

- Crain, W.M. and Ekelund, R.B. 1976. Chadwick and Demsetz on competition and regulation. *Journal of Law and Economics* 19: 149–62.
- Demsetz, H. 1968. Why regulate utilities? *Journal of Law and Economics* 11: 55–66.
- Dnes, A.W. 1994. The scope of Chadwick's bidding scheme. *Journal of Institutional and Theoretical Economics* 150: 524–36.
- Galal, A., Jones, L., Tandon, P. and Vogelsant, I. 1994. *Welfare Consequences of Selling Public Enterprise*. Washington DC: World Bank.
- McAfee, P. and McMillan, J. 1987. Auctions and bidding. *Journal of Economic Literature* 25: 699–738.
- Peacock, A.T. and Rowley, C.K. 1972. Welfare economics and the public regulation of natural monopolies. *Journal of Public Economics* 1: 227–44.
- Priest, G.L. 1993. The origins of utility regulation and the 'theories of regulation' debate. *Journal of Law and Economics* 36: 289–323.
- Schmalensee, R. 1979. *The Control of Natural Monopolies*. Lexington, IL: D.C. Heath & Co.
- Telser, L.G. 1969. On the regulation of industry: a note. *Journal of Political Economy* 77: 937–52.
- Vickrey, W. 1961 Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance* 16: 8–37.
- Williamson, O.E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Zupan, M.A. 1989. The efficacy of franchise bidding schemes in the case of cable television: some systematic evidence. *Journal of Law and Economics* 32: 401–56.

public goods. See FISCAL FEDERALISM; PRISONERS' DILEMMA AND THE THEORY OF THE STATE.

publicity. See PRIVACY.

punitive damages. This essay concerns *punitive damages*, an important form of damages that are sometimes awarded to plaintiffs in addition to compensatory damages. (The term 'punitive damages' is somewhat inapt because the purpose of such damages is only partly, and perhaps not even mainly, to punish; we nevertheless use the term because it is conventional.) In the United States, punitive damages are awarded in approximately six percent of all cases in which plaintiffs prevail. While punitive damages are granted mainly in tort cases, they are increasingly employed in contract disputes and other areas of litigation; see generally Eisenberg et al. (1997) for an evaluation of the empirical significance of punitive damages. Outside the United States, punitive damages and other forms of extra-compensatory damages are of lesser importance; see Stoll (1983: 99–106). Much of what we have to say, however, is relevant to publicly imposed penalties that exceed harm, which are common in all countries.

In considering the justification for awarding punitive damages, we will refer to two broad social goals: deterrence and punishment. By deterrence we mean the use of sanctions to influence behaviour, so as to maximize the following measure of social welfare: the benefits parties obtain from their actions, less the costs of precautions, the harm done, and the expenses due to use of the legal system. By punishment, we mean the imposition of sanc-

tions to satisfy a desire for retribution against wrongdoers. We generally do not consider compensation and the reduction of risk as social goals; this simplification is not of great consequence because punitive damages are extra-compensatory; thus, there is no need to insure victims (although the bearing of liability risk by injurers remains an issue, as will be noted).

The remainder of this essay is organized as follows. In section 1, we review the basic theory of deterrence, and in sections 2 through 5, we discuss the main deterrence-related justifications for punitive damages: the possibility of escaping sanctions; underestimation of harm; socially illicit gains; and inducing parties to bargain rather than acting unilaterally to cause harm. Then in section 6 we examine the punishment goal and how it is served by punitive damages, and in section 7 we consider how the punishment and deterrence goals should jointly determine damages. Finally, in section 8 we address a variety of extensions to the analysis and certain legal doctrines that bear on the award of punitive damages.

1. OPTIMAL DAMAGES WHEN INJURERS ARE FOUND LIABLE FOR SURE: THE BASIC THEORY OF DETERRENCE. We summarize here the basic principles of the economic theory of deterrence and liability assuming that, whenever a party causes harm, he will be sanctioned for sure. (For integrated treatments of the standard theory of liability and deterrence, see Landes and Posner (1987) and Shavell (1987).) In this setting, the point on which we want to focus is that *the proper magnitude of damages is the harm that the party has caused*. (The term 'damages' means the magnitude of liability payments.) We first discuss this point when liability is strict – when injurers are liable for harm regardless of the care they took – and then when liability is based on the negligence rule – when injurers are liable for harm only if they were at fault. Readers familiar with the basic theory of deterrence and liability may want to proceed directly to section 2.

There are two basic reasons why it is best for damages to equal harm under strict liability. The first concerns the *level of precautions* taken by parties, where the term 'precautions' is to be interpreted generally (including, for example, the use of safety devices, attention to hazards, and the monitoring of employees by firms). If damages equal harm, parties will have socially correct incentives to take precautions; they will be induced to invest in precautions if and only if the cost is less than the resulting reduction in expected harm. If, however, damages are less than harm, precautions will tend to be inadequate, and if damages exceed harm, precautions will tend to be excessive.

The second reason why it is desirable for damages to equal harm involves parties' *level of activity* – the extent to which individuals and firms participate in risky activities. A party's level of activity affects the magnitude of expected harm, whatever precautions are taken by the party when engaging in the activity. For example, the more miles a person drives (his level of activity), the greater the number of accidents that he is likely to cause, whatever is his level of care when he drives. Similarly, the more units of a product produced and sold by a firm (its level of activity),

the greater the number of accidents that will be caused by the product, whatever are the safety features of the product (which affect the expected harm per unit sold).

If damages equal harm, parties will have socially correct incentives to engage in risky activities. In particular, because an individual's expected damages will equal the expected harm he causes by participating in an activity (such as driving), he will participate in the activity if and only if the benefit he obtains from the activity exceeds the resulting expected harm. Likewise, a firm will produce a product if and only if its value, as reflected in the willingness of customers to pay for it, exceeds the full cost of its production, including the expected harm that it causes. (This is because the price of the product will equal its full cost of production, assuming for simplicity that the firm produces in a competitive environment.) However, if damages are less than harm, levels of activity will tend to be socially excessive, and if damages exceed harm, levels of activity will tend to be too low.

Let us turn now to the negligence rule, under which a party whose level of precautions is below a specified standard is said to be negligent and must pay damages. Assume that the negligence standard is set equal to the optimal level of precautions (the level that minimizes the sum of precaution costs and expected harm). Then, if damages for negligence equal harm, parties will decide to comply with the negligence standard and thus will take appropriate precautions. However, if damages are less than harm, parties might not meet the standard. If damages exceed harm, parties will have a more-than-adequate incentive to meet the standard, and no reason to exceed it, assuming that the negligence determination is accurate.

Realistically, however, there will be errors in the negligence determination. For example, courts may err in determining the negligence standard or in assessing parties' behaviour. Because of the risk of such mistakes, parties may have an incentive to take greater precautions than they would otherwise, in order to reduce the chance that they will incorrectly be found negligent. If the chance of mistake leads parties to take excessive precautions, raising the level of damages will exacerbate this problem.

Next consider the relationship between damages and the level of activity under the negligence rule. In the absence of mistakes, the negligence rule will tend to cause parties to participate in risky activities to a socially excessive extent. This is because, once a party takes the precautions required by the negligence standard, he will not be found liable for any harms that he causes. For example, a person who drives with reasonable care will not be found negligent, and therefore will not have to pay for any harm caused by his driving; consequently, he will drive more than is socially desirable. However, because non-negligent parties will sometimes be found liable by mistake, they will sometimes bear damages. In principle, this could ameliorate the problem of excessive participation in risky activities under the negligence rule. It is also possible, however, that finding parties negligent by mistake will result in their bearing damages in excess of the harm they have caused, and thereby overly discourage their participation in activities. This effect, if it occurs, will be exacerbated by raising the level of damages.

The preceding discussion shows that there is not a simple, theoretically correct answer to the question of what level of damages is optimal under the negligence rule. We will assume for simplicity that optimal damages under the negligence rule are equal to the harm, as under the rule of strict liability. Accordingly, we generally will not distinguish between the two rules in our subsequent discussion.

In passing, we want to note that the conclusion that damages should equal harm depends on our implicit assumption that parties are risk neutral. If injurers are risk averse and cannot purchase liability insurance, the optimal level of damages tends to be lower than harm, both to reduce the imposition of risk on injurers and because damages do not need to be as high to induce injurers to behave appropriately. But if, as is realistic, liability insurance is available (even if only partially available due to moral hazard), the optimal level of damages remains equal to the harm. Also, publicly held firms should be treated as approximately risk neutral – implying that damages should equal harm – if their shareholders have well-diversified portfolios, which often, if not usually, will be the case.

We next turn to various deterrence-based rationales for setting damages in excess of harm – that is, for imposing punitive damages. The first and most important of these arises when injurers might escape liability.

2. OPTIMAL DAMAGES WHEN INJURERS MIGHT ESCAPE LIABILITY. There are several reasons why injurers sometimes escape liability for harms for which they should be liable under a liability rule. First, it may be difficult for the victim to determine that the harm was the result of some party's act – as opposed to simply being the result of nature, of bad luck. This might be the case, for instance, if an individual develops a form of cancer that could have been caused by exposure to a naturally occurring carcinogen but which was in fact caused by exposure to a man-made carcinogen. Second, even if the victim knows that he was injured by a person's conduct and not by nature, it might be difficult for him to prove who caused the harm. The owner of a parked car that was damaged might know that it had been struck by another vehicle but not be able to identify the injurer or be able to establish his identity in court. Third, even if the victim knows both that he was wrongfully injured and who injured him, he might not sue the injurer because of the effort and expense a suit would entail.

The consequences of the possibility that injurers can escape liability are clear. If damages merely equal harm, injurers' motivations to take precautions will be inadequate and their incentive to participate in risky activities will be excessive. To remedy these problems, the damages that are imposed in those instances when injurers are found liable should be raised sufficiently so that injurers' expected damages will equal the harm they cause. This implies that total damages should equal the harm multiplied by the reciprocal of the probability that the injurer will be found liable when he ought to be. Formally, if h is harm and p is the probability of being found liable, the injurer should pay $h \times (1/p) = h/p$ when he is found liable; his expected damages therefore will be $p \times (h/p) = h$. We will refer to $1/p$ as the *total damages multiplier*. For example, if the

probability of being found liable is .25, the total damages multiplier is 4 ($= 1/.25$), so the injurer should pay, in total, four times the level of harm if he is found liable.

The excess of total damages over compensatory damages can be labelled *punitive damages*. Thus, the *optimal* level of punitive damages is the optimal level of total damages less compensatory damages. If the harm is \$100,000 and the probability of being found liable is .25, implying a total damages multiplier of 4, total damages should be \$400,000; since \$100,000 of this total represents compensatory damages, the \$300,000 remainder is the optimal punitive damages amount. This amount also can be described as a multiple of harm or, equivalently, of compensatory damages. Since optimal total damages are h/p , optimal punitive damages are $h/p - h$, which can be rewritten as $[(1 - p)/p]h$. The term in brackets – the ratio of the injurer's chance of escaping liability to the injurer's chance of being found liable – is the *punitive damages multiplier*. In the preceding example, the punitive damages multiplier is 3 ($= .75/.25$), which, when multiplied by the harm of \$100,000, yields the \$300,000 punitive damages amount.

Note that the award of punitive damages may itself raise the probability of suit, and therefore the probability that an injurer will be found liable. This effect, when applicable, should be taken into account. In general, there will be a level of damages that, given the resulting probability of suit, will lead to optimal deterrence. Basing punitive damages on the relatively low probability of suit that would occur if just compensatory damages were awarded would tend to lead to excessive damages.

The general point that, to achieve proper deterrence, sanctions must be inflated if injurers can escape liability, dates back at least to Bentham ([1838–43] 1962: 401–2) and has been applied to the subject of punitive damages by many commentators. The first explicit references to the factor of escaping liability as a justification for punitive damages apparently are Posner (1973: 77–8) and Ellis (1982: 25–6); this justification has been developed most thoroughly by Cooter (1989) and Polinsky and Shavell (1998).

3. OPTIMAL DAMAGES WHEN HARM IS UNDERESTIMATED. Even if injurers are always found liable when they are responsible for harm, if the magnitude of harm is underestimated, compensatory damages will be less than harm and deterrence will be inadequate. This possibility is realistic because hard-to-measure components of harm (such as nonpecuniary losses) are often excluded from damages. Such missing components of harm are commonly mentioned as a reason to impose punitive damages; see especially Ellis (1982: 26–31) and Galligan (1990).

However, as emphasized in Polinsky and Shavell (1998: 84–6), there is a problem with employing punitive damages as a substitute for missing components of compensatory damages. Namely, a component of harm might be excluded from compensatory damages because of the difficulties and expense that would be encountered in its estimation. For example, were the pain and suffering experienced by the friends of a person who dies included in compensatory awards, the number of claimants in cases of wrongful death

could become quite large, and the cost of litigation would also increase as parties contested the degree of their psychological losses. It may well be best, then, for the law to exclude from compensatory damages many such speculative, difficult-to-determine elements of harm, even though these elements are real and their omission does undesirably dilute deterrence. If a component of loss is excluded from compensatory damages for such reasons, arguably it should be excluded from punitive damages for the same reasons.

Conversely, if a component of loss should have been included in compensatory damages, despite the costs of doing so, this mistake in legal policy should be rectified by incorporating the component in such damages. Including the component only in punitive damages would still result in underdeterrence, for the component would remain omitted in the large majority of cases in which only compensatory damages are awarded. Moreover, the component of loss would probably be more poorly measured as a form of punitive damages because the calculation of such damages is not disciplined by the procedures and evidentiary requirements common to the determination of compensatory damages.

4. OPTIMAL DAMAGES WHEN INJURERS' GAINS ARE SOCIALLY ILLICIT. We have implicitly assumed to this point that the gains that parties obtain from committing harmful acts count in social welfare, whereas here we consider the situation when their gains are not counted in social welfare because they are treated as *socially illicit*. Suppose that a person, out of spite, punches another individual. Society might well deem the pleasure the injurer obtains from this act to be socially illicit. This view of an injurer's gains would seem especially plausible when the injurer's utility derives solely from causing harm (that is, when the injurer's act is malicious). However, certain conduct that is not intended to cause harm might also be treated as socially illicit, for instance, driving at high speed for the fun of it.

If an injurer's utility from an act is considered socially illicit (whatever the explanation for this), it is desirable for the act to be deterred completely. To accomplish this, damages must exceed the injurer's utility from committing the act. And since the injurer's utility could be greater than the harm, the required level of damages might exceed the harm. For instance, if the illicit gain from an act is equivalent to \$500 to the injurer and the harm is \$100, damages of at least \$500 are necessary to deter the act. This justification for punitive damages was first noted by Ellis (1982: 31–3) and Cooter (1982: 86–9); for more formal treatments, see Shavell (1987: 159–61) and Diamond (1997a: 8–12).

It should be noted, though, that the present justification for punitive damages is limited in scope. Many, if not most, socially undesirable acts committed by individuals, including some very reprehensible ones, do not seem to be associated with socially illicit utility; often this is because such acts are not committed with the intention of causing harm. Similarly, most conduct of firms is unlikely to be associated with socially illicit utility, since the goal of firms is to make profit, not cause harm.

5. OPTIMAL DAMAGES WHEN, PARTIES CAN BARGAIN AND TRANSACT IN THE MARKETPLACE. In some circumstances it is possible for a party to communicate with a potential victim before causing harm. This would usually be so, for example, when a firm contemplates infringing on another's copyright. When prior communication is possible, a potential injurer could negotiate in advance with the potential victim to purchase the right to engage in the harm-creating conduct. The firm deliberating about the copyright violation could secure a licence to use the copyrighted material.

In such circumstances, it may be socially desirable to induce a potential injurer to bargain and purchase the right to engage in harm-creating conduct – by threatening to impose punitive damages if the injurer acts unilaterally to cause harm. This point apparently originated with Calabresi and Melamed (1972) and was further developed by Landes and Posner (1981), Haddock, McChesney and Spiegel (1990), Biggar (1995) and Kaplow and Shavell (1996).

To amplify on this rationale for punitive damages, suppose that compensatory damages alone are employed and that they are underestimated. A potential injurer then might cause harm when doing so is socially undesirable – because the benefit to the injurer might be less than the harm done, but greater than the low estimate of compensatory damages.

There may be additional undesirable repercussions from underestimating compensatory damages. If injurers can take property from victims without having to pay its full value, injurers will devote effort to identifying and taking such property (copyright violators will seek out material to copy), and victims will expend effort to protect their property (copyright owners will invest resources in preventing duplication of their material). Such efforts are socially wasteful; they are similar to those associated with the theft of property.

The foregoing problems can be avoided if punitive damages are imposed for unilaterally causing harm. If the level of such damages is set so that total damages substantially exceed the value of the property at issue, a potential injurer will be induced to bargain with the property owner – it will be cheaper to pay an agreed upon price than to pay damages. Consequently, property will be exchanged only if the injurer's benefit exceeds the property owner's loss, and the wasteful incentives to take and to protect property will be eliminated.

Another possible reason to employ punitive damages to encourage bargaining and market transactions concerns administrative costs. If compensatory damages are used alone, harm and the taking of property will tend to be mediated through the legal system by the bringing of lawsuits. But if punitive damages are used as a threat, harm and the shifting of property interests will be much more likely to occur through voluntary transactions, the costs of which are likely to be lower than those associated with litigation.

The preceding arguments favouring the use of punitive damages to promote negotiation and market transactions obviously do not apply if bargaining between parties is not possible or if there are substantial impediments to it. Suppose, for instance, that a hiker lost in the mountains discovers an unoccupied cabin. The benefit he would obtain from using the cabin and consuming the food in it

presumably would exceed the loss borne by the cabin's owner. But because there is no opportunity for the hiker to bargain with the owner, the threat of punitive damages might discourage the hiker from using the cabin, which would be undesirable. Hence, when parties cannot bargain, it may be better just to employ compensatory damages (despite the possibility of errors in estimation). Additionally, even if bargaining is feasible, there may be other impediments to efficient exchange – such as bargaining failures due to strategic behaviour – that could justify relying solely on compensatory damages.

Because many harms cannot, as a practical matter, be resolved beforehand by bargaining – including most harms due to accidents between strangers, such as automobile accidents – and because bargaining failures are important, the present justification for punitive damages often will not be relevant.

6. OPTIMAL DAMAGES AND PUNISHMENT. Having discussed the use of punitive damages to accomplish proper deterrence, let us now turn to the punishment objective. We treat this objective as deriving from the desire of individuals to have blameworthy parties appropriately punished. We equate blameworthiness with the reprehensibility of a party's conduct, that is, with its maliciousness or the extent to which it reflects disregard for the safety of others. Given the degree of a party's blameworthiness, we assume that there is a correct level of punishment, and that either higher or lower punishment detracts from satisfaction of the punishment objective.

When the defendant is an individual, the connection between imposition of punitive damages and accomplishment of the punishment objective is conceptually straightforward: if, after assessing the blameworthiness of an individual's act, appropriate punitive damages are imposed, the punishment objective is achieved.

However, when the defendant is a firm, the role of punitive damages in relation to the punishment objective involves a number of complexities; these have been considered in Polinsky and Shavell (1998: 95–101). One is that there are different ways of viewing the objective of punishment: the goal may be to punish firms as *entities*, that is, without reference to whether anyone within a firm behaved inappropriately or was punished as a consequence; or the goal may be to punish firms only as a means of punishing culpable *individuals* in the firms. We find the former conception of the punishment goal unappealing both because it requires a definition of blameworthiness of a firm that is divorced from the behaviour of any individuals who are affiliated with it, and because it necessitates believing that people would, after reflecting about the matter, want to impose a penalty on what ultimately is an artificial legal construct. Notwithstanding these reservations, it is possible that people do want to personify firms and punish them as entities, and the reader can make up his or her mind about the importance of this way of defining the punishment objective.

Now consider the alternative reason for punishing firms – as a means of punishing blameworthy individuals within them. Supposing that this is the purpose of punishment, we turn to the question of the extent to which the

imposition of punitive damages on firms will in fact result in the punishment of blameworthy employees. Because firms clearly have an interest in discouraging culpable conduct by their employees that could give rise to punitive damages, firms can be expected to seek to control such conduct through the use of internal sanctions (such as demotion or dismissal). However, several considerations suggest that the imposition of punitive damages on firms will have a smaller effect on the punishment of blameworthy employees than might at first be supposed.

First, culpable employees may not be punished by firms because firms may have difficulty identifying them. Second, even if culpable individuals within a firm can be identified and punished by the firm, imposing punitive damages on firms often will have little or no marginal effect on their punishment. That is, the internal sanction imposed on such employees may not be much (if at all) greater as a result of the firm's bearing both punitive and compensatory damages than if the firm had borne compensatory damages alone, because the latter may result in the firm imposing the maximum internal sanction on the employee. Additionally, there may not even exist culpable employees in the firm to punish: responsibility for a decision may be so dispersed that no one person would be considered blameworthy with respect to it; and even if there are such persons, they may have changed jobs, retired, or died by the time a judgment is rendered.

A further point is that imposing punitive damages on firms often penalizes the firms' shareholders and customers, who ordinarily would not be thought to deserve punishment. This adverse consequence of punitive damages must be weighed against the beneficial effects of such damages in furthering the punishment goal.

To amplify, shareholders, as residual claimants on a firm's profits, obviously will be made worse off when punitive damages are imposed on a firm. The question, however, is whether they should be punished. If a shareholder owns a significant share of a firm's stock, participated actively in the firm's decisions and acted egregiously, then his position would be much the same as that of a blameworthy employee with decision-making power; each would be morally culpable. But if a shareholder owns a minuscule fraction of the stock of the firm, and was a passive investor with no direct involvement in the firm's decision-making processes, then his degree of blameworthiness is small, if it exists at all.

A firm's customers also will suffer from the imposition of punitive damages on the firm if such damages cause the prices of the firm's products or services to rise. This can occur because firms may regard punitive damages as an additional cost of doing business – a cost that, with a positive probability, will be borne by them beyond their ordinary costs. To cover the added cost of punitive damages, therefore, firms will have to raise their prices, which will cause the welfare of their customers to decline. It seems clear, however, that customers would not ordinarily be considered blameworthy, because they do not exert direct control over the actions of firms that pose risks to other persons.

Thus, assuming that the punishment objective with respect to firms is to ensure that blameworthy individuals

are penalized, punitive damages do not accomplish this objective in a direct way and also tend to penalize parties who are not blameworthy.

7. OPTIMAL DAMAGES IN THE LIGHT OF BOTH OBJECTIVES. The levels of damages that are optimal from the perspective of the two separate objectives of deterrence and of punishment will generally be different. Notably, the level that is best for deterrence is likely to exceed that which is best for punishment if the chance of being found liable is low and the magnitude of punitive damages necessary for deterrence therefore is high. Conversely, the level that is best for punishment is likely to be higher if the chance of being found liable is high, because then optimal damages for purposes of deterrence are approximately equal to harm, but the reprehensibility of the defendant's act presumably calls for extra-compensatory damages to serve the punishment objective.

It is evident that the optimal level of damages overall – that which maximizes a measure of social welfare combining both objectives – is a compromise between the levels that are optimal when each objective is considered independently, as noted in Polinsky and Shavell (1998: 103) and developed in Diamond (1997b).

8. EXTENSIONS OF THE ANALYSIS. In this section we will consider a variety of additional topics, focusing on the deterrence objective and usually restricting attention to the rationale for punitive damages stemming from the chance that the injurer can escape liability. We will, however, mention the punishment objective when it seems of particular importance.

Reprehensibility of conduct. The law requires that a defendant be found to have acted in a reprehensible manner before punitive damages can be imposed on him. However, as emphasized in Galligan (1990: 62–64) and Polinsky and Shavell (1998: 44–48), this legal policy is often inconsistent with the deterrence objective. On one hand, emphasis on reprehensibility may lead to imposition of punitive damages when such damages are not needed to achieve deterrence because the injurer is virtually certain to be found liable (as when a surgeon fails to remove a surgical tool from his patient). On the other hand, the converse problem may arise: an individual's harmful conduct may not be reprehensible but nevertheless may be unlikely to result in his liability (as when a truck inadvertently spills toxic waste onto a highway at night).

Notwithstanding the preceding observations, basing the level of punitive damages on the reprehensibility of the defendant's conduct may be proper with respect to the deterrence objective for acts leading to gains that are socially illicit: such acts usually are considered reprehensible and, as observed in section 2, punitive damages may be necessary to deter them.

From the perspective of the punishment objective, the focus on reprehensibility is clearly sensible, because reprehensibility of conduct is essentially synonymous with the actor's blameworthiness and thus with the need for punishment.

Wealth of injurers. The courts often state that a defendant's financial condition is a relevant factor in setting a punitive damages award, with the understanding that higher punitive damages may be appropriate for defendants with higher wealth.

With regard to deterrence, however, damages usually should not depend on the injurer's wealth; see, for example, Abraham and Jeffries (1989), Cooter (1989: 1176-7), and Polinsky and Shavell (1998: 50-54). The reason, in essence, is that if parties make decisions about precautions and choice of activity based on the expected value of their liability – that is, if they act in a risk-neutral way – their decisions will not depend on their wealth, and thus there is no reason to link damages to wealth. This point generally applies to corporations since, for the reason discussed at the end of section 1, they can be treated as risk neutral. It also applies to individuals if they are risk neutral or have access to liability insurance.

A qualification is that if individuals are risk averse and cannot obtain liability insurance, then optimal damages may depend on their wealth. To elaborate, the optimal level of damages for such individuals tends to be lower than that indicated by the multiplier formula presented in section 2. (We noted an analogous point at the end of section 1, where we observed that, for uninsured, risk-averse individuals who are found liable for sure, optimal damages are less than harm.) Further, the more risk averse an individual is, the lower the optimal level of damages. Assuming that poor individuals are more risk averse than rich individuals, this implies that the optimal level of punitive damages is lower for poorer individuals. Equivalently, punitive damages should be higher for wealthier individuals.

Another qualification to the conclusion that a defendant's wealth should not bear on the level of damages needed for proper deterrence arises when an injurer's gain is considered socially illicit. An injurer's wealth may then be relevant because the sanction necessary to offset his illicit gain will be higher the higher is his wealth, assuming that his marginal utility of money declines with his wealth.

Similarly, an individual's wealth may be relevant to the level of punitive damages that will achieve appropriate punishment. For to impose a given disutility on an individual, he must pay more if he is wealthy than if he is not, again assuming that his marginal utility of money declines with his wealth. But if the injurer is a firm and the punishment objective is concerned with punishing culpable employees, the firm's wealth generally would not be relevant to satisfaction of the punishment objective. This is because there is no general reason to believe that the penalties that a firm imposes on its employees for misbehaviour will be a function of the firm's wealth, and thus no reason to think that achievement of the punishment objective will be served by linking punitive damages to a firm's wealth.

Whether victims are strangers or customers. Although we have so far implicitly assumed that the parties harmed by injurers are 'strangers' – parties who have no market or contractual relationship with the injurer – victims of harm are often customers of defendant firms. The status of victims either as

strangers or as customers is important to consider, although courts generally do not observe this distinction.

When customers might be harmed by the products (or services) they buy, firms will tend to be concerned that customers may not be willing to pay as much for the products or that they may stop purchasing the products altogether. Given that firms have this market-based incentive to be attentive to the risk of harm to their customers, the need for liability in general, and for punitive damages in particular, to control injurer behaviour is diminished. The more knowledgeable customers are about product hazards, the less the need for punitive (or any) damages. Obviously, this market mechanism cannot operate if the victims are strangers to the defendant. On punitive damages and the customer relationship, see Craswell (1996) and Polinsky and Shavell (1998: 78-80).

Litigation costs. Litigation costs may be relevant to the calculation of punitive damages because they may influence the probability of suit, and therefore the chance of escaping liability. If litigation costs are significant relative to the expected gain from suit, the probability of suit may be small, and this fact may justify imposing punitive damages on the injurer. However, litigation costs often will be insignificant in relation to the expected gain from suit, so that the probability of suit may be presumed to be very high. Then, consideration of litigation costs does not provide a basis for imposing punitive damages.

Note that punitive damages should not be awarded for the purpose of spurring suit. The damage multiplier formula is designed to achieve appropriate deterrence when suit does not always occur, so it is not necessary to award damages to increase the probability of suit (provided that the probability of suit is not so low that the implied level of damages exceeds the defendant's ability to pay). Indeed, encouraging lawsuits would increase social costs and therefore is socially undesirable, other things equal.

The tendency of higher damage awards to increase litigation costs lends appeal to the policy of decoupling punitive damages, that is, awarding the plaintiff only a part of the punitive damages judgment paid by the defendant, with the remainder going to the state. Use of decoupling allows society to discourage excessive spending on litigation (the plaintiff receives less than otherwise) without diluting deterrence (the defendant can still be made to pay an appropriate penalty). On punitive damages and litigation costs generally, see Polinsky and Shavell (1998: 62-6); see also Kahan and Tuckman (1995) for a discussion of punitive damages and decoupling.

Insurability. The question whether liability insurance for punitive damages should be permitted is of interest, in part because legal policy on this matter varies among the states. The basic answer to this question is that punitive damages should be insurable when the justification for punitive damages is that injurers might escape liability.

The reasons for allowing liability insurance for punitive damages are essentially the same as those for allowing liability insurance for compensatory damages. These reasons are easiest to explain when liability is strict and harm is solely monetary. In that case, the sale of liability insurance

cannot hurt potential victims, since they will be fully compensated for any loss; and the insurance must raise the well-being of injurers if they elect to buy it. The arguments for allowing liability insurance in other contexts are more complicated. A potential qualification to all of these arguments arises if injurers are judgment proof; the availability of insurance could worsen injurers' behaviour in these circumstances (although insurance also could improve matters). For discussions of insurance and punitive damages, see Priest (1989) and Polinsky and Shavell (1998: 75–8).

Tax treatment. If punitive damages are imposed on an injurer as a result of his engaging in a business activity, such damages are generally tax deductible, just as are compensatory damages in those circumstances. This policy is socially desirable given the deterrence objective. For if punitive damages were not deductible, their after-tax cost to injurers would be artificially inflated relative to the costs of taking precautions, which are deductible. Consequently, injurers would be induced to take excessive precautions (and, for similar reasons, to be overly deterred from participating in risky activities). The general point that damages should be deductible, given that precaution costs are deductible, originated with Png and Zolt (1989) and carries over to the situation when parties might escape liability and punitive damages might be imposed as a result.

A. MITCHELL POLINSKY AND STEVEN SHAVELL

See also CORPORATE CRIME AND ITS CONTROL; CRIMINAL JUSTICE; DUE CARE; ECONOMIC THEORY OF LIABILITY RULES; JUDGMENT-PROOFNESS; LEGAL STANDARDS OF CARE; PUBLIC ENFORCEMENT OF LAW; THIRD-PARTY LIABILITY; TYPES OF TORT; VICARIOUS LIABILITY.

Subject classification: 5d(iv); 5e.

BIBLIOGRAPHY

- Abraham, K.S. and Jeffries, J.C., Jr. 1989. Punitive damages and the rule of law: the role of defendant's wealth. *Journal of Legal Studies* 18: 415–25.
- Bentham, J. 1838–43. Principles of Penal Law. In *The Works of Jeremy Bentham* Vol. 1, ed. J. Bowring, New York: Russell & Russell Inc., 1962.
- Biggar, D. 1995. A model of punitive damages in tort. *International Review of Law and Economics* 15: 1–24.
- Calabresi, G. and Melamed, A.D. 1972. Property rules, liability rules,

- and inalienability: one view of the cathedral. *Harvard Law Review* 85: 1089–1128.
- Cooter, R.D. 1982. Economic analysis of punitive damages. *University of Southern California Law Review* 56: 79–100.
- Cooter, R.D. 1989. Punitive damages for deterrence: when and how much? *Alabama Law Review* 40: 1143–96.
- Craswell, R. 1996. Damage multipliers in market relationships. *Journal of Legal Studies* 25: 463–92.
- Diamond, P.A. 1997a. Efficiency effects of punitive damages. Unpublished manuscript, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.
- Diamond, P.A. 1997b. Integrating punishment and efficiency concerns in punitive damages for reckless disregard of risks to others. Unpublished manuscript, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.
- Eisenberg, T., Goerd, J., Ostrom, B., Rottman, D. and Wells, M.T. 1997. The predictability of punitive damages. *Journal of Legal Studies* 26: 623–661.
- Ellis, D.D., Jr. 1982. Fairness and efficiency in the law of punitive damages. *Southern California Law Review* 56: 1–78.
- Galligan, T.C., Jr. 1990. Augmented awards: the efficient evolution of punitive damages. *Louisiana Law Review* 51: 3–85.
- Haddock, D.D., McChesney, F.S. and Spiegel, M. 1990. An ordinary economic rationale for extraordinary legal sanctions. *California Law Review* 78: 1–51.
- Kahan, M. and Tuckman, B. 1995. Special levies on punitive damages: decoupling, agency problems, and litigation expenditures. *International Review of Law and Economics* 15: 175–85.
- Kaplow, L. and Shavell, S. 1996. Property rules versus liability rules: an economic analysis. *Harvard Law Review* 109: 713–90.
- Landes, W.M. and Posner, R.A. 1981. An economic theory of intentional torts. *International Review of Law and Economics* 1: 127–54.
- Landes, W.M. and Posner, R.A. 1987. *The Economic Structure of Tort Law*. Cambridge, MA: Harvard University Press.
- Png, I.P.L. and Zolt, E.M. 1989. Efficient deterrence and the tax treatment of monetary sanctions. *International Review of Law and Economics* 9: 209–17.
- Polinsky, A.M. and Shavell, S. 1998. Punitive damages: an economic analysis. *Harvard Law Review* 111 (page citations are to the July 1997 pre-publication version circulated as Working Paper No. 143, John M. Olin Program in Law and Economics, Stanford Law School, Stanford, CA).
- Posner, R.A. 1973. *Economic Analysis of Law*. Boston, MA: Little, Brown & Co.; 1st edn.
- Priest, G.L. 1989. Insurability and punitive damages. *Alabama Law Review* 40: 1009–1035.
- Shavell, S. 1987. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.
- Stoll, H. 1983. Consequences of liability: remedies. In *International Encyclopedia of Comparative Law* 11, ed. A. Tunc, Tübingen: J.C.B. Mohr (Paul Siebeck) and Dordrecht: Martinus Nijhoff Publishers.