Economic Analysis of Punitive Damages

Robert D. Cooter

Berkeley Law

Follow this and additional works at: https://scholarship.law.berkeley.edu/facpubs

Part of the Law Commons

Recommended Citation

Economic Analysis of Punitive Damages, 56 S. Cal. L. Rev. 79 (1982)
Litigating a tort dispute involving punitive damages, much like navigating the Straits of Magellan, runs the risk of incurring grave losses from colliding with unseen objects. There is no clear standard for deciding when punitive damages are appropriate or for computing their magnitude when awarded. This paper develops such standards with the help of economic theory.

Imagine any potentially harmful activity in which liability is governed by a fault rule. The legal standard of care represents a threshold, on one side of which the injurer is liable and on the other side he is not liable. Crossing this threshold causes an abrupt jump in the costs which the injurer expects to bear, with profound effects upon rational behavior. The economic model developed in this paper is based upon an analysis of these threshold effects, which have not been explained previously in a systematic fashion.

I begin with a brief, nontechnical explanation of the policy conclusions of this analysis. For most potential injurers, it is far cheaper to comply with the law than risk liability, so noncompliance will usually be unintentional. If fault is unintentional, then imposing punitive damages in addition to compensatory damages is both unnecessary for deterrence and undeserved as punishment. However, there may be a small group of unusual persons who derive illicit pleasure from noncompliance or incur exceptional costs from compliance. Such people may cross the threshold of fault intentionally and, once crossed, it usually pays them to cross it a long way. The situation is analogous to the historical conflict between France and Germany in which armies did not cross the Rhine merely to occupy the far bank. This fact provides a behavioral test for intentionality: Intentional fault is usually aggra-
vated and constitutes gross negligence, willful and wanton disregard for others, and the like. Three policy conclusions follow from this analysis: (1) punitive damages should be restricted to intentional faults; (2) a criterion for identifying intentional fault is that it is gross or repeated; and (3) punitive damages should be computed to offset the injurer's illicit pleasure from noncompliance or exceptional cost of compliance.

In Part I of this paper, the distinction between intentional and unintentional fault is explained by the distinction between equilibrium and disequilibrium behavior, and the concept of a threshold is explained by discontinuities in cost functions. In Part II, the theory is used to determine whether the purpose of punitive damages is to punish the defendant, to deter potential injurers, to reward victims for suing, or to correct a shortfall between actual harm and allowable damages. In Part III, the difference between negligence and strict liability is explained by the presence or absence of threshold effects. Finally, Part IV discusses some applications of the model, including the conclusion that consumers injured by defective products should not be awarded punitive damages unless they prove that the manufacturer knew that his behavior was negligent and likely to cause a defect. The paper concludes with a mathematical appendix.

I. FAULT RULES AND INCENTIVES

Economists often explain a person's behavior by observing that the costs and benefits of a change are perfectly balanced at the margin, so there is no gain from acting differently. However, a threshold in costs can create an unbalanced situation in which the costs of a change substantially exceed the benefits. For example, the cost of complying with a negligence rule will be far less than the cost of noncompliance for most people. In technical language, the difference is between an "internal equilibrium" where the marginal costs and marginal benefits of change are equal, and a "corner solution" where the costs of change exceed the benefits. This distinction is important for liability law because a fault rule creates a corner solution in which the potential injurer will not change his precaution in spite of moderate changes in the costs or benefits or precaution.

A simple numerical illustration is provided by Tables 1 and 2, which depict the costs associated with exploding pop bottles. In this example, precaution against exploding bottles takes the form of testing bottles for weaknesses in the glass. There are two possible levels of testing: sampling one bottle in fifty or sampling one bottle in one hun-
dred. An increase in the sampling ratio from 1/100 to 1/50 causes the cost of precaution to rise from 40 to 65, as indicated in column 1. However, this increase in precaution causes the cost of the harm suffered by victims to fall from 80 to 40, as indicated in column 2. The social cost is the sum of the cost of precaution and the cost of the harm, as indicated in column 3. Efficiency requires the bottler to sample at the ratio which minimizes social costs, namely 1/50.

TABLE 1: COSTS OF EXPLODING POP BOTTLES

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of Precaution</td>
<td>Cost of Accidents</td>
<td>Social Costs</td>
<td>Cost of Precaution Plus 50% Accident Costs</td>
</tr>
<tr>
<td>Sample 1 in 50</td>
<td>65</td>
<td>40</td>
<td>105</td>
</tr>
<tr>
<td>Sample 1 in 100</td>
<td>40</td>
<td>80</td>
<td>120</td>
</tr>
</tbody>
</table>

*aSocial cost is the sum of the cost of precaution and the cost of accidents.*

TABLE 2: COSTS BORN BY BOTTLER

| | Negligence Rule | Strict Liability Rule |
| | 100% Liability | 50% Liability | 100% Liability | 50% Liability |
| Sample 1 in 50 | 65 (not negligent) | 65 (not negligent) | 105 | 85 |
| Sample 1 in 100 | 120* (negligent) | 80* (negligent) | 120 | 80 |

*aThe cost to a negligent bottler is the sum of his cost of precaution and the costs of the accidents for which he is sued (100% or 50% of the total number, as the case may be).*

Table 2 depicts the incentives for precaution created by a rule of strict liability and a negligence rule. Assuming that the standard of negligence requires a sampling ratio of 1/50, a negligence rule presents the bottler with a choice between sampling at the ratio 1/50 and escaping liability, in which case his costs are 65, or sampling at the ratio 1/100 and assuming liability. If the bottler assumes liability, then his costs are 120 or 80, depending upon whether he is liable for 100 percent or 50 percent of the harm. The important point is that the bottler's costs are lower under a negligence rule when he complies, even if fault exposes him to liability for only half of the actual harm. Thus, in this example, a negligence rule provides efficient incentives even though a negligent injurer does not bear the full cost of accidents.

By contrast, consider the bottler's costs under a rule of strict liability. If he is liable for 100 percent of the harm, then it is cheaper to
sample at the ratio 1/50 and bear costs of 105, than to sample at the ratio 1/100 and bear costs of 120. However, if he is liable for 50 percent of the harm, then it is cheaper to sample at the ratio 1/100 and bear costs of 80, than to sample at the ratio 1/50 and bears costs of 85. This illustrates the fact that strict liability does not provide efficient incentives to the injurer unless he bears 100 percent of the cost of harm.

The reader may wonder why a negligence rule may provide efficient incentives without fully internalizing costs, whereas a strict liability rule cannot provide efficient incentives unless the injurer bears 100 percent of accident costs. The reason is that the level of precaution defined in the legal standard of negligence is a threshold at which costs jump abruptly, rather than increase marginally. In order to understand this fundamental difference between negligence and strict liability, it is necessary to use some simple graphs and mathematical notation.

Let $x$ denote the injurer's expenditure on precaution, let $A$ denote the dollar value of the harm caused by an accident, and let $p$ denote the

![Figure 1: Social Costs of Accidents](image)
probability of avoiding an accident. The social cost, $SC$, is the expenditure on precaution plus the expected harm, or $SC = x + (1-p)A$. Figure 1 graphs the relationship between social cost and precaution. The probability $p$ of avoiding an accident is an increasing function of the expenditure on precaution, $x$, or $p = p(x)$, where $p>0$. At low levels of precaution, a little more precaution saves more than it costs, but the opposite is true at high levels of precaution. The efficient level of precaution, $x_e$, is an intermediate value at which social costs are minimized.

It is easy to represent graphically the injurer's costs under a fault rule. For simplicity, assume that negligent injurers must fully compensate their victims, and assume that the legal standard of fault, $x^*$, corresponds to the efficient level of precaution, $x_e$, that is, $x^* = x_e$.

\[ \text{FIGURE 2: INJURER'S COSTS} \]

---

1. Social costs, $x + (1-p(x))A$, are minimized when $x$ is chosen to satisfy the inequality:
As depicted in Figure 2, if the injurer's precaution is less than the legal standard, \( x < x^* \), then he bears the social cost of accidents. Thus, the injurer's cost curve \( C(x) \) in Figure 2 is identical to the cost curve in Figure 1 for values of \( x \) below \( x^* \). If his precaution is above the legal standard, \( x \geq x^* \), then he only bears the cost of his precaution, in which case his costs \( C(x) \) equal his expenditure on precaution, \( x \). Thus, his cost curve is in two pieces, with a jump or discontinuity at \( x = x^* \). The

![Figure 3: Incomplete Liability](image)

\[ C(x) = x + (1-p)A \]

Solving this inequality for \( x \) gives the efficient level of precaution \( x_e \). This formula also gives the legal standard of fault, \( x^* \), prescribed by the marginal Hand Rule. This rule was formulated in Brown, *Toward an Economic Theory of Liability*, 2 J. LEGAL STUD. 323, 332 (1973). There is a fundamental misconception, however, in Brown's account of the Hand Rule in that he concludes that courts cannot find the correct legal standard when they have limited information about the costs of precaution. This misconception is addressed in Cooter, Kornhauser & Lane, *Liability Rules, Limited Information, and the Role of Precedent*, 10 BELL J. ECON. 366, 369-73 (1979) (where variations in injurer behavior determine the information available to the court, rather than the legal standard itself, a sequence of precedents tends toward efficiency).
injurer’s costs jump from $x$ to $x + [1-p(x)]A$ at the point where he crosses the threshold of liability.

The cost curve in Figure 2 reaches its lowest point at $x=x^*$, so a cost-minimizing injurer will exactly satisfy the legal standard of care. But Figure 2 is drawn under the assumption that negligent injurers must fully compensate their victims. Instead, assume that imperfections in the law prevent complete internalization of social costs. For example, suppose that damages cover only a fraction of the harm $A$, or that only a fraction of victims actually bring suit. Figure 3 depicts such a situation. If the injurer’s liability equals half the actual harm, then his costs are still minimized by satisfying the legal standard, $x=x^*$. If the injurer’s liability equals one-third of the actual harm, then his costs are minimized by taking precaution at the level denoted $x_1$, which is well below the legal standard, $x^*$. Figure 3 illustrates the fact that, under a negligence rule, liability must fall far short of the actual harm before the cost-minimizing injurer will reduce his precaution below the legal standard of care.

A. UNINTENTIONAL FAULT

In the preceding examples, the injurer complies with fault rules even though damages do not fully compensate victims and some victims do not sue. Would such an injurer ever be at fault? The answer is yes, because such an injurer might be at fault unintentionally. For instance, he might be mistaken about the legal standard, $x^*$, and believe that he is complying, when in fact he is not. Such mistakes can easily occur since the standard of “reasonable care” is so vague in tort law.

In economic jargon, unintentional fault corresponds to “disequilibrium behavior.” An economic equilibrium is a condition in which expectations are fulfilled, or, equivalently, in which actions have their intended effects. An economic disequilibrium is a situation in which expectations are disappointed, intentions are unfulfilled, decisionmakers are surprised, and so on. If the injurer thinks he is complying with the law, but at trial he learns that he is liable because he is in fact not complying, then the injurer will be disappointed and will want to revise his behavior. An equilibrium will be reached when he per-

---

2. For a general discussion of this topic, consult any microeconomic textbook. For an advanced treatment, see K. Arrow & F. Hahn, General Competitive Analysis (1971).
ceives the legal standard accurately and satisfies it, ending the possibility of further surprises.

As explained, unintentional fault can be corrected by awarding compensatory damages. Of course, it is also true that unintentional fault will be eliminated if the injurer bears more than the full cost of the harm that he causes. Adding punitive damages provides a second reason for compliance where the first suffices. When punitive damages are imposed for unintentional fault, the cost of making errors increases, causing injurers to invest more in avoiding mistakes. Thus, punitive damages for unintentional fault will not change the injurer's equilibrium behavior, but the movement from disequilibrium to equilibrium will be more rapid.

An argument can be made against imposing punitive damages for unintentional faults because punishment is unjust where no harm was intended. This claim is important, but this paper will not consider such moral arguments. Instead, it considers next the small number of cases where fault is intentional.

B. INTENTIONAL FAULT

Suppose the courts face an unusual individual who intended to breach his legal duty, despite liability for the actual damage. Requiring him to pay compensation will not cause surprise or disappointment. In a situation such as this, the court confronts equilibrium fault. Although intentions are difficult to prove in court, there is a behavioral guide that often distinguishes intentional from unintentional fault. Unintentional fault usually involves a moderate degree of noncompliance, whereas intentional fault usually involves gross noncompliance. Intentional fault is gross because a person who deliberately exposes himself to liability crosses a threshold, after which aggravation of the fault is less costly.

Table 3 illustrates why intentional fault is usually gross. It shows the options of a father who will miss his daughter's wedding if he complies with traffic laws. Assume the father is willing to pay up to $1,000

3. For elaboration of this point, see discussion of Figure 5 in the text.
4. The scope and purpose of imposing punitive damages is discussed in K. REDDEN, PUNITIVE DAMAGES §§ 2.2 (origin and development of punitive damages), 4.1-4.36 (cases in which punitive damages may be awarded) (1980 & Supp. 1981).
5. This paper is concerned with intentional fault, rather than intentional harm. Thus, a driver may intend to speed but hope that his speeding will not result in harm. Several meanings of an intentional tort are distinguished in Landes & Posner, An Economic Theory of Intentional Torts, 1 INT'L REV. L. & ECON. 127, 127-39 (1981).
ECONOMIC ANALYSIS

not to miss the wedding. If he does not speed, he will miss the whole wedding and suffer a loss valued at $1,000. If he speeds a little, then he will miss much of the wedding, with a loss valued at $400. If he speeds a lot, then he will miss some of the wedding, with a loss valued at $300. Even a little speeding makes him liable for all the damage resulting from a collision with another motorist who is not at fault. Thus, his expected civil liability jumps from $0 to $500 when he speeds a little. If he decides to speed a lot, the expected civil liability increases from $500 to $550 to reflect the increase in the probability of an accident. There is a threshold in liability costs when he decides to speed a little, but there is no abrupt jump when he decides to speed a lot. If it pays to cross the threshold at all, then it pays to cross it a long way. Thus, the father will minimize his costs by speeding a lot, as is apparent from column 3 of Table 3.

To explain this point more fully, suppose there is a special class of injurers who derive an illicit benefit from wrongful behavior or who experience exceptional costs from compliance. For notational simplicity, let $b$ represent such illicit benefits, where the benefit $b$ is a decreasing function of precaution $x$, or $b=b(x)$. If the injurer satisfies the legal standard, his costs are precaution costs less illicit benefits, or $x-b(x)$. If the injurer is at fault, then he bears these costs minus the illicit benefits, or $x-b(x)+[1-p(x)]A$. This situation is depicted in Figure 4, where there is a discontinuity or jump in costs at the legal standard, $x^*$. The difference between Figure 4 and Figure 2 is that the existence of illicit benefits causes the injurer to minimize his costs at the level of precaution, $\bar{x}$, a point at which he is at fault, rather than at the legal standard, $x^*$.

It is clear from Figure 4 that the level of precaution, $\bar{x}$, that minimizes the injurer's costs cannot be just a little below the legal standard,
The jump or discontinuity guarantees that the injurer's cost-minimizing level of precaution, $\bar{x}$, must be far below the legal standard, $x^*$. In technical terms, intentional fault must be gross because the discontinuity in costs at the legal standard creates a nonconvexity in the injurer's cost function. Consequently, there are two local minima, one at the legal standard and one far below it.\(^6\)

\(^{6}\) If gross fault triggers punitive damages, then an injurer may attempt to avoid punitive damages liability through noncompliance at a level of fault that is not gross enough to trigger punitive damages. For example, newspapers that specialize in gossip may publish libelous material that stops just short of causing public figures to sue. Repetition of mild fault is an indication that the fault is intentional, even though each individual act is not gross. A more detailed treatment of strategic behavior is beyond the scope of this paper.

\(^{7}\) For a discussion of nonconvexities and the law, see Cooter, *How the Law Circumvents Starrett's Nonconvexity*, 22 J. Econ. Theory 499, 500-01 (1980) (courts can distinguish between local and global optima by assigning liability rights according to a calculation of benefits and costs).
If punitive damages are added to compensatory damages, it is possible to restore a situation in which compliance minimizes the injurer's costs. This is accomplished by imposing punitive damages equal to the illicit benefits of non-compliance, $b$. If punitive damages exactly offset the injurer's illicit benefits, then his cost function will look just like that of an ordinary injurer in the zone of liability, as depicted in Figure 2. Thus, a good measure by which to set punitive damages is the illicit benefit or extraordinary cost that prompts the intentional fault. If this measure is adopted, then the amount of money awarded in punitive damages will usually be large, because the illicit benefit or extraordinary cost of gross fault is large.

It is important to consider the conditions under which intentional fault should be deterred. From the utilitarian or economic perspective, intentional fault should be deterred if its social cost exceeds its social benefit, but not otherwise. If the illicit pleasure from noncompliance is given no social weight, then the social cost of intentional fault will exceed the social benefit and deterrence will be appropriate. For example, the pleasure of defaming your enemy or breaking his nose is given no social weight by the courts, and punitive damages are therefore assessed to deter such behavior. However, a person who breaks into a cabin in the woods and eats food to keep from starving will not face punitive damages, even though his fault is intentional. Similarly, most courts would not excuse speeding on the grounds that the driver was late for his daughter's wedding, but speeding to bring a pregnant woman to the hospital might be excused under the emergency doctrine.

II. PURPOSES OF PUNITIVE DAMAGES

The model presented above highlights three broad characteristics of an economic view of punitive damages: (1) punitive damages should be restricted to intentional faults; (2) these faults will usually violate the legal standard of care by a wide margin; and (3) the awards for punitive damages should be large. These characteristics are essential to the

---

8. It may be best for reasons not stated in this model that punitive damages should internalize social costs exactly. To internalize social costs fully, the punitive damages should be set equal to the illicit benefit, $b$, plus the expected shortfall of compensatory damages from actual damages. If $D$ denotes compensatory damages, $q$ denotes the probability that the injurer will be held liable, and $P$ denotes punitive damages, then

Social Costs = $x + (1-p)A$
Injurer's Costs = $x - b + (1-p)q(D+P)$.

Punitive damages internalize social costs if the preceding values are equal, which implies:

$$P = \frac{A-qD}{q} + \frac{b}{(1-p)q}$$
institution of punitive damages and can clarify the purposes of that institution. Purposes that have been suggested for imposing punitive damages include punishing offenders, deterring injurers, rewarding plaintiffs who otherwise would not sue, and fully compensating victims. In this section, I consider which of these purposes are consistent with the three characteristics of punitive damages identified by economic analysis.

The first characteristic to consider is intentionality. As explained, compensatory damages are sufficient to deter unintentional faults and insufficient to deter intentional faults. Furthermore, punishment is often regarded as unjust if the fault is unintentional. Thus, restricting punitive damages to intentional faults is consistent with the purposes of deterrence and punishment. By contrast, there is no necessary relationship between the injurer’s intentions and the completeness of compensatory damages or the victim’s reluctance to bear the cost of a suit. If the purpose of punitive damages were to achieve complete compensation or to reward plaintiffs for suing, then the intentionality restriction would make no sense.

The second characteristic, restriction to gross faults, also seems consistent with punishment and deterrence. The fact that fault must be gross to trigger punitive damages can be viewed as a safeguard against mistakes in civil proceedings, which afford less protection to defendants than criminal proceedings. Such a safeguard seems desirable when punishment is at issue. Furthermore, restriction of punitive damages to gross fault is consistent with my analysis of deterrence based upon threshold effects, as explained in the preceding section.

The fact that punitive damages are often large is also consistent with deterrence and punishment. As noted, a substantial increment in damages is needed to offset the illicit pleasure from noncompliance or the exceptional costs of compliance which prompt intentional fault. Furthermore, intentional fault is more serious morally than unintentional fault, so an increase in the sanction is warranted. By contrast, small awards of punitive damages would be appropriate if the purpose were to reward plaintiffs for suing or to fully compensate victims.

Therefore, the three characteristics of punitive damages are consistent with the purposes of punishment and deterrence of exceptional injurers, but inconsistent with the purposes of rewarding plaintiffs for suing or fully compensating victims. I will now consider the difference

---

9. There is a troublesome question which remains: Why give the money to the victim,
between strict liability and negligence.

III. STRICT LIABILITY VERSUS NEGLIGENCE

The economic model developed in this paper is based upon the observation that a negligence rule creates a threshold or jump in the potential injurer’s costs. No such threshold or jump exists in the cost function of an injurer under a rule of strict liability. This observation can be developed into a theory about the difference between strict liability and negligence rules.

To develop this theory, Figure 5 depicts the precaution of a potential injurer under a negligence rule as a function of the extent of his liability. If negligent injurers are liable for only a small fraction of
accident costs, $A$, then a cost minimizing injurer will take a level of precaution, $x$, which is below the legal standard of care, $x^*$. As the fraction of liability increases, the cost minimizing level of precaution increases, which is represented in Figure 5 by the upward sloping line to the left of the switching point. Eventually, a switching point is reached at which damages are high enough so that the cost minimizing level of precaution is the legal standard, $x^*$. The switching point is by definition the lowest level of damages at which compliance with the legal standard of care is cost minimizing. Once total damages are high enough to induce compliance, further increases in damages do not influence the injurer’s precaution. This fact is represented in Figure 5 by the horizontal line $x=x^*$ to the right of the switching point.

![Injurer's Precaution vs. Damages](image)

**FIGURE 6: CARE AND DAMAGES WITH STRICT LIABILITY**

Figure 6 illustrates the relationship between damages and the amount of precaution taken by an injurer under a rule of strict liability. The injurer is liable at all levels of care, so precaution increases continuously, without jumps, as damages increase. If social costs are perfectly internalized, so that damages equal the harm $A$ suffered by
accident victims, then the injurer’s precaution will be at the socially optimal level $x_*$. If some costs are externalized, however, then the injurer will not take enough precaution. In either case, an increase in damages causes the injurer’s precaution to increase.

There is a disadvantage of fault rules which is illustrated by a comparison of Figures 5 and 6. Whenever damages exceed the switching point in Figure 5, the injurer will satisfy exactly the legal standard of care. If the standard is too high, then his precaution will be too high, and if the standard is too low, then his precaution will be too low. In brief, fault rules make the injurer’s precaution sensitive to variations in the legal standard and insensitive to variations in damages. By contrast, strict liability makes the injurer’s precaution sensitive to variations in the legal standard and insensitive to variations in damages. Thus, the danger with fault rules lies in imperfect standards of fault, not imperfect computation of damages, while the opposite is true for strict liability rules.

From an efficiency viewpoint, the choice between fault rules or strict liability rules should be guided by considering the kinds of errors the courts are likely to make. A fault rule is preferable if the courts are likely to make errors in setting damages and unlikely to make errors in setting the legal standard of care. Conversely, a strict liability rule is preferable if the courts are likely to make errors in setting the legal standard of care and unlikely to make errors in setting damages. Computing damages involves measuring the harm suffered by the victim, whereas setting an efficient legal standard of care involves balancing the marginal cost and benefit of precaution. Thus, the choice between negligence and strict liability should be guided by whether the court is more likely to make errors in computing the harm suffered by the victim or in computing the marginal value of precaution.

Courts are likely to set damages below the actual harm if part of the harm is speculative or intangible. Harm which is speculative or intangible is difficult to prove in court and may be excluded from damages as a matter of law. By contrast, the courts are likely to set the legal standard of care incorrectly if the costs and benefits of marginal changes in the injurer’s precaution are difficult to observe. Returning

---

10. As explained in the first footnote, the marginal Hand Rule sets the legal standard, $x^*$, at the value of $x$ which solves the inequality:

$$0 \leq I - p'(x)A.$$ 

If damages, $D$, underestimate accident costs, $A$, however, so that $D < A$, then courts may solve the inequality this way:

$$0 \leq I - p'(x)D.$$
to the example of exploding pop bottles, the courts may have difficulty determining how many accidents would be eliminated if the bottler sampled at a higher rate. Such a determination is particularly hard for the courts if it involves understanding technology and evaluating statistics.

IV. SOME APPLICATIONS OF THE MODEL

In the preceding analysis, the economic model of punitive damages concerned gross negligence, but the original cases involved intentional torts such as assault or slander, and many contemporary cases involve consumer injuries for which the rule to be applied is strict liability. Thus, one issue is whether the formal model applies to cases involving a rule other than negligence.

The applicability of the model to assault is obvious after reinter-

\[
C = A(x) + k
\]

in which case \( x^* \) will be below the efficient value.

FIGURE 7: INJURER'S COST OF ASSAULT
ECONOMIC ANALYSIS

preparing its variables. Assault can be scaled from mild to aggravated, so the variable $x$ can be reinterpreted to denote the level of assault. Any level of assault is prohibited, so legal standard $x^*$ is taken to be nil, $x^*=0$. Crossing the threshold creates the possibility of a costly dispute and criminal sanctions, so let $k$ denote the threshold costs. Besides the threshold costs, the injurer’s costs increase with the seriousness of the assault, that is, $A=A(x)$. The increase in costs is due to an increase in civil and criminal liability that goes with a more serious assault. The injurer’s costs can be graphed as in Figure 7, with a threshold in costs at $x^*$, just as in Figure 2.

From a mathematical viewpoint, there is little difference between Figures 2 and 7. This is why the formal model presented here applies to assault as well as to negligence. The argument that most potential injurers will minimize costs by compliance with the law is still true, as well as the fact that a few injurers will intentionally violate the standard because they either enjoy illicit pleasure from noncompliance or incur exceptional costs from compliance. The only substantial difference between the model of assault and the model of negligence is the manner in which the jump in costs is interpreted. With assault, the jump is caused by the injurer’s risk of criminal sanctions and his need for defense. By contrast, the jump in costs in negligence cases is caused by the injurer’s abrupt assumption of liability when he crosses the threshold of fault.

Another aspect of the model that warrants attention is its application to consumer injuries.\textsuperscript{11} The conventional rule of law holds the manufacturer strictly liable for accidents caused by defective products. To recover compensatory damages, the injured consumer must prove that a defective product caused his injury. However, to recover punitive damages, the economic model suggests that the consumer should also be required to prove that the manufacturer knowingly caused the defect. If there is no fault in the manufacturer’s quality control, but a defective product slips through and injures someone, then punitive damages are unwarranted. As noted in the preceding section, awarding supercompensatory damages distorts the incentives of potential injurers when the rule is strict liability. If punitive damages are awarded to consumers without requiring proof of intentional fault, then manufacturers will take excessive precaution.

Before concluding this paper, there are several applications which should be mentioned in passing, even though space does not permit a detailed treatment. In modern cases, the injurer is often the agent of a corporation, which raises the question of vicarious liability for the intentional faults of employees. From an economic standpoint, the appropriateness of vicarious liability turns upon whether it is cheaper for courts to control the employees by punishing them directly or to do so indirectly by punishing the employers. Generally, increasing an employer's liability will cause the employer to increase his vigilance over the actions of his employees. It is impossible to discuss in detail the question of efficient vigilance, but economic theory favors vicarious liability for punitive damages if the cost of the resulting increase in employer vigilance is less than the harm that it averts.\footnote{Application of the Coase Theorem to this issue leads to the conclusion that it does not matter which rule of law is used in vicarious liability cases, as long as that rule presents no impediment to private contracts. The Coase Theorem suggests that the law regulating vicarious liability should be structured to save transaction costs in private contracts.}

Another issue that warrants attention concerns the propriety of allowing liability insurance to cover punitive damages. The primary objection to allowing these types of insurance contracts to be written is the fear that this practice will lead to an increase in the very behavior that the courts are trying to deter. In most situations, however, it would seem that such insurance would be desirable to both injurers and victims. Injurers can decide for themselves whether they are better off with or without the insurance, so prohibiting its sale would not be in their best interest. And, when punitive damages are awarded in addition to compensatory damages, victims are usually overcompensated. Thus, even if critics are correct and such insurance leads to an increase in injuries, victims who receive compensatory plus punitive damages may be better off than if they never suffered the accident at all. If both parties are better off as a result of insurance, it should not be prohibited. Of course, insurance companies may be unwilling to write such contracts,\footnote{If punitive damages are awarded only for cases of intentional fault, then any injurer who wishes to insure against a punitive damages assessment must, at the least, conceive of a situation in which he intends to be at fault. This prospect is enough to cause any insurance company to approach such a contract with trepidation.} but there is no need for the state to prohibit them.

The insurance issue raises another important problem: the eager victim. The eager victim is one who wants to suffer an injury in order to recover damages. If punitive damages overcompensate, they can create disincentives for precaution by victims. This problem has been...
extensively investigated in the economic analysis of law. Generally, a defense such as contributory negligence will prevent inadequate precaution by victims.

The arguments against awarding punitive damages when the rule of law is strict liability explain why punitive damages are rarely awarded for breach of contract. Absent specific defenses or excuses, a party who breaches a contract is strictly liable for the resulting harm. There is no threshold of fault that must be crossed to assume liability. Consequently, the potential injurer's precaution against events that would cause him to breach increase continuously with his liability. If punitive damages are imposed in addition to damages that compensate the victim, then the precaution against breach will be excessive.

As a final remark, there is a disparity between the economic model of negligence developed in this paper and the actual rules used by courts. The standard of fault discussed in my model, namely that level which minimizes social costs, is clear and precise, which is why it could be represented by the variable $x^*$. On the other hand, the standard used by courts, the behavior of a "reasonable" person, is vague. The potential injurer cannot always predict whether or not courts will hold him negligent for exercising a particular level of care. Fortunately, vagueness in the standard of fault does not alter the conclusions of the model, although demonstrating this fact involves some technical arguments.

CONCLUSION

It is characteristic of moral and legal systems that fault is a discrete judgment: either the behavior crosses the threshold of fault or it does not. If liability is contingent on fault, then the potential injurer's costs will jump at the threshold of fault. Consequently, an injurer concerned

14. The economic analysis has usually allowed precaution to be bilateral, by which I mean both the injurer and the victim can reduce the probability or extent of an accident. See, e.g., Brown, supra note 1; Cooter, Kornhauser & Lane, supra note 1.


The similarity of structure between the model of torts and contracts is discussed in R. Cooter, Utility in Torts, Contracts, and Property (Fall 1981) (unpublished manuscript).

16. For a detailed exposition of this point, see infra Mathematical Appendix.

with minimizing costs is likely to be slightly at fault only through bad judgment or an honest mistake. An injurer whose fault is intentional is likely to be grossly at fault, and a large punishment is needed to deter him. Thus, the economic model shows the relationship among the three broad characteristics of punitive damages noted above: the fault must be intentional, the fault usually must be gross or repeated, and the punitive damage award must be large.

The economic model suggests that punitive damages serve the purposes of deterring future injuries and rewarding plaintiffs for undertaking the additional burden of proving that fault was intentional. The economic model, however, does not support the view that punitive damages should correct the shortfall between actual harm and allowed compensation, or reward plaintiffs who otherwise would not sue. The social efficiency of fault rules is not disturbed by externalization of costs, provided its size is moderate. Rather, the efficiency of fault rules is disturbed by failure to compute correctly the legal standard of fault. By contrast, rules of strict liability are socially inefficient even if only some costs are externalized.

The economic model does suggest how courts might improve the consistency of their punitive damages rules. First, punitive damages should be regarded as an unusual measure, appropriate only for gross, intentional fault. Punitive damages should not be used to correct small imperfections in computing damages or in bringing suits. Second, punitive damages should be computed at a level that offsets the illicit pleasure of noncompliance or the exceptional costs of compliance that motivated the injurer. Fortunately, this computation does not have to be exact to achieve deterrence. Finally, punitive damages should be awarded in strict liability cases only if the plaintiff can prove intentional fault.
MATHEMATICAL APPENDIX

This appendix provides a brief formal statement of the model used in this paper. The notation used is as follows:

- \( x \) = injurer's precaution, where \( x \geq 0 \) (dollars);
- \( x^* \) = precaution required by law (dollars);
- \( p \) = probability of no harm, where \( p=p(x) \) and \( 0 \leq p \leq 1 \);
- \( q \) = probability that the victim will press his claim against the injurer;
- \( A \) = dollar value of harm (dollars);
- \( D \) = damages (dollars);
- \( b \) = extraordinary compliance costs (or illicit benefit from non-compliance), where \( b=b(x) \) (dollars);
- \( C \) = injurer's costs, where \( C=C(x) \) (dollars).

**Behavior of the Injurer**

Assume there is a legal standard, \( x^* \), and a rule that the injurer is liable if and only if \( x < x^* \). The injurer's costs, \( C(x) \), under this fault rule would be:

\[
C(x) = \begin{cases} 
  x + b(x) + [1-p(x)]qD, & \text{if } x < x^* \\
  x + b(x), & \text{if } x \geq x^*.
\end{cases}
\]

The cost function \( C(x) \) has a discontinuity at \( x^* \) whose size is given by

\[
-[x^* + b(x^*]) + [x^* + b(x^*) + (1-p(x^*))qD] = (1-p(x^*))qD.
\]

If the injurer's cost minimizing precaution \( x \) is \( x^* \) for some values of \( q \) and \( D \), then the cost minimizing precaution will not vary with moderate changes in \( q \) and \( D \).

By contrast, the injurer's costs under a rule of strict liability are:

\[
C(x) = x + b(x) + [1-p(x)]qD.
\]

The value of \( x \) that minimizes the expression is sensitive to variations in \( q \) and \( D \).

**Social Costs**

The social costs are:

\[
SC = \begin{cases} 
  x + [1-p(x)]A, & \text{or} \\
  x + b(x) + [1-p(x)]A,
\end{cases}
\]

depending upon whether weight is given to \( b(x) \). Let \( x_e \) minimize \( SC \).

Let \( \tilde{x} \) denote the value of \( x \) which solves \( \inf_{x} C(x) \). Incentives are efficient when \( x_e \) is equal to \( \tilde{x} \). Efficient incentives are achieved under a fault rule for a range of values of \( q \) and \( D \) if the legal standard is efficient, or \( x^* = x_e \).
Suppose that fault is an equilibrium behavior:

\[
\inf_{x < x^*} \left[ x + b(x) + (1-p(x))qD \right] < \min_{x \geq x^*} \left[ x + b(x) \right].
\]

The purpose of punitive damages is to reverse the inequality. This can be achieved by a sufficiently large increase in \( D \).

**Unclear Standards**

This explanation of punitive damages has relied upon a discontinuity in the injurer's costs at the level of care equal to the legal standard, or at \( x = x^* \). Legal standards are vague, however, so injurers may not face a sharp discontinuity. I want to show that the same conclusions can be reached if standards are unclear. When standards are unclear, the injurer's cost function will contain a nonconvexity that has the same effect as a discontinuity.

Reformulating the model with a vague standard of negligence is straightforward. Let \( \pi(x^*) \) be the probability that the standard of negligence applied by the courts is \( x^* \). Thus, the probability that precaution level \( x \) will be judged to be nonnegligent can be written as:

\[
\pi(x) = \int_0^x \pi(x^*) \, dx^*.
\]

The expected cost function for the injurer is the same as before, except that it contains a new probability term:

\[
C(x) = x + [1-\pi(x)][1-p(x)]qD,
\]

or

\[
C(x) = x + [1-\pi(x)][1-p(x)]qD - b(x).
\]

Assume the standard of negligence is vague, but not wholly unknown. Thus, the probability density \( \pi(x^*) \) clusters around a certain value of \( x^* \). The discontinuity has disappeared from \( C(x) \) and has been replaced by the cluster of probability. In this case the cost function \( C(x) \) will be nonconvex as depicted in Figure 8. There will be two local minima, one near the cluster point for the legal standard and the other at the far lower level of precaution. The nonconvexity in Figure 8 plays the role of the discontinuity in Figure 2. As before, potential offenders will be divided into those who try to satisfy the legal standard but sometimes fail because a very high standard is sometimes applied by courts, and those whose fault is intentional. The policy conclusions and analysis of the purposes of punitive damages are the same when legal standards are vague as when the standards are clear.
ECONOMIC ANALYSIS

\[ C = x + (1 - \pi(x))(1 - p)qD - b(x) \]

FIGURE 8: INJURER'S COSTS WITH UNCLEAR STANDARDS