

A defendant found not guilty by reason of insanity is automatically committed to an insane asylum. Release is dependent on a judicial hearing in the presence of a public prosecutor and on the presentation of a medical expert's report that, in his or her view, the patient no longer constitutes a danger to self or others.

#### 4.2 Israel

The Israeli Penal Law updated in 1983 reads as follows: 'A person shall not bear the criminal responsibility for an act that he has committed if, by reason of a mental illness or defect, he is incapable of choosing between performing the act and refraining from doing so.'

#### 4.3 Kenya

Kenya's standard also reads much like M'Naghten: 'Where an act or omission is charged against a person as an offence, and it is given in evidence in the trial of that person for that offence that he was insane so as not to be responsible for his acts or omissions at the time when the act was done or the omission made, then if it appears to the court before which the person is tried that he did the act or made the omission charged but was insane at the time he did or made it, the court shall make a special finding to the effect that the accused was guilty of the act or omission charged but was insane when he did the act or made the omission.'

If a defendant is found to be insane, the case is reported to the president; the accused is placed in custody—in either a mental hospital or a prison—and remains under the jurisdiction of the court.

#### 4.4 Japan

The Japanese Criminal Code—amended in 1954—states simply: 'An Act of a person of unsound mind is not punishable. Punishment shall be reduced for acts of weak-minded persons.'

#### 4.5 South Korea

The South Korean Criminal Code—amended in 1960—reads as follows: 'A person is not punishable if, because of mental disorder, he is unable to pass rational judgment or to control his will.'

South Korea also allows for diminished responsibility: 'The punishment of a person who, due to a mental disorder, is deficient in the capacity mentioned in the preceding Section, shall be mitigated.'

*See also:* Crime and Class; Criminal Defenses; Criminal Law and Crime Policy; Deterrence: Legal Perspectives; Expert Testimony; Guilt

#### Bibliography

- Allen F 1964 *The Borderline of Criminal Justice*. University of Chicago Press, Chicago, p. 105  
Simon R J, Aaronson D E 1988 *The Insanity Defense*. Praeger, New York, p. 22  
Stone A 1975 *Mental Health and Law: A System in Transition*. DHEW pub. no. (ADM) 75-176. NIMH, Rockville, MD, p. 218

R. J. Simon

## Law: Economics of its Public Enforcement

In this article we consider the economic theory of the public enforcement of law—the use of public agents (inspectors, tax auditors, police, prosecutors) to detect and to sanction violators of legal rules. Economically-oriented analysis of public law enforcement dates from the eighteenth century contributions of Montesquieu (1748), Beccaria (1767), and, especially, Bentham (1789). Curiously, after Bentham (1789), the subject of enforcement lay essentially dormant in economic scholarship until the late 1960s, when Becker (1968) published a highly influential article that has led to a voluminous literature. In Sects. 1 through 3, we present the basic elements of the theory of public enforcement. Our concern is with the probability of imposition of sanctions, the magnitude and form of sanctions, and the rule of liability. In Sects. 4 through 14 we then examine a variety of extensions of the central theory, including accidental harms, costs of imposing fines, mistake, marginal deterrence, settlement, self-reporting, repeat offenses, and incapacitation. (A more expansive treatment of the subject of this article is contained in Polinsky and Shavell 2000.)

### 1. The Basic Framework

An individual (or a firm) chooses whether to commit an act that for simplicity is assumed to cause harm with certainty. If he commit the act, he obtains some gain, and also faces the risk of being caught, found liable, and sanctioned. The rule of liability could be either strict—under which he is definitely sanctioned; or fault-based—under which he is sanctioned only if his behavior fell below a fault standard. The sanction

that he suffers could be a monetary fine, a prison term, or a combination of the two.

Whether an individual chooses to commit a harmful act is determined by an expected utility calculation. He will commit the act if that would raise his expected utility, taking into account the gain he would derive and the subsequent probability of being caught and sanctioned. We will usually first examine the assumption that individuals are risk neutral with respect to sanctions, that is, that they treat an uncertain sanction as equivalent to its expected value; but we will also consider alternative assumptions.

Social welfare 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 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. If individuals are risk neutral, social welfare can be expressed simply as the gains individuals obtain from committing his acts, less the arms caused, and less the costs of law enforcement.

We assume, as is conventional, that fines are socially costless to employ because they are mere transfers of money, whereas imprisonment involves positive social costs because of the expense associated with the operation of prisons and the disutility due to imprisonment.

The enforcement authority's problem is to maximize social welfare by choosing enforcement expenditures, or, equivalently, a probability of detection, the level of sanctions and their form (a fine, prison term, or combination), and the rule of liability (strict or fault-based).

## 2. Optimal Enforcement Given the Probability of Detection

We consider here optimal enforcement given the assumption that the probability of detection is fixed (the probability will be treated as a policy instrument in the next section). Thus, we ask about the optimal form and level of sanctions under strict and fault-based liability, and about how the two liability rules compare.

### 2.1 Strict Liability

Assume initially that fines are the form of sanction and that individuals are risk neutral. Then the optimal fine is the harm  $h$  divided by the probability of detection  $p$ , that is,  $h/p$ ; for then the expected fine equals the harm (observe that  $p(h/p) = h$ ). If, for example, the harm is \$1,000 and the probability of detection is 25 percent, then the optimal fine is \$4,000, and the expected fine is \$1,000. This fine is optimal because, when the expected fine equals the harm, an individual will commit a

harmful act if, and only if, the gain he would derive from it exceeds the harm he would cause. Essentially this basic formula was noted by Bentham (1789, p. 173) and it has been commented upon by many others since.

If individuals are risk averse with regard to fines, the optimal fine would tend to be lower than in the risk-neutral case for two reasons. First, reducing 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.

Next assume that imprisonment is the form of sanction. In this case, there is not a simple formula for the optimal imprisonment term. The optimal term could be such that there is either underdeterrence or overdeterrence, compared to socially ideal behavior. On one hand, a relatively low imprisonment term, implying underdeterrence, might be socially desirable because it means that imprisonment costs are reduced with respect to those individuals who commit harmful acts. On the other hand, a relatively high term, implying overdeterrence, might be socially desirable because it means that imprisonment costs are reduced due to fewer individuals committing harmful acts.

Now consider the combined use of fines and imprisonment. Here, the main 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 maximal fine might be interpreted as the wealth of an individual.) 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. This point is noted by Bentham (1789, p. 183) and Becker (1968, p. 193); see also Polinsky and Shavell (1984).

### 2.2 Fault-based Liability

Assume again that fines are the form of liability. Then the same formula for the fine that we said was optimal under strict liability—namely,  $h/p$ , the harm divided by the probability of detection—will lead to compliance with the fault standard (assuming that the fault standard is optimally selected).

If individuals are risk averse, they are deterred more easily than if they are risk neutral, so the fine does not need to be as high to induce compliance with the fault standard. Moreover, assuming that compliance occurs, no one actually is sanctioned because no one is found at fault (provided that there are no mistakes). Thus, fault-based liability has the attractive feature that it can accomplish desired deterrence of harm-creating conduct without imposing risk on risk-averse individuals (Shavell 1982).

Next, consider imprisonment as the sanction; see Shavell (1987a). Here, for essentially the reasons given in the case of fines, any sanction above a threshold level will ensure compliance with the fault standard, and the minimum sanction necessary to induce compliance is higher the lower is the probability of detection. Also (Shavell 1985), fault-based liability can accomplish deterrence without the actual imposition of costly imprisonment sanctions.

Finally, consider the joint use of fines and imprisonment. In this case, it does not matter what the combination of sanctions is, provided that the sanctions achieve compliance with the fault standard.

### 2.3 Comparison of Liability Rules

Because sanctions are not imposed under fault-based liability (in the absence of mistakes), this form of liability has an advantage over strict liability when the sanction is a fine and individuals are risk-averse, or when the sanction is imprisonment. However, fault-based liability is more difficult to administer. Namely, to apply fault-based liability, the enforcement authority must have more information than under strict liability: it must be able to calculate optimal behavior to determine the fault standard and it must ascertain whether the fault standard was met. Under strict liability, the authority need only ascertain harm. (Moreover, for reasons we discuss in Sect. 6 below, strict liability encourages better decisions by injurers regarding their level of participation in harm-creating activities.)

## 3. Optimal Enforcement Including the Probability of Detection

We now consider the optimal system of enforcement when expenditures on enforcement, and hence the probability of detection, are allowed to vary. Consideration of this issue originated with Becker (1968).

### 3.1 Strict Liability

Assume first that the sanction is a fine and that individuals are risk neutral. Then the optimal level of the fine is maximal and the optimal probability is low (in a sense to be described). The basic explanation for this conclusion is that 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. Suppose, for example, that the fine initially is \$4,000 and that the probability of detection is 25 percent. Now raise the fine to \$10,000, presuming that the maximal fine is at least this high, and lower the probability of detection

to 10 percent. Then the expected fine remains equal to \$1,000, so that deterrence is maintained, but expenditures on enforcement are reduced significantly, implying that social welfare rises. This process can be continued, and social welfare augmented, whenever the fine is below the maximal level  $f_m$ . Becker (1968) suggested this result; Carr-Hill and Stern (1979, pp. 280–309) and Polinsky and Shavell (1979) note it explicitly.

The optimal probability is low in the sense that there is some underdeterrence; that is, the optimal  $p$  is such that the expected fine  $pf_m$  is less than the harm  $h$ . The reason for this result is that 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. Lowering  $p$  will be socially beneficial because these individuals cause no net social losses (their gains essentially equal the harm), but reducing  $p$  saves enforcement costs.

If individuals are risk averse, the optimal fine generally is less than maximal, as first shown in Polinsky and Shavell (1979) (and elaborated upon in Kaplow 1992). This is because the use of a very high fine would impose a substantial risk-bearing cost on individuals who commit the harmful act.

Next, assume that the sanction is imprisonment and that individuals are risk neutral in imprisonment, that is, the disutility of imprisonment is the same for each additional year. Then the optimal imprisonment term is maximal. 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.

Suppose instead that individuals are risk averse in imprisonment. In other words, the disutility of each year of imprisonment grows with the number of years in prison, perhaps because imprisonment becomes increasingly difficult to tolerate. In this case there is a stronger argument for setting the imprisonment sanction maximally than when individuals are risk neutral (Polinsky and Shavell 1999). 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 social costs of imposing imprisonment sanctions decline because the expected prison term falls.

Last, suppose that individuals are risk preferring in imprisonment, that is, the disutility of each year of imprisonment falls with the number of years in prison. This assumption seems particularly

important: the first years of imprisonment may create special disutility, due to brutalization of the prisoner, or due to the stigma effect of having been imprisoned at all. Additionally, the fact that individuals discount the future disutility of imprisonment makes earlier years of imprisonment more important than later ones. If individuals are risk preferring in imprisonment, the optimal sanction may be less than maximal (Polinsky and Shavell 1999). In particular, the type of argument used above does not necessarily apply. 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 of imposing imprisonment sanctions might exceed the savings in enforcement expenditures from lowering the probability, the optimal prison term might not be maximal.

When the probability of detection is set optimally, together with the imprisonment term, underdeterrence may well result, not only to save enforcement expenditures, but also to reduce the costs of imposing imprisonment sanctions.

### 3.2 Fault-based Liability

The least expensive way to accomplish compliance with the fault standard is to use 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. 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.

### 3.3 Comparison of Liability Rules

As we emphasized earlier, under fault-based liability sanctions are not actually imposed (in the absence of mistakes), which often is an advantage over strict liability. However, this advantage of fault-based liability would have to be weighed against the disadvantages of this rule that we mentioned at the end of Sect. 2.

## 4. Accidental Harms

Until now, we have assumed 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. For instance, if a firm stores toxic chemicals in a substandard tank, the firm only creates the probability of a harmful spill.

Essentially all that we have said above applies in a straightforward way when harms are accidental. If parties are risk-neutral, sanctions are monetary, and the expected sanction equals harm, 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; and so forth.

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. In principle, either approach can achieve optimal deterrence. If liability is based on the dangerous act, the expected fine should equal the expected harm, while if liability is based on actual harm, the expected fine should equal the actual harm.

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 to accomplish a given level of deterrence (because expected harm is less than actual harm), and thus are more likely to be able to be paid. Second, because act-based sanctions can be lower, they tend to be preferable when parties are risk averse. Third, act-based and harm-based sanctions may differ in the ease with which they can be applied. 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); or harm-based sanctions may be easier to impose (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, thereby favoring harm-based liability.

## 5. Costs of Imposing Fines

Now suppose that there are costs borne by enforcement authorities in imposing fines. Our principal observation is that such costs should raise the level of the fine. To see why, suppose for simplicity that the probability of detection is fixed, that liability is strict, and that individuals are risk neutral. In this setting, recall from Sect. 2 that the optimal fine is  $h/p$ , the harm divided by the probability of detection. Now let there be a public cost  $k$  of imposing a fine. The optimal fine then becomes  $h/p+k$ ; in other words, the cost  $k$  should be added to the fine that would otherwise be desirable (Becker 1968, p. 192, Polinsky and Shavell 1992). The intuition behind this result 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 the act. Hence, he will behave in a socially appropriate way.

### 6. Level of Activity

We have been assuming that the sole decision that an individual makes is whether to act in a way that causes harm when engaging in some activity. In many contexts, however, an individual also makes a choice about his activity level—that is, whether to engage in that activity, or, more generally, at what level to do so. For example, besides deciding how to behave when driving (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.

The socially optimal activity level is such that the individual's marginal utility from engaging in 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.

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 generally 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. Under fault-based liability, however, parties generally do not pay for the harm they cause because, as we have discussed, they will tend to behave so as not to be found at fault. Consequently, when deciding on their level of activity, they will not take into account the harm that their participation in the activity causes, and therefore they will participate 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. 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.

The tendency of parties to choose an excessive level of activity under fault-based liability, but not under strict liability, was first emphasized in Shavell (1980) and Polinsky (1980).

### 7. Mistakes

An individual who should be found liable might mistakenly not be found liable—a Type I error.

Alternatively, 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 the probabilities of these errors be  $\varepsilon_1$  and  $\varepsilon_2$ , respectively.

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  $g$  net of his expected fine if they do commit it exceeds his expected fine if they do not commit it, namely, when  $g - p(1 - \varepsilon_1)f > -p\varepsilon_2f$ , or, equivalently, when  $g > (1 - \varepsilon_1 - \varepsilon_2)pf$ .

The first point to note is that, as emphasized in Png (1986), both types of error reduce deterrence: the term  $(1 - \varepsilon_1 - \varepsilon_2)pf$  is declining in both  $\varepsilon_1$  and  $\varepsilon_2$ . 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 lowers deterrence because it reduces the difference between the expected fine from violating the law and not violating it—in effect, making a violation less costly to the individual.

Because mistakes dilute deterrence, they tend to reduce social welfare. Specifically, to achieve any level of deterrence, the probability  $p$  must be higher to offset the effect of errors.

If individuals are risk averse, 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 (Block and Sidak 1980).

When liability is based on fault, an important implication of mistake is that some individuals will bear sanctions even if they comply with the fault standard, tending to make fault-based liability operate like strict liability. Moreover, as stressed by Craswell and Calfee (1986), individuals will often have a motive to take excessive precautions in order to reduce the chance of erroneously being found at fault.

### 8. 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. To investigate such situations, suppose that a single probability of detection applies uniformly to all harmful acts, regardless of the magnitude of the harm.

When enforcement is general in this sense, the optimal sanction rises with the severity of the harm and is maximal only for relatively high harms; this point was first made in Shavell (1991) (Mookherjee and Png 1992 is closely related). Suppose, for example, 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 given general probability of detection  $p$ , the optimal fine schedule is  $h/p$ , provided that  $h/p$  is feasible; otherwise the optimal fine is maximal. This schedule is optimal given  $p$  because it implies that the expected fine equals harm, thereby inducing ideal behavior, whenever that is possible. 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.

### 9. *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. In such contexts, the threat of sanctions influences which harmful acts individuals choose to commit. Deterrence of a more harmful act because its sanction exceeds that for a less harmful act is sometimes referred to as marginal deterrence (apparently so named by Stigler 1970).

Other things being equal, as observed by Beccaria (1767, p. 32) and Bentham (1789, p. 171), it is socially desirable that enforcement policy creates marginal deterrence, so that law violators have a reason to moderate the amount of harm they cause. This suggests that sanctions should rise with the magnitude of harm and, therefore, that sanctions generally should not be maximal. Note that marginal deterrence also can be promoted by increasing the probability of detection for more severe harms. For formal analyses of marginal deterrence, see Shavell (1992), Wilde (1992), and Mookherjee and Png (1994).

### 10. *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.

When harm is caused by the agent of a principal, many of our prior conclusions are not fundamentally altered; they simply carry over to the sanctioning of principals. Notably, if a risk-neutral principal faces an expected fine equal to harm done, he will in effect be in the same position vis-à-vis his agent as society is vis-à-vis a single violator of law (see Newman and Wright 1990 on a closely related point). Consequently, the principal will behave socially optimally in controlling his agent, and in particular will contract with him and monitor him in ways that will give the agent socially appropriate incentives to reduce harm (but see Arlen 1994).

The allocation of a financial sanction between the principal and the agent would not matter if, as would

be the natural presumption, the parties can reallocate the sanction through their own contract. The allocation of the sanction would matter, however, if it would allow the parties to reduce their total burden, for example, if the agent is unable to pay a fine because his assets are less than the fine; see Sykes (1981) and Kornhauser (1982). Then, the fine should be imposed on the principal.

The imposition of imprisonment sanctions on agents may be desirable when their assets are less than the harm that they can cause; see Polinsky and Shavell (1993). Because an agent's assets are limited, the principal may be unable to control him adequately through use of contractually determined penalties, which can only be monetary. It may be socially valuable to use the threat of personal criminal liability and a jail sentence to remedy this problem.

### 11. *Settlements*

We consider here how settlements 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 (see generally Cooter and Rubinfeld 1989, and regarding settlement in a criminal context, see, for example, Reinganum 1988 and Miceli 1996). First, a trial is costly in terms of time and/or money. Second, settlements eliminate the risks inherent in the trial outcome. For these reasons, settlement tends to be socially valuable.

But a complicating factor is that settlements dilute deterrence: for if injurers desire to settle, it must be because the expected disutility of sanctions is lowered for them (see generally Polinsky and Rubinfeld 1988). The state may be able to offset this effect by increasing the level of sanctions; if so, settlements are socially desirable for the reasons mentioned in the previous paragraph.

### 12. *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 to enforcement authorities. For example, firms often report violations of environmental and safety regulations. There are two basic reasons why self-reporting is socially advantageous; see Kaplow and Shavell (1994b). 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. For example, drivers bear less risk because they know that if they cause an accident, they can 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. Self-reporting can be induced by lowering the sanction for individuals who disclose their own infractions. Although this will tend to reduce deterrence, the reward for self-reporting can be made small enough that deterrence is only negligibly reduced.

### 13. Repeat Offenders

In practice, the law often sanctions repeat offenders more severely than first-time offenders. This may be beneficial for two reasons. First, as developed in Polinsky and Shavell (1998), raising the sanction for repeat offenders may create additional deterrence: if getting caught violating the law implies not only an immediate sanction, but also a higher sanction for any future violation, an individual will be more deterred from committing a violation presently. Second, as studied, for example, in Rubinstein (1979) and Polinsky and Rubinfeld (1991), making sanctions depends on offense history and allows society to take advantage of information about the propensity of individuals to commit offenses: 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. It also may be desirable to incapacitate such individuals by imprisoning them (see the next section).

### 14. Incapacitation

We have focused on the use of sanctions to reduce harm by deterring individuals from causing harm. However, an entirely different way to reduce harm is by incapacitating individuals so that they cannot cause harm. Imprisonment is the primary incapacitative sanction; on the economic theory of incapacitation, see Shavell (1987b).

If the sole function of sanctions were to incapacitate, the optimal sanction would be determined by comparing the expected harm, net of gains, an offender would cause if not in prison to the private and public costs of imprisonment. Jail should only be used to incapacitate individuals whose net harm is relatively high.

*See also:* Crime: Sociological Aspects; Crime, Sociology of; Deterrence; Law and Economics; Law, Sociology of; Norms

### Bibliography

Arlen J A 1994 The potentially perverse effects of corporate criminal liability. *Journal of Legal Studies* **23**: 833–67  
Beccaria C 1995 [1767] *On Crimes and Punishments, and Other Writings*. In: Bellamy R (ed.) [trans. Davies R et al.] Cambridge University Press, New York

Becker G S 1968 Crime and punishment: An economic approach. *Journal of Political Economy* **76**: 169–217  
Bentham J 1789/1973 An introduction to the principles of morals and legislation. In: Bentham J (ed.) *The Utilitarians*. Anchor Books, Garden City, New York  
Block M K, Sidak J G 1980 The cost of antitrust deterrence: Why not hang a price fixer now and then? *Georgetown Law Journal* **68**: 1131–39  
Carr-Hill R A, Stern N H 1979 *Crime, the Police and Criminal Statistics: An Analysis of Official Statistics for England and Wales Using Econometric Methods*. Academic Press, London  
Cooter R D, Rubinfeld D L 1989 Economic analysis of legal disputes and their resolution. *Journal of Economic Literature* **27**: 1067–97  
Craswell R, Calfee J E 1986 Deterrence and uncertain legal standards. *Journal of Law, Economics, and Organization* **2**: 279–303  
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, Shavell S 1994 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**: 1345–92  
Miceli T J 1996 Plea bargaining and deterrence: An institutional approach. *European Journal of Law and Economics* **3**: 249–64  
Montesquieu C 1748/1977 *The Spirit of the Laws*. University of California Press, Berkeley, CA  
Mookherjee D, Png I P L 1992 Monitoring vis-à-vis investigation in enforcement of law. *American Economic Review* **82**: 556–65  
Mookherjee D, Png I P L 1994 Marginal deterrence in enforcement of law. *Journal of Political Economy* **102**: 1039–66  
Newman H A, Wright D W 1990 Strict liability in a principal-agent model. *International Review of Law and Economics* **10**: 219–31  
Png I P L 1986 Optimal subsidies and damages in the presence of judicial error. *International Review of Law and Economics* **6**: 101–5  
Polinsky A M 1980 Strict liability vs. negligence in a market setting. *American Economic Review* **70**: 363–7  
Polinsky A M, Rubinfeld D L 1988 The deterrent effects of settlements and trials. *International Review of Law and Economics* **8**: 109–16  
Polinsky A M, Rubinfeld D L 1991 A model of optimal fines for repeat offenders. *Journal of Public Economics* **46**: 291–306  
Polinsky A M, Shavell S 1979 The optimal tradeoff between the probability and magnitude of fines. *American Economic Review* **69**: 880–91  
Polinsky A M, Shavell S 1984 The optimal use of fines and imprisonment. *Journal of Public Economics* **24**: 89–99  
Polinsky A M, Shavell S 1992 Enforcement costs and the optimal magnitude and probability of fines. *Journal of Law and Economics* **35**: 133–48  
Polinsky A M, Shavell S 1993 Should employees be subject to fines and imprisonment given the existence of corporate liability? *International Review of Law and Economics* **13**: 239–57  
Polinsky A M, Shavell S 1998 On offense history and the theory of deterrence. *International Review of Law and Economics* **18**: 305–24

- Polinsky A M, 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, Shavell S 2000 The economic theory of public enforcement of law. *Journal of Economic Literature* **38**: 45–76
- Reinganum J F 1988 Plea bargaining and prosecutorial discretion. *American Economic Review* **78**: 713–28
- Rubinstein A 1979 An optimal conviction policy for offenses that may have been committed by accident. In: Brams S J, Schotter A, Schwödiauer G (eds.) *Applied Game Theory*. Physica-Verlag, Würzburg, Germany, pp. 406–13
- 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**: 120–32
- Shavell S 1985 Criminal law and the optimal use of nonmonetary sanctions as a deterrent. *Columbia Law Review* **85**: 1232–62
- Shavell S 1987a The optimal use of nonmonetary sanctions as a deterrent. *American Economic Review* **77**: 584–92
- Shavell S 1987b A model of optimal incapacitation. *American Economic Review* **77**: 107–10
- 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**: 345–55
- Shavell S 1993 The optimal structure of law enforcement. *Journal of Law and Economics* **36**: 255–87
- Stigler G J 1970 The optimum enforcement of laws. *Journal of Political Economy* **78**: 526–36
- 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**: 333–44

A. M. Polinsky and S. Shavell

Copyright © 2001 Elsevier Science Ltd.

All rights reserved.

## Law Firms

The overwhelming majority of lawyers in private practice in most countries around the world practice alone or in very small and relatively informal groupings (Abel and Lewis 1988–95). Beginning in the 1960s, however, there has been a steady trend for lawyers who represent business interests to practice in larger and more bureaucratically organized law firms. Although lawyers for individuals have also attempted to organize larger firms, these efforts have been largely unsuccessful (Van Hoy 1997). As a result, at the end of the twentieth century the large corporate law firm is the dominant organizational model for private practitioners.

### 1. The Evolution of the Large Law Firm: From Cravathism to Megalawyering

The American law firm of Cravath, Swaine, and Moore is widely credited with creating the blueprint for the modern large law firm in the first decade of the

twentieth century. The ‘Cravath System’ consists of four interrelated practices. First, lawyers, typically called ‘associates,’ are hired directly out of law school, paid a fixed salary, and required to work exclusively for the firm. Second, these new recruits are hired for a probationary period during which they are trained by the senior lawyers in the firm, called ‘partners,’ to handle matters of increasing responsibility for the firm’s clients. Third, at the end of the probationary period, the partners select only the best associates for promotion to partnership, requiring those who are not selected to leave the firm. Finally, partners share in both the firms’ profits and management, including, most importantly, the selection of new partners (Swaine 1946).

Firms patterned on the Cravath model grew dramatically in size and geographic scope from 1960 to the end of the twentieth century. In 1960, there were 38 US law firms with more than 50 lawyers, with the largest consisting of 125 attorneys (Smigel 1969). Forty years later, there were 250 law firms in the USA with more than 100 lawyers, with several exceeding 1,000 attorneys (National Law Journal 1999). In 1960, virtually all firms consisted of a single office. By 2000, most large firms had multiple offices, with many practicing in several jurisdictions. In 1960, only a handful of US firms had foreign offices. Forty years later, foreign offices of the top 250 American law firms were located in 72 foreign cities and employed almost 5,000 US and foreign lawyers.

During this same period, ‘megafirms’ (Galanter and Palay 1991) patterned on the Cravath model began to emerge in the UK, Europe, Asia, and other commercial centers. England was the first to develop comparable firms to those found in the USA (Flood 1989). By the end of the century, these firms were among the largest and most globalized firms in the world. Lawyers in the Netherlands (Blankenburg and Bruinsma 1994), Germany (Gerber 1999), Spain (Stewart 1991), Canada (Arthurs et al. 1988), Australia (Galanter and Palay 1991), China (Alford 1995), and Venezuela (Perez Perdomo 1988) have also developed their own versions of the large law firm. Although the structure and culture of these firms reflect important national and regional differences (Trubek et al. 1994), the Cravath System continues to exert a powerful influence on the development of large law firms around the world. As a result, theories about how US law firms are structured are likely to have important implications for firms in other countries that are patterned on the American model.

### 2. Explaining Law Firm Growth

Theorists offer three related explanations for law firm growth: tournament theory, portfolio theory, and demand theory. Although each theory highlights important truths, none captures the dynamic interaction between markets, institutions, and lawyer