BASIC FRAMEWORK FOR ANALYZING PUBLIC ENFORCEMENT. An individual who commits a harmful act obtains a gain and also faces the risk of being caught and sanctioned. The form of sanction could be a fine or a prison term. Fines generally will be treated as socially costless because they are mere transfers of money, whereas imprisonment will be considered as socially costly because of the expense of operating prisons and the disutility suffered by those imprisoned. The higher is the probability of detecting violators, the more resources the state must devote to enforcement.

We assume that social welfare equals the sum of individuals' expected utilities. If individuals are risk neutral, social welfare can be expressed as the gains individuals obtain from committing their harmful acts, less the harms caused, and less the costs of law enforcement. The enforcement authority's problem is to maximize social welfare by choosing enforcement expenditures (or, equivalently, a probability of detection), the form of sanctions, and their level.

FINES. Suppose that the sanction is a fine and that individuals are risk neutral. If the probability of detection p is taken as fixed, then the optimal fine is the harm h divided by the probability, that is, h/p; for then the expected fine $p(h/p)$ equals h. This fine is optimal because, facing it, an individual will commit a harmful act if, and only if, the gain he would derive exceeds the harm he would cause. Such behavior is first-best. The fundamental formula h/p essentially was noted by Bentham (1789) and it has been observed by many others since.

If the probability of detection can be varied, the optimal fine is maximal, f_M, as

emphasized by Becker (1968). If the fine were not maximal, society could save enforcement costs by simultaneously raising the fine and lowering the probability without affecting the level of deterrence. If $f < f_M$, then raise the fine to f_M and lower the probability from p to $(f/f_M)p$; the expected fine is still pf, so that deterrence is maintained, but expenditures on enforcement are reduced, implying that social welfare rises.

The optimal probability p of imposing a fine is low in the sense that it results in some underdeterrence; that is, the optimal p is such that the expected fine pf_M is less than the harm h (Polinsky and Shavell, 1984). The reason is to economize on enforcement resources. In particular, if pf_M equals h, behavior will be ideal, meaning that the individuals who are just deterred obtain gains essentially equal to the harm. These are the individuals who would be led to commit the harmful act if p were lowered slightly. That in turn must be socially beneficial because these individuals cause no net social losses (their gains essentially equal the harm), but reducing p saves enforcement costs. How much pf_M should be lowered below h depends on the saving in enforcement costs from reducing p compared to the net social costs of underdeterrence that will result if p is lowered non-trivially.

If individuals are risk averse, the optimal fine may be well below the maximal fine, as first shown in Polinsky and Shavell (1979); see also Kaplow (1992). This is because a high fine would impose substantial risk-bearing costs on individuals who commit harmful acts. If $f < f_M$, it is still true that f can be raised and p lowered so as to maintain deterrence, but because of risk aversion, this now implies that pf falls, meaning that fine revenue falls. The reduction in fine revenue reflects the disutility caused by imposing greater risk on risk-averse individuals. The decline in fine revenue could more than offset the savings in enforcement expenditures, causing social welfare to be lower.

IMPRISONMENT. Now suppose that the sanction is imprisonment. If the probability of detection is fixed, there is not a simple formula for the optimal imprisonment term; see Polinsky and Shavell (1984). The optimal term could be such that there is either underdeterrence or overdeterrence. On one hand, a relatively low imprisonment term, implying underdeterrence, might be socially desirable because imprisonment costs are reduced for those individuals who commit harmful acts. On the other hand, a relatively high term, implying overdeterrence, might be socially desirable because imprisonment costs are reduced due to fewer individuals committing harmful acts, even if some of these deterred individuals would have obtained gains exceeding the harm.

If the probability of detection can be varied and individuals are risk neutral in imprisonment, then the optimal imprisonment term is maximal. The reasoning is similar to that employed above: if the imprisonment term were not maximal, it could be raised and the probability of detection lowered so as to keep the expected prison term constant; neither individual behavior nor the costs of imprisonment are affected, but enforcement expenditures fall.

If, instead, individuals are risk averse in imprisonment (the disutility of each additional year of imprisonment grows with the number of years in prison), there is a stronger argument for setting the imprisonment sanction maximally (Polinsky and Shavell, 1999). Now when the imprisonment term is raised, the probability of detection can be lowered more than in the risk-neutral case without reducing deterrence. Thus, not only are there greater savings in enforcement expenditures, but also the costs of imposing imprisonment sanctions decline because the expected prison term falls.

Last, suppose that individuals are risk preferring in imprisonment (the disutility of each additional year of imprisonment declines with the number of years in prison). This possibility seems particularly important: the first years of imprisonment may create unusually high disutility, due to brutalization of the prisoner or to the stigma of having been imprisoned at all. Individuals' positive time discount rates, which are thought to be especially significant for criminals, also make the disutility of later years less significant. In the case of risk-preferring individuals, the optimal prison term may well be less than maximal: if the sentence were raised, the probability that maintains deterrence could not be lowered proportionally, implying that the expected prison term would rise. Thus, although there would be enforcement cost savings, they might not be great enough to offset the increased sanctioning costs.

When the sanction is imprisonment, the optimal probability of detection may be such that there is either underdeterrence or overdeterrence. On one hand, the motive to lower the probability is reinforced relative to the case of fines because imprisonment costs, as well as detection costs, decline if fewer offenders are caught. On the other hand, raising the probability of detection results in fewer offenders, which, everything else equal, decreases imprisonment costs because fewer are imprisoned. Either effect may dominate.

FINES VERSUS IMPRISONMENT. Fines generally are preferable to prison terms as a means of deterrence, since fines are socially cheaper sanctions to impose (Becker, 1968; Polinsky and Shavell, 1984). Hence, fines should be employed to the greatest extent possible — until a party's wealth is exhausted — before imprisonment is imposed. Further, imprisonment should be used as a sanction only if the harm prevented by the added deterrence is sufficiently great.

FAULT-BASED LIABILITY. Our discussion thus far has presumed that liability is strict — is imposed whenever harm occurs — but liability may instead be based on fault — imposed only when behavior was found to be socially undesirable. Fault-based liability, like strict liability, can induce individuals to behave properly, but fault-based liability possesses an advantage when individuals are risk averse: if they act responsibly, they will not be found at fault, so will not bear the risk of being sanctioned. Similarly, fault-based liability is advantageous when the sanction is imprisonment, for then again individuals may be led to behave optimally without the actual imposition of sanctions, and thus without social costs being incurred (Shavell, 1987b). To the extent that mistakes are made in determining fault, however, these two advantages are reduced.

Fault-based liability is more difficult to implement because it requires more information than strict liability. To apply fault-based liability, the enforcement authority must be able to determine the proper fault standard — that is, socially desirable behavior — and it must ascertain whether the defendant's conduct was in compliance with the fault standard. Under strict liability, the authority need only measure harm. (Moreover, for reasons we discuss below, strict liability encourages better decisions by injurers regarding their level of participation in harm-creating activities.)

* * *

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

ACCIDENTAL HARMS. We have been implicitly assuming that individuals decide whether or not to commit acts that cause harm with certainty, that is, they decide whether or not to cause intentional harms. In many circumstances, however, harms are accidental — they occur only with a probability. Essentially all that we have said above applies in a straightforward way when harms are accidental.

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 an act that increases the chance of harm — storing chemicals in a substandard tank — or on the basis of the actual occurrence of harm — if the tank ruptures and results in a spill. In principle, either approach can achieve optimal deterrence — by setting the (expected) sanction equal to expected harm if liability is imposed whenever a dangerous act is committed, or equal to actual harm if liability is imposed only if harm occurs.

Several factors are relevant to the choice between act-based and harm-based sanctions (Shavell, 1993). First, act-based sanctions need not be as high as harm-based sanctions to accomplish a given level of deterrence (expected harm is less than actual harm), and thus offer an advantage because of parties' limited assets. Second, because act-based sanctions can accomplish a given level of deterrence with lower sanctions, they are preferable when parties are risk averse. Third, either 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), or harm-based sanctions may be easier to implement (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, favoring harm-based sanctions.

COSTS OF IMPOSING FINES. The costs borne by enforcement authorities in imposing fines should be reflected in the fine. Recall that if the probability of detection is taken as fixed and individuals are risk neutral, the optimal fine is h/p, the harm divided by the probability of detection. Now suppose there is a public cost k of imposing a fine. The optimal fine then becomes $h/p + k$; the cost k should be added to the fine that would otherwise be desirable (Becker, 1968; Polinsky and Shavell, 1992). 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 $h/p + k$, the individual's expected fine is $p[h/p + k] = h + pk$, leading him to commit the harmful act if and only if his gain exceeds the expected total social cost of his act.

Not only does the state bear 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.

LEVEL OF ACTIVITY. In many settings in which harm may occur, an individual not only chooses whether to commit a harmful act when engaging in an activity, but also the level at which to engage in the activity. Drivers decide how careful to be while driving, as well as how many miles to drive; similarly, firms choose safety precautions as well as their level of output. The socially optimal activity level is such that the actor's marginal utility from the activity just equals the marginal expected harm caused by the activity (assuming optimal care is taken).

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.

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 they tend to behave so as not to be found at fault. As a consequence, they will choose an excessive level of activity (Shavell, 1980). Driving more miles increases expected harm, but this effect generally will be ignored under fault-based liability.

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. 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 level of output will be excessive (Polinsky, 1980b).

Relatedly, 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 liability. Hence, regulations of this sort 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.

MISTAKES. An individual who should be found liable might mistakenly be acquitted. Conversely, an individual who should not be found liable might mistakenly be convicted. For an individual who has been detected, let the probabilities of these errors be ε_A and ε_C, respectively.

Given the probability of detection p and the chances of these types of error, an individual will commit the wrongful act if and only if his gain g net of his expected fine if he does commit it exceeds his expected fine if he does not commit it, namely, when $g - p(1 - \varepsilon_A)f > -p\varepsilon_C f$, or, equivalently, when $g > (1 - \varepsilon_A - \varepsilon_C)pf$.

As emphasized by Png (1986), both types of error reduce deterrence: the term $(1 - \varepsilon_A - \varepsilon_C)pf$ is declining in both ε_A and ε_C. The first type of error diminishes deterrence because it lowers the expected fine if an individual violates the law. The second type of error lowers deterrence because it reduces the difference between the expected fine from violating the law and not violating it — the greater is ε_C, the smaller the increase in the expected fine if one violates the law.

Because mistakes dilute deterrence, they reduce social welfare. Specifically, to achieve any level of deterrence, the probability p must be higher to offset the effect of errors. Mistaken convictions have the additional effect of discouraging socially desirable participation in the activity. Consequently, expenditures made to reduce errors may be socially beneficial (Kaplow and Shavell, 1994a).

Two other points regarding the implications of mistake are worth noting. First, if individuals are risk averse, the possibility of mistakes of either type generally lowers optimal sanctions (Block and Sidak, 1980). Second, as stressed by Craswell and Calfee (1986), individuals will often have a motive to take excessive precautions under fault-based liability in order to reduce the chance of being found erroneously at fault.

GENERAL ENFORCEMENT. In many settings, enforcement may be said to be general in the sense that several different types of violations will 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 one who goes through a red light or who speeds, and a tax auditor may detect a variety of infractions when he examines a tax return. (In contrast, if enforcement is specific, the probability is chosen independently for each type of harmful act.)

When enforcement is general, the optimal sanction rises with the level of harm and is maximal only for relatively high harms (Shavell, 1991; Mookherjee and Png, 1992). To see why, 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 (that is, p applies regardless of h), the optimal fine schedule is h/p, provided that h/p is feasible; otherwise the optimal fine is maximal. This schedule is obviously optimal given p because it implies that the expected fine equals harm, thereby inducing ideal behavior whenever that is possible. That sanctions should rise with the severity of harm up to a maximum when enforcement is general also holds if the sanction is imprisonment and if liability is fault-based.

MARGINAL DETERRENCE. In many circumstances, a person 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 to kidnap a person or also to kill the kidnap victim. In such contexts, sanctions influence which harmful acts individuals choose to commit (as well as whether to commit any harmful act). Marginal deterrence is said to occur when a more harmful act is deterred because its sanction exceeds that for a less harmful act (Stigler, 1970; Shavell, 1992; Wilde, 1992; Mookherjee and Png, 1994).

Other things being equal, it is socially desirable that enforcement policy creates marginal deterrence so that, when harmful acts do occur, less harm is done. One way to accomplish

marginal deterrence is for sanctions to rise with the magnitude of harm, which means that sanctions generally will not be maximal. However, fostering marginal deterrence may conflict with achieving overall deterrence: in order 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.

Note that marginal deterrence also can be promoted by increasing the probability of detection. Kidnappers can be better deterred from killing their victims if more police resources are devoted to apprehending kidnappers who murder their victims than to those who do not.

PRINCIPAL-AGENT RELATIONSHIP. Although we have assumed that an injurer is a single actor, injurers often are more appropriately characterized as collective entities, and specifically as 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 prior conclusions carry over to the sanctioning of principals. Notably, if a risk-neutral principal faces an expected fine equal to harm done, he 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 appropriate incentives to reduce harm (Newman and Wright, 1990; but see Arlen, 1994).

An issue that arises when there are principals and agents concerns the allocation of financial sanctions between the two parties. It is apparent that the particular allocation of sanctions does not matter when the parties can reallocate the 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 is imposed on the principal, he

will retain it and not impose an internal sanction on the agent. Thus, the post-contract sanctions that the agent bears are not 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 (Sykes, 1981; Kornhauser, 1982). 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 harm that they can cause, even if the principal's assets are sufficient to pay the optimal fine (Polinsky and Shavell, 1993). That an agent's assets are limited means that the principal may be unable to control him adequately through the use of contractually-determined penalties, which can only be monetary. In such circumstances, it may be socially valuable to use the threat of a jail sentence to better control agents' misconduct.

SETTLEMENTS. It is common for lawbreakers to settle with public enforcement authorities prior to being found liable in a 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.) Both parties might prefer an out-of-court settlement to avoid the cost of a trial and to eliminate the risks inherent in the trial outcome (Cooter and Rubinfeld, 1989; and on plea bargaining, see Reinganum, 1988, and Miceli, 1996).

These advantages 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 (Polinsky and Rubinfeld, 1988). The state may be able to offset this effect by increasing the level of sanctions.

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, settlements will tend to reflect the average harm caused (resulting in high-harm injurers being underdeterred, and vice versa), whereas trial outcomes may better approximate the actual harm. Second, settlements hinder the amplification and development of the law through the setting of precedents. Third, if the sanction is imprisonment and defendants are risk averse, settlements necessitate longer terms than the expected sentence at trial in order to maintain deterrence, and thus increase public expenditures. On the social welfare evaluation of settlement, see, for example, Shavell (1997) and Spier (1997).

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

Self-reporting can be induced by lowering the sanction for individuals who disclose their own violations (Kaplow and Shavell, 1994b). Moreover, the reward for self-reporting can be

made small, so that deterrence is only negligibly reduced. For example, if a risk-neutral individual commits a violation and does not self report, his expected fine is *pf*. If he self-reports, the fine can be set just below *pf*, say at $pf - \varepsilon$, where $\varepsilon > 0$ is small. Then the individual will want to self-report but the deterrent effect of the sanction will be essentially the same as if he did not self-report.

There are several advantages of self-reporting. First, self-reporting reduces enforcement costs because the enforcement authority does not have to identify and prove who the violator was. Second, self-reporting reduces risk (a relatively high sanction imposed with a relatively low probability is replaced by a certain punishment), and thus is advantageous if injurers are risk averse. Third, self-reporting may allow harm to be mitigated (early notice of an oil spill may facilitate its containment).

REPEAT OFFENDERS. In practice, the law often sanctions repeat offenders more severely than first-time offenders. This policy cannot be socially advantageous if deterrence always induces first-best behavior. For if the expected sanction for an offense equals its harm, then raising the sanction because an offender has a record of sanctions would overdeter him. Only if deterrence is inadequate is it possibly desirable to condition sanctions on offense history to increase deterrence. But as we observed above, it usually will be worthwhile for the state to tolerate some underdeterrence in order to reduce enforcement expenses.

If there is underdeterrence, making sanctions depend on offense history may be beneficial. First, the use of offense history may create an additional incentive not to violate the law: if getting caught implies not only an immediate sanction, but also a higher sanction for any future violation, an individual will, everything else equal, be deterred to a greater extent

(Polinsky and Shavell, 1998). 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 to impose higher sanctions on them (Rubinstein, 1979; Polinsky and Rubinfeld, 1991). In addition, if repeat offenders have higher propensities to commit violations, they are more likely to be worth incapacitating by imprisonment (see below).

IMPERFECT KNOWLEDGE ABOUT THE PROBABILITY AND MAGNITUDE OF SANCTIONS. Individuals might not know the true probability of a sanction because the enforcement authority refrains from publishing information about the probability (perhaps hoping that individuals will believe it to be higher than it is in fact); or because the probability depends on factors that individuals do not fully understand; or because probabilities are difficult to assess. Also, individuals may have incomplete knowledge of the true magnitude of sanctions, particularly if the levels of sanctions are discretionary.

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 a sanction, account must be taken of the relationship between the actual and the perceived variables (Bebchuk and Kaplow, 1992; Kaplow, 1990). For example, if enforcement resources are increased in order to raise the probability of detection, there might be a delay before this increase is perceived by individuals, making such an investment less worthwhile.

INCAPACITATION. Society may reduce harm not only through deterrence, but also 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, businesses can lose their rights to operate in certain markets, and the like.

Suppose that the sole function of imprisonment is to incapacitate. Then it will be desirable to keep someone imprisoned as long as the reduction in criminal harm from incapacitating him exceeds the cost of imprisonment (Shavell, 1987c). Although this condition could hold for a long period, it often will not because the proclivity to commit crimes appears to decline sharply with age.

As a matter of economic logic, the incapacitation rationale might imply that a person should be imprisoned even if he has not committed a crime, because the danger he poses to society makes incapacitating him worthwhile. In practice, however, the commission of a harmful act may be a good basis for predicting a person's future behavior, in which case the incapacitation rationale would suggest imprisoning an individual only if he has committed such an act.

Two observations are worth noting about the relationship between the incapacitation goal and the deterrence goal. First, when enforcement is based on incapacitation, the optimal magnitude of the sanction is independent of the probability of apprehension, which contrasts with the case when enforcement is based on deterrence. Second, when enforcement is deterrence-oriented, the probability and magnitude of sanctions depend on the ability to deter, and if this ability is limited (as, for instance, with the insane), a low expected sanction may be optimal, whereas a high sanction still might be called for to incapacitate.

CORRUPTION. One form of corruption in the enforcement process is bribery, in which an enforcer accepts a payment in return for not reporting a violation (or for reducing the mandated sanction for the violation). A second form of corruption is framing and framing-related extortion, in which an enforcement agent may frame an innocent individual or threaten to frame him in order to extort money from him. On corruption of law enforcement, see Bowles and Garoupa (1997) and Polinsky and Shavell (2001) (and on corruption more generally, see, for example, Shleifer and Vishny, 1993, and Rose-Ackerman, 1999).

Bribery dilutes deterrence of violations of law because it results in a lower payment by an individual than the sanction for the offense. Framing and framing-related extortion also dilute deterrence. The reason is that framing and extortion imply that those who act innocently face an expected sanction, so that the difference between the expected sanction if an individual commits a violation and if he does not is lessened. (This point is essentially the same as the earlier observation that mistaken convictions dilute deterrence.)

One way to reduce corruption is to impose fines (or imprisonment sentences) on individuals caught engaging in bribery, extortion, and framing. Corruption also can be reduced by paying enforcers rewards for reporting violations. Such payments will reduce their incentive to accept bribes because they will sacrifice their rewards if they fail to report violations. But high rewards give enforcers a greater incentive to frame innocent individuals. A third way to control corruption is to pay enforcers more than their reservation wage (that is, to pay them an efficiency wage). Then they would have more to lose if punished for corrupt behavior and denied future employment.

A natural question is whether the deterrence-diluting effects of corruption can be offset

by raising the fine on offenders. In the basic risk-neutral model of enforcement, it is not possible to raise the fine because the optimal fine is maximal. More realistically, however, the optimal fine is less than maximal for a variety of reasons, including those related to risk aversion, marginal deterrence, and general enforcement. While it would then be possible to raise the fine to offset the deterrence-diluting effects of corruption, doing so would lead to social costs (for example, by imposing greater risk).

COSTLY OBSERVATION OF WEALTH. Individuals and firms may be able to hide assets from government enforcers, including by hoarding cash, transferring assets to relatives or related legal entities, or moving money to offshore bank accounts. Consequently, an individual's level of wealth might not be able to be observed at all, or only after a costly audit.

Suppose first that the enforcement authority employs fines as sanctions and can audit an individual who claims that he cannot pay the fine (Polinsky, 2004). The optimal fine for misrepresenting one's wealth level equals the fine for the offense divided by the audit probability, and therefore generally exceeds the fine for the offense. This is a natural generalization of the formula for the optimal fine when the probability of detection is fixed, which is the harm divided by the probability. Auditing is valuable because it reduces misrepresentation of wealth and thereby increases deterrence.

Next suppose that the enforcement authority cannot observe wealth because the cost of an audit is prohibitively high (Levitt, 1997; Polinsky, 2006). If the authority would have used fines alone if it could have observed wealth at no cost, it would have imposed a higher fine on higher-wealth individuals. It obviously cannot do this when wealth is unobservable. Instead, it may be desirable to use the threat of an imprisonment sentence to induce individuals capable of

paying a higher fine to do so. Alternatively, the enforcement authority might have used both fines and imprisonment if it could have observed wealth at no cost. Perhaps surprisingly, the inability to observe wealth might not be detrimental in this case. The reason is that the mix of fines and imprisonment that would be chosen when wealth is observable might impose a higher burden (though a lower fine) on low-wealth individuals. Then, high-wealth individuals will naturally want to identify themselves. Specifically, they will prefer to pay a higher fine and bear a shorter imprisonment sentence than to masquerade as low-wealth individuals, who will bear longer imprisonment sentences and a higher overall burden.

SOCIAL NORMS. To some extent, social norms and morality are substitutes for public law enforcement because they encourage in significant ways the attainment of desired behavior (McAdams and Rasmusen, 2007; Posner, 1997; and Shavell, 2002). Social norms influence behavior partly through internal incentives: when a person obeys a moral rule, he will tend to feel virtuous, and if he disobeys the rule, he will tend to feel guilty. Social norms also affect behavior through external incentives: when a person is observed by another party to have obeyed a moral rule, that party may bestow praise on the first party, who will enjoy the praise; and if the person is observed by the other party to have disobeyed the rule, the second party will tend to disapprove of the first party, who will dislike the disapproval. Because social norms channel behavior in this way, some socially desirable conduct can be encouraged reasonably well without employing the legal system.

Notwithstanding these observations, there will, of course, often be a need for formal law enforcement. First, much conduct that society desires cannot be controlled through moral incentives alone. One reason is that the private gains from undesirable conduct are often large

and dominate the moral incentives. Another reason is that external moral sanctions might be imposed only with a low probability (the robber, tax cheat, or polluter might not be spotted by others). A second rationale for formal law enforcement is that the social harm from failing to control an act through moral incentives may be large. This makes the expense of law enforcement worth incurring (as in the case of controlling robbery, but not of cutting in line at movie theaters).

FAIRNESS. To this point we have not considered the possibility that individuals have opinions about the fairness of sanctions or the arbitrariness of enforcement (Polinsky and Shavell, 2000b; Kaplow and Shavell, 2002). Suppose, first, that individuals believe that the magnitude of sanctions should reflect the gravity of the acts. As discussed previously, if individuals are risk neutral, the usual solution to the enforcement problem consists of the highest possible sanction and a relatively low probability of detection. When the issue of fairness is added to the analysis, however, the usual solution generally is not optimal because a very high sanction will be seen as unfair.

A consequence of the desire to keep sanctions at fair levels, meaning at quite constrained levels for acts that are not very harmful, is that the socially optimal probability of detection changes. The optimal probability could be higher than the conventionally optimal probability: to achieve a desired level of deterrence with a lower fairness-restricted sanction, the probability has to rise, perhaps significantly. Alternatively, the optimal probability could be lower than in the conventional case: the additional deterrence from raising the probability might be relatively low because the sanction is relatively low; and the lower the deterrent benefit from raising the probability, the lower would be the social incentive to devote resources to enforcement.

Another aspect of fairness concerns the probability of detection rather than the magnitude of sanctions. Suppose that individuals consider it unfair for some lawbreakers to be sanctioned when others, who were lucky enough not to be caught, are not sanctioned. Then the optimal probability would be higher, and therefore the optimal sanction would be lower, than in the absence of this fairness concern.

A further notion of fairness involves the form of liability, whether liability is strict or based on fault. Individuals might prefer fault-based liability because sanctions are imposed on parties only if they behaved in a socially inappropriate way.

A final issue concerns the relevance of fairness considerations when firms, as opposed to individuals, are sanctioned. If what matters in terms of fairness is that the individuals responsible for harmful acts bear sanctions, as opposed to the artificial legal entity of a firm, one would want to identify the sanctions actually suffered by such persons within a firm if the firm bears a sanction. Note, too, that the imposition of sanctions on firms often penalizes individuals who are unlikely to be considered responsible for the harm, namely shareholders and customers.

CRIMINAL LAW. The subject of criminal law may be viewed in the light of the theory of public law enforcement (Posner, 1985; Shavell, 1985). First, the fact that the acts in the core area of crime (robbery, murder, rape, and so forth) are punished by the sanction of imprisonment makes basic sense. Were society to rely on fines alone, deterrence of the acts in question would be grossly inadequate. This is because the probability of detecting many of these acts is low, making the money sanction necessary for deterrence high, but the assets of individuals who commit these acts often are insubstantial. Hence, the threat of prison is needed for deterrence.

Moreover, the incapacitative aspect of imprisonment is valuable because of the difficulty of deterring individuals who are prone to commit criminal acts.

Second, many of the doctrines of criminal law appear to enhance social welfare. This seems true of the basic feature of criminal law that punishment is not imposed on all harmful acts, but instead is usually confined to those that are especially undesirable. (For example, murder is subject to criminal sanctions, but some accidental killing is not.) As we have stressed, when the socially costly sanction of imprisonment is employed, the fault system is desirable because it results in less frequent imposition of punishment than strict liability. Also, the focus on intent in criminal law as a precondition for imposing sanctions may serve to foster deterrence because those who intend to do harm are more likely to conceal their acts, and may be harder to discourage because of the benefits they anticipate. An additional example of a welfare-enhancing doctrine in criminal law concerns attempts. That attempts to do harm are punished is an implicit way of raising the likelihood of sanctions for undesirable acts.

References

Arlen, J.A. 1994. The potentially perverse effects of corporate criminal liability. *Journal of Legal Studies* 23, 833-867.

Bebchuk, L.A., and Kaplow, L. 1992. Optimal sanctions when individuals are imperfectly informed about the probability of apprehension. *Journal of Legal Studies* 21, 365-370.

Becker, G. S. 1968. Crime and punishment: an economic approach. *Journal of Political Economy* 76, 169-217.

Becker, G.S., and Stigler, G.J. 1974. Law enforcement, malfeasance, and compensation of enforcers. *Journal of Legal Studies* 3, 1-18.

Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation,* in: *The Utilitarians*, 1973 ed. Garden City, NY: Anchor Books, 5-398.

Block, M.K., and Sidak, J.G. 1980. The cost of antitrust deterrence: why not hang a price fixer now and then? *Georgetown Law Journal* 68, 1131-1139.

Bowles, R., and Garoupa, N. 1997. Casual police corruption and the economics of crime. *International Review of Law and Economics* 17, 75-87.

Cooter, R.D., and Rubinfeld, D.L. 1989. Economic analysis of legal disputes and their resolution. *Journal of Economic Literature* 27, 1067-1097.

Craswell, R., and Calfee, J. E. 1986. Deterrence and uncertain legal standards. *Journal of Law, Economics, & Organization* 2, 279-303.

Kaplow, L. 1990. Optimal deterrence, uninformed individuals, and acquiring information about whether acts are subject to sanctions. *Journal of Law, Economics, & Organization* 6, 93-128.

Kaplow, L. 1992. The optimal probability and magnitude of fines for acts that definitely are undesirable. *International Review of Law and Economics* 12, 3-11.

Kaplow, L., and Shavell, S. 1994a. Accuracy in the determination of liability. *Journal of Law and Economics* 37, 1-15.

Kaplow, L., and Shavell, S. 1994b. Optimal law enforcement with self-reporting of behavior. *Journal of Political Economy* 102, 583-606.

Kornhauser, L.A. 1982. An economic analysis of the choice between enterprise and personal liability for accidents. *California Law Review* 70, 1345-1392.

Landes, W.M., and Posner, R.A. 1975. The private enforcement of law. *Journal of Legal Studies* 4, 1-46.

Landes, W.M., and Posner, R.A. 1987. *The Economic Structure of Tort Law*. Cambridge, MA: Harvard University Press.

McAdams, R.H., and Rasmusen, E.B. 2007. Norms in law and economics. *Handbook of Law and Economics*, vol. 2, Polinsky, A.M., and Shavell, S. eds. Amsterdam: Elsevier (forthcoming).

Miceli, T. 1996. Plea bargaining and deterrence: an institutional approach. *European Journal of Law and Economics* 3, 249-264.

Mookherjee, D., and Png, I.P.L. 1992. Monitoring vis-a-vis investigation in enforcement of law. *American Economic Review* 82, 556-565.

Mookherjee, D., and Png, I.P.L. 1994. Marginal deterrence in enforcement of law. *Journal of Political Economy* 102, 1039-1066.

Newman, H.A., and Wright, D.W. 1990. Strict liability in a principal-agent model. *International Review of Law and Economics* 10, 219-231.

Png, I.P.L. 1986. Optimal subsidies and damages in the presence of judicial error. *International Review of Law and Economics* 6, 101-105.

Polinsky, A.M. 1980a. Private versus public enforcement of fines. *Journal of Legal Studies* 9, 105-127.

Polinsky, A.M. 1980b. Strict liability vs. negligence in a market setting. *American Economic Review* 70, 363-370.

Polinsky, A.M. 2004. Optimal fines and auditing when wealth is costly to observe. Working Paper No. 289, John M. Olin Program in Law and Economics, Stanford Law School (forthcoming in the *International Review of Law and Economics*).

Polinsky, A.M. 2006. The optimal use of fines and imprisonment when wealth is unobservable. *Journal of Public Economics* 90, 823-835.

Polinsky, A.M., and Rubinfeld, D.L. 1988. The deterrent effects of settlements and trials. *International Review of Law and Economics* 8, 109-116.

Polinsky, A.M., and Rubinfeld, D.L. 1991. A model of optimal fines for repeat offenders. *Journal of Public Economics* 46, 291-306.

Polinsky, A.M., and Shavell, S. 1979. The optimal tradeoff between the probability and magnitude of fines. *American Economic Review* 69, 880-891.

Polinsky, A.M., and Shavell, S. 1984. The optimal use of fines and imprisonment. *Journal of Public Economics* 24, 89-99.

Polinsky, A.M., and Shavell, S. 1992. Enforcement costs and the optimal magnitude and probability of fines. *Journal of Law and Economics* 35, 133-148.

Polinsky, A.M., and Shavell, S. 1993. Should employees be subject to fines and imprisonment given existence of corporate liability? *International Review of Law and Economics* 13, 239-257.

Polinsky, A.M., and Shavell, S. 1998. On offense history and the theory of deterrence. *International Review of Law and Economics* 18, 305-324.

Polinsky, A.M., and Shavell, S. 1999. On the disutility and discounting of imprisonment and the theory of deterrence. *Journal of Legal Studies* 28, 1-16.

Polinsky, A.M., and Shavell, S. 2000a. The economic theory of public enforcement of law. *Journal of Economic Literature* 38, 45-76.

Polinsky, A.M., and Shavell, S. 2000b. The fairness of sanctions: some implications for optimal enforcement policy. *American Law and Economics Review* 2, 223-237.

Polinsky, A.M., and Shavell, S. 2001. Corruption and optimal law enforcement. *Journal of Public Economics* 81, 1-24.

Polinsky, A.M., and Shavell, S. 2007. The theory of public enforcement of law. *Handbook of Law and Economics*, vol. 1, Polinsky, A.M., and Shavell, S. eds. Amsterdam: Elsevier (forthcoming).

Posner, R.A. 1985. An economic theory of the criminal law. *Columbia Law Review* 85, 1193-1231.

Posner, R.A. 1997. Social norms and the law: an economic approach. *American Economic Review: Papers and Proceedings* 87, 365-369.

Reinganum, J.F. 1988. Plea bargaining and prosecutorial discretion. *American Economic Review* 78, 713-728.

Rose-Ackerman, S. 1999. *Corruption and Government: Causes, Consequences and Reform.* New York: Cambridge University Press.

Rubinstein, A. 1979. An optimal conviction policy for offenses that may have been committed by accident. *Applied Game Theory*, Brams, S.J., Schotter, A., and Schwodiauer, G. eds. Wurzburg: Physica-Verlag, 406-413.

Shavell, S. 1980. Strict liability versus negligence. *Journal of Legal Studies* 9, 1-25.

Shavell, S. 1982. On liability and insurance. *Bell Journal of Economics* 13, 120-132.

Shavell, S. 1985. Criminal law and the optimal use of nonmonetary sanctions as a deterrent. *Columbia Law Review* 85, 1232-1262.

Shavell, S. 1987a. The optimal use of nonmonetary sanctions as a deterrent. *American Economic Review* 77, 584-592.

Shavell, S. 1987b. A model of optimal incapacitation. *American Economic Review: Papers and Proceedings* 77, 107-110.

Shavell, S. 1987c. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.

Shavell, S. 1991. Specific versus general enforcement of law. *Journal of Political Economy* 99, 1088-1108.

Shavell, S. 1992. A note on marginal deterrence. *International Review of Law and Economics* 12, 345-355.

Shavell, S. 1993. The optimal structure of law enforcement. *Journal of Law and Economics* 36, 255-287.

Shavell, S. 1997. The fundamental divergence between the private and the social motive to use the legal system. *Journal of Legal Studies* 26, 575-612.

Shavell, S. 2002. Law versus morality as regulators of conduct. *American Law and Economics Review* 4, 227-257.

Shavell, S. 2004. *Foundations of Economic Analysis of Law.* Cambridge, MA: Harvard University Press.

Shleifer, A., and Vishny, R.W. 1993. Corruption. *Quarterly Journal of Economics* 108, 599-617.

Spier, K. 1997. A note on the divergence between the private and the social motive to settle under a negligence rule. *Journal of Legal Studies* 26, 613-621.

Stigler, G. 1970. The optimum enforcement of laws. *Journal of Political Economy* 78, 526-536.

Sykes, A.O. 1981. An efficiency analysis of vicarious liability under the law of agency. *Yale Law Journal* 91, 168-206.

Wilde, L.L. 1992. Criminal choice, nonmonetary sanctions, and marginal deterrence: a normative analysis. *International Review of Law and Economics* 12, 333-344.