International Review of Law and Economics 24 (2004) 241–254

# On the definition and efficiency of punitive damages

C.Y. Cyrus Chu[a,*], Chen-Ying Huang[b]

[a] *Institute of Economics, Academia Sinica, 128 Academia Road, Sec. 2, Nankang, Taipei, Taiwan*
[b] *National Taiwan University, Taiwan*

**Abstract**

We construct a model of optimal deterrence to address the issue concerning the form and efficiency of punitive damages. Because the same amount of damages means less to wealthy injurers than to poor ones, without punitive damages, it might happen that poor injurers abide by the law while wealthy injurers do not. Thus the court can increase social welfare by choosing its tolerance level and imposing punitive damages on injurers for any precautionary effort below this tolerance level. The use of tolerance together with punitive damages accords well with the empirical observation that punitive damages are normally awarded only for "outrageous" misconduct. We show that the optimal negligence standard, the tolerance level and the punitive damages crucially depend on the wealth distribution in the society. This explains that when juries are instructed to relate the awarded damages to an outrageous injurer's wealth, it may be an efficient design rather than a pure prejudice against people's wealth. Some comparative statics results are also provided.


*Keywords:* Avoidance actions; Punitive damages; Punishment

## 1. Introduction

Perhaps due to the increasing controversy concerning the imposition of punitive damages in tort cases, in the past 20 years there has been an increasing number of articles discussing the rationale, qualifications, reasonable amount, and limitations of punitive damages. When the focus of discussion is on the *optimal deterrence* that damages bring about, one conventional argument is the following. Suppose (1) we refer to punitive damages as the amount in

* Corresponding author.
*E-mail address:* cyruschu@sinica.edu.tw (C.Y.C. Chu).

0144-8188/$ – see front matter 
doi:10.1016/j.irle.2004.08.008

addition to compensatory damages; (2) there are no illicit gains attributed to the injurer; and (3) it is costless to identify the injurer, bring suit, and collect full damages. Then, punitive damages would not be necessary.[1]

Many previous theoretical contributions in fact concentrate on relaxing supposition (3) listed above from various angles. Specifically, observing that very often the harm caused by a tort is related to some other variables, several economists were able to find the positive impact of punitive damages on changing these variables. For instance, Biggar (1995) argues that if punitive damages are in force, victims can be ensured to be compensated. Thus they will not be induced to undertake some "avoidance actions," which may reduce efficiency. Daughety and Reinganum (1997) show that when punitive damages are absent, there is a broader range of parameters where no separating (but possibly efficient) equilibrium exists. More recently, Boyd and Ingberman (1999) show that punitive damages may cause a firm to reduce the amount of wealth exposed to liabilities and blunt the incentives to invest in safety. Finally, Png (1987) and Spier (1997) study the function of awarding punitive damages on changing the suit-settlement decision of the victims, thereby serving the purpose of reducing ex ante social costs.

An alternative view of punitive damages is to treat them as a penalty designed to *completely deter* potential offenders by removing any prospect of gain.[2] Under this view, punitive damages are essentially *sanctions* instead of *prices*, in terms of the distinction pointed out by Cooter (1984). We will come back to this point later.

Despite the fact that the discussion on punitive damages has been broad, there are three practices observed in reality which, to our knowledge, have not been satisfactorily addressed. The first observation is that punitive damages are normally awarded only for "outrageous, malicious, unusual, or wanton" misconduct.[3] In the related literature of the economic analysis of law, however, we have not seen a satisfactory definition of punitive damages which matches well with the outrageousness of an injurer's conduct. A second and related point missing in the literature is the connection between the *behavioral criterion* of outrageousness and the objective measurement of the amount of compensatory damages. According to the conventional argument, punitive damages should be roughly equal to the product of a multiplier, presumably reflecting how difficult to make an injurer compensate the victim, and the amount of compensatory damages. The multiplier is used to adjust for the low probability of sentencing an injurer's liability, reflecting an objective friction in the litigation system. This conflicts with the observation that for different cases, different criteria for judging whether conduct is outrageous are often used. Moreover, there seems to be no common-to-all-cases multiplier that exists for awarding punitive damages. The third observation concerns whether it is efficient to allow punitive damages to depend on the defendant's wealth. Jury instructions in many states of the US have stated explicitly that the defendant's wealth can be taken into account in determining punitive damages.[4] However,

[1] For a detailed discussion, see Cooter (1982), Polinsky and Shavell (1998) (p. 891) and Sunstein, Kahneman, and Schkade (1998) (p. 2082), and the references therein.

[2] See Hylton (1998) for details.

[3] For instance, see Ashland Dry Goods Co. versus Wages, 302 Ky. 577, 582–583, 195 S.W. 2nd 312, 315 (1946) (Kentucky Court of Appeals), or the model jury instruction in Sunstein et al. (1998), p. 2147.

[4] See for instance, the model jury instructions of Arkansas, California, and Wisconsin.

there has not been any theoretical explanation of how this instruction could possibly be justified on the ground of efficiency.[5]

The purpose of this paper is to construct a model of optimal deterrence to address these three observations. We consider a setup with four groups of people: poor injurers and their victims, and wealthy injurers and their victims. The liability criterion is assumed to be the negligence rule, where a common negligence standard applies to everyone.[6] Facing a negligence rule, potential injurers make their decisions of precautionary effort, determining their respective probability of causing an accident. Taking this into account, the court chooses a socially optimal negligence standard to maximize social welfare.

When the precautionary cost does not decrease too much with wealth, wealthy injurers rationally choose to ignore the negligence standard while poor ones abide by it. This is because paying the small amount of compensatory damages simply does not matter much to wealthy injurers. But this is not ideal, since the joint welfare of wealthy injurers and their victims may be improved if the former take more care.[7] When the socially optimal care level of wealthy injurers is lower than the negligence standard, the court should "tolerate" any effort in-between, since it implies that the wealthy injurers have taken enough care. On the other hand, any effort below this optimal care level should then be defined as "outrageous."[8] For such an outrageous effort, a large enough amount of punitive damages should be levied, as such punitive damages serve to make wealthy injurers find it worthwhile to undertake more effort.

The three observations we raised can now be re-evaluated. First, the use of tolerance together with punitive damages in the above construction accords well with the observation that punitive damages are awarded only for outrageous behaviors.[9] Second, since the court's tolerance (and punitive damages) depends on the difference between the socially optimal negligence standard and the optimal effort level of wealthy injurers, it is no wonder that punitive damages and tolerance should be judged case by case instead of finding a common-to-all-cases multiplier. Third, as in any social optimization problem, the wealth distribution of all four groups of people in the society matters in determining the socially optimal negligence standard and the optimal effort level of wealthy injurers. This helps explain why in many jury instructions, the defendant's wealth can be taken into account in determining the amount of punitive damages.

[5] Cooter (1982) provides an argument similar to ours; he points out that punitive damages should be imposed to offset the injurers' illicit pleasure, which may lead them to exercise a care level far below the legal threshold. Although our argument differs from Cooter's in that we aim to provide a positive account of important features of the law on punitive damages, we should note that there is a key formal similarity between the two models. Specifically, Cooter notes that his model applies to instances in which the injurer enjoys an illicit pleasure from being careless or experiences an exceptional cost from taking care. In our model, the injurer's wealth per se may cause him to experience such a exceptional cost of taking care.

[6] This is the same as the general practices. See Polinsky and Shavell (1998), p. 910.

[7] This is because higher efforts, to the first order, do not affect the wealthy injurers' utility but they reduce the probability of causing an accident.

[8] This definition is also consistent with Cooter's (1989) criterion of "incentive inadequacy."

[9] A similar inquiry is raised by Biggar (1995) who argues that, if punitive damages are merely a scale-up of compensatory damages, then the level of liability should be scaled up altogether, instead of restricted to cases with grossly negligent or outrageous injurers. Arlen (1992) notices the importance of the risk-aversiveness of injurers, but the discrepancy between the injurer's reprehensibility and the size of punitive damages is not completely addressed.

We hasten to clarify the link between punitive damages and wealth. The wealthy are not punished more heavily simply because they are wealthy; they are punished more heavily because their wealth leads them to undertake less effort. That is, punitive damages depend positively on convicted injurers' wealth because the former depends on injurers' precautionary effort, which in turn, depends on the latter.[10] Our whole argument provides a mere *theoretical* possibility of justifying the current practices. Its *empirical* persuasiveness hinges upon the validity of the assumptions behind our theory. More related work along this line is needed in the future.

Polinsky and Shavell (1998) (p. 911) argue that relating the size of damages to the injurers' wealth may cause the injurers to take excessive precaution. In this paper, since the damages amount is calculated to maximize social welfare, "excessive" precaution is by definition ruled out as suboptimal. Many authors have objected to the connection of the level of damages and the defendant's wealth. For instance, Abraham and Jeffries (1989) argue correctly that punishment should depend on what the defendant has done, not on who the defendant is. They also argue that if more damages are needed due to inadequate deterrence, then the cure "is to raise the expected costs to all defendants, not just to those with greater wealth" (p. 418). The key reason why we arrive at a different result is because we consider a model where potential injurers have different wealth levels. The different wealth levels per se are not important, but the difference in the resulting wealth-specific, optimally chosen efforts makes the problem of optimal deterrence more complex. Finally, we should emphasize that our analysis applies only to *individual* injurers; if punitive damages on *corporations* are in question, controversial arguments concerning the principal–agent relationship within corporations are unavoidable.

The rest of this paper is arranged as follows. Section two presents the basic setup and potential injurers' decision framework. The third section discusses the socially optimal negligence standard. Section 4 introduces punitive damages and studies their efficiency-enhancing property. The last section summarizes some comparative statics and concludes.

## 2. The model

Consider a society composed of four groups of people: poor injurers and their victims, and wealthy injurers and their victims. Suppose a potential injurer can take some precaution $x \geq 0$ to prevent an accident from happening. If he chooses the precautionary level $x$, let $p(x)$ be the probability that an accident still happens. Since taking precaution requires effort, the cost when $x$ is chosen is denoted by $c(x, y)$, where $y \geq 0$ is his wealth level.[11] We make the following assumptions:

[10] The argument here is not dissimilar to the reasoning why imprisonment is needed in combination with fines to deter a group of potential offenders with different wealth. Since richer people have higher opportunity costs and are more afraid of imprisonment, an imprisonment term serves the purpose of deterring these rich people, despite the fact that exercising fines is economically more efficient. See Chu and Jiang (1993) for more details.

[11] Let us denote the differentiation by a prime, and partial derivatives by subscripts. For example, $c_1$ stands for the partial derivative with respect to the first argument in function $c$.

(A1) $p'(x) < 0$ and $p''(x) > 0$ for all $x$.
(A2) $c_1(x, y) > 0$ for all $x > 0$ and $y$.
(A3) $c_{11}(x, y) > 0$ for all $x$ and $y$.
(A4) $c_1(x, y) = 0$ when $x = 0$ for all $y$.

(A1) means that a higher precautionary effort reduces the probability of an accident, but its marginal effect is decreasing. (A2) and (A3) imply that efforts are costly and costs are strictly convex in precautionary efforts. (A4) implies zero marginal cost when no precaution is taken.

Notice that the cost function may depend on wealth. The sign of $c_{12}(x,y)$ determines whether it is more or less costly for the wealthy to take extra precaution. Intuitively, if taking precautions is time-consuming, the wealthy might find it very costly. On the other hand, if taking precautions involves only a monetary payment, the wealthy might be more likely to take precautionary measures. The dependence of the cost on wealth crucially determines the effectiveness of punitive damages, as we will see later.

### *2.1. The injurer's decision framework*

Assume that all potential injurers have an additively separable utility function of the following form:

$$U(x, y) = [p(x)bu(y - I \cdot A) + (1 - p(x)b)u(y)] - c(x, y). \tag{1}$$

The positive constant $b$ stands for the probability that an injurer is caught when an accident happens and $u(\cdot)$ is the utility function of the injurer. We assume:

(A5) $u'(y) > 0$ and $u''(y) < 0$ for all $y$.
(A6) $y > A$.[12]

Eq. (1) can be explained as follows. If the injurer is judged liable for the accident, then $A$ is the compensatory damages he has to pay to the victim. When the injurer is not liable for the accident, he can keep all his wealth $y$. An index I identifies whether the injurer is judged liable. Since the liability rule is a negligence rule, there exists a negligence standard $\bar{x}$ such that if the precaution effort is less than $\bar{x}$, then the injurer is judged negligent and thus liable. This implies:

$$I = \begin{cases} 1, & \text{if he is liable; i.e. } x < \bar{x} \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

The negligence standard $x$ is chosen by the court to maximize social welfare, which we shall discuss later.

[12] For cases where an injurer does not have enough wealth to pay the compensatory damages $A$, refer to Chu and Huang (2002).

The meaning of (1) is now clear: an injurer's wealth is $y - IA$ with probability $p(x)b$; his wealth is $y$ with the remaining probability $1 - p(x)b$. Undertaking effort $x$ entails cost $c(x, y)$.

### 2.2. *When the Injurer Is liable*

Suppose the injurer is liable and hence $I = 1$. The first-order condition of (1) with respect to precaution $x$ is:

$$U_1(x, y) = p'(x)b(u(y - A) - u(y)) - c_1(x, y) = 0. \tag{3}$$

The optimizer of $x$ is denoted by $x^*(y)$. Substituting it back into (1), we get the indirect utility for an injurer with wealth $y$ when he is judged liable. We denote this by $V^{\mathrm{l}}(y)$ so that:

$$V^{\mathrm{l}}(y) \equiv U(x^*(y), y), \tag{4}$$

where superscript l indicates that this is the liable case so that $I$ is 1 in (1).

A simple comparative statics analysis from (3) implies:

$$\frac{\mathrm{d}x^*(y)}{\mathrm{d}y} = \frac{-U_{12}(x^*(y), y)}{U_{11}(x^*(y), y)}. \tag{5}$$

It can be shown that the sign of $U_{12}(x^*(y), y)$ depends on that of $c_{12}(x^*(y), y)$*:* If $c_{12}(x^*(y), y)$ is not too negative, then $U_{12}(x^*(y), y)$ will be negative, implying a negative relationship between $y$ and $x^*(y)$. Intuitively, there are two effects in determining the relationship between $y$ and $x^*(y)$. The first effect, termed the wealth effect, exists because the damages $A$ mean less in utility terms to the wealthy than to the poor due to the strict concavity of the utility function. This implies the wealthy are less driven to prevent the accident. The second effect, termed the *cost effect*, captures whether it is more costly for the wealthy to take extra prevention. This is summarized in the cross derivative $c_{12}(x^*(y), y)$. If the wealthy have a cost disadvantage or a mild cost advantage, then the wealth effect will dominate the cost effect. This encourages the wealthy to choose less precaution than the poor. There might then be room for punitive damages because extra damages can make the wealthy increase their precautionary level. We summarize the result as follows:[13]

**Proposition 1.** *When $c_{12}(x, y) \geq 0$ or when $|c_{12}(x, y)|$ is small, the injurers' precautionary level is a decreasing function of their wealth if they are only liable for compensatory damages $A$.*

Since we aim at providing an efficiency argument for punitive damages, in the following we will restrict our attention to the case where Proposition 1 holds.[14] Therefore, we assume:

(A7) Either $c_{12}(x, y) \geq 0$ or $|c_{12}(x, y)|$ is small for all $x$ and $y$.

---

[13] See Chu and Huang (2002) for details of the proof.

[14] For cases where Proposition 1 does not hold, refer to Chu and Huang (2002) for discussions.

We now proceed to the injurer's decision of whether to comply with the negligence standard.

### 2.3. When the injurer is not liable

The analysis using the first-order condition in (3) is incomplete, because when $x^*(y) \geq \bar{x}$ (the negligence standard), the injurer is not liable. To take this into account, we first write down the injurer's utility when he is *not* liable, and then compare it with $V^{\mathrm{l}}(y)$ in (4).

When the injurer is not liable, $I = 0$ in (1). This happens only when he has taken a sufficient precautionary effort $x \geq \bar{x}$. Since $c_1(x, y) > 0$ by (A2), an injurer has no incentive to exert anything more than $\bar{x}$. Thus the injurer's indirect utility function in this case, denoted by $V^{\mathrm{nl}}(\bar{x}, y)$, is:

$$V^{\mathrm{nl}}(\bar{x}, y) = u(y) - c(\bar{x}, y), \tag{6}$$

where superscript nl indicates no liability. We now summarize some helpful properties in Proposition 2 below.

**Proposition 2.** (a) $V^{\mathrm{nl}}(x, y)$ *is a decreasing function of* $\bar{x}$; (b) *When* $I = 1$, $U(x, y)$ *is a concave function of x with the maximum at* $x^*(y)$; (c) $V^{\mathrm{nl}}(x^*(y), y) > V^{\mathrm{l}}(y)$.

Properties (a) and (b) are straightforward. Property (c) holds because if the negligence standard is exactly $x^*(y)$, the no-liability case involves no damages payment while the liability case involves damages payment. Since the precautionary efforts are the same ($x^*(y)$) in both cases, the utility in the former is higher.

### 2.4. The injurer's optimal choice

By Proposition 2, there exists a critical negligence standard, denoted by $\bar{x}(y) > x^*(y)$, such that (see Fig. 1):

$$V^{\mathrm{nl}}(\bar{x}(y), y) = V^{\mathrm{l}}(y). \tag{7}$$

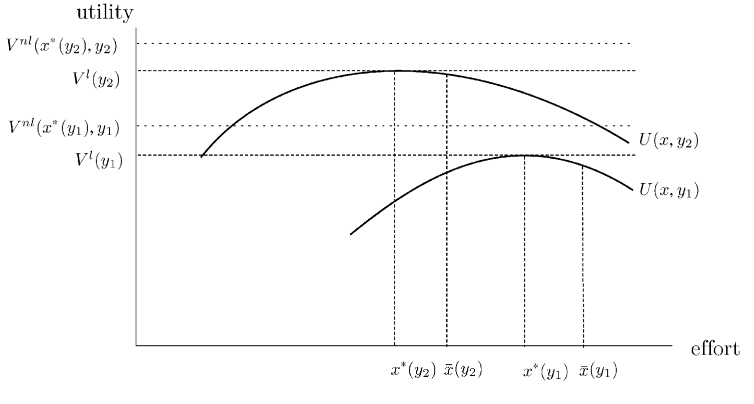


Fig. 1. The utility functions $U(x, y)$ of injurers when they are liable, i.e. $I = 1$.

For any negligence standard $\bar{x}$, the injurer will compare $V^{\mathrm{nl}}(\bar{x}, y)$ with $V^{\mathrm{l}}(y)$ to decide what he wants to do. Since $\bar{x}(y)$ is the critical negligence standard where the injurer is indifferent between abiding by the rule or ignoring it, if the actual standard $\bar{x}$ is set below it, the injurer will abide by the rule. On the other hand, when $\bar{x}$ is above it, the injurer will ignore the rule and choose $x^*(y)$. In other words, $\bar{x}(y)$ is *the highest possible precautionary effort of an injurer with wealth y*. Denote $x(\bar{x}, y)$ as *the precautionary effort by an injurer with wealth y when the negligence standard is* $\bar{x}$. The discussion above leads to the following result:

**Proposition 3.**

$$x(\bar{x}, y) = \begin{cases} x^*(y) \quad \text{if } \ \bar{x} > \bar{x}(y), \\ x^*(y) \ \text{ or } \ \bar{x} \quad \text{if } \ \bar{x} = \bar{x}(y), \\ \bar{x}, \quad \text{if } \ \bar{x} < \bar{x}(y). \end{cases}$$

By (A2), (A5), (A7) and the Envelope Theorem, a total differentiation of (7) yields:

$$\frac{\mathrm{d}\bar{x}(y)}{\mathrm{d}y} = \frac{V_1^{\mathrm{l}}(y) - V_2^{\mathrm{nl}}(\bar{x}(y), y)}{V_1^{\mathrm{nl}}(\bar{x}(y), y)} < 0. \tag{8}$$

This implies when wealth becomes higher, the critical negligence standard $\bar{x}(y)$ decreases.

In the following analysis we shall assume that there are two groups of potential injurers. Their wealth levels are $y_1$ and $y_2$. To be consistent with (A6), assume that $A < y_1 < y_2$. By (8), $\bar{x}(y_2) < \bar{x}(y_1)$. Therefore, if the negligence standard lies between $\bar{x}(y_2)$ and $\bar{x}(y_1)$, poor injurers (wealth $y_1$) will abide by the rule while wealthy injurers (wealth $y_2$) will ignore it. As we shall see, this is exactly why punitive damages are observed to be levied on the wealthy only. Fig. 1 helps visualize the idea of this section.

## 3. The socially optimal negligence standard

We now turn to the case for victims. A victim's utility depends on his wealth and the probability of an accident, which in turn relies on the precautionary effort by the injurer. When a victim faces an injurer who takes precautionary effort $x$, the victim's utility function is:

$$V(x, y^{\mathrm{v}}) = p(x)bv(y^{\mathrm{v}} - A + I \cdot A) + p(x)(1 - b)v(y^{\mathrm{v}} - A) + (1 - p(x))v(y^{\mathrm{v}}),$$

where $I$ is the index function in (2), $y^{\mathrm{v}}$ is the victim's wealth and $v(\cdot)$ is the victim's utility function. The victim's expected utility function takes this form because with probability $p(x)b$, an accident happens and the injurer is caught. In this event the victim's wealth is dependent on whether the injurer is judged liable. With probability $p(x)(1 - b)$, an accident happens but the injurer is not caught. In this case the victim's wealth falls by $A$. With the remaining probability $(1 - p(x))$, there is no accident and the potential victim keeps all his wealth $y^{\mathrm{v}}$. The utility function of the victims is assumed to satisfy the standard assumptions:

(A8) $v'(y^{\mathrm{v}}) > 0$ and $v''(y^{\mathrm{v}}) < 0$ for all $y^{\mathrm{v}}$.

As mentioned in Section 1, we assume that there are two groups of victims. Group 1 consists of victims of potential injurers with wealth $y_1$. The wealth level of this group is denoted by $y_1^{\mathrm{v}}$. Group 2 consists of victims of potential injurers with wealth $y_2$. The wealth level of this group is denoted by $y_2^{\mathrm{v}}$.

Taking all four groups into account, the court has a social welfare function of the following form:

$$W(\bar{x}) = U(x(\bar{x}, y_1), y_1) + U(x(\bar{x}, y_2), y_2) + V(x(\bar{x}, y_1), y_1^{\mathrm{v}}) + V(x(\bar{x}, y_2), y_2^{\mathrm{v}}), \quad (9)$$

and it chooses an optimal negligence standard $\bar{x}$ to maximize the social welfare $W(\bar{x})$. Denote the optimum by $\bar{x}^*$.[15] Intuitively, if the court can raise $\bar{x}$ effectively, the victims benefit because the probability of an accident decreases. However, the injurers are worse off because they need to undertake more precautionary efforts. The court hence needs to balance the marginal benefit and marginal cost carefully. This is further complicated by the fact that there is a highest negligence standard that the injurers will pay attention to. For wealthy injurers, it is $\bar{x}(y_2)$, and for poor injurers, it is $\bar{x}(y_1)$. If the negligence is higher than what the injurers will pay attention to, any further increase has no effect on the injurers' behavior. Since by (8) $\bar{x}(y_2) < \bar{x}(y)$, we have:

**Lemma 4.** *If the negligence standard is lower than $\bar{x}(y_2)$, then both injurers will abide by the rule. If it lies between $\bar{x}(y_2)$ and $\bar{x}(y_1)$, only poor injurers will abide by the rule; wealthy injurers will choose $x^*(y_2)$ instead. If it is higher than $\bar{x}(y_1)$, then neither will abide by the rule: wealthy injurers will choose $x^*(y_2)$ and poor injurers will choose $x^*(y_1)$. The optimal negligence standard cannot exceed $x(y_1)$ since any further increase has no effect.*

## 4. Introducing punitive damages

In this section we shall argue that: (1) there are situations where imposing punitive damages increases social welfare; (2) more importantly, punitive damages are levied only on wealthy injurers and are related to their wealth as a result. These two results together provide an efficiency rationale for the jury instructions in many states where they state explicitly that the defendant's wealth can be taken into account in determining punitive damages.[16]

To facilitate the analysis, we define two ancillary social welfare functions:

$$W_1(\bar{x}) = U(x((\bar{x}, y_1), y_1)) + V(x(\bar{x}, y_1), y_1^{\mathrm{v}}), \quad (10)$$

and

$$W_2(\bar{x}) = U(x((\bar{x}, y_2), y_2)) + V(x(\bar{x}, y_2), y_2^{\mathrm{v}}). \quad (11)$$

[15] A priori, there is not too much we can say about the optimal negligence standard because it depends on the distribution of wealth of the victims $(y_1^{\mathrm{v}}, y_2^{\mathrm{v}})$ relative to that of the injurers $(y_1, y_2)$. The risk aversion of the victims, captured by the concavity of $v(\cdot)$, and that of the injurers, reflected by the concavity of $u(\cdot)$, also play a role.

[16] Refer to footnote 4.

$W_1(\bar{x})$ is the overall welfare of poor injurers and their victims, and $W_2(\bar{x})$ is that of wealthy injurers and their victims. We denote the negligence standard optimizing $W_1$ by $\bar{x}_1^*$ and that optimizing $W_2$ by $\bar{x}_2^*$.

The reason why punitive damages might improve social welfare is as follows. If the optimal negligence standard $\bar{x}^*$ turns out to lie above the highest standard which wealthy injurers can bear, they will optimally choose insufficient care. Since victims always benefit from an increase in precautionary efforts, if the court levies additional punishment, the wealthy injurers are induced to increase their efforts. If so, the social welfare will increase because to the first order, the wealthy injurers are not affected. This is exactly the function that punitive damages serve: to bring the efforts of wealthy injurers more into line. However, the court does not want wealthy injurers to choose an excessively high effort either. Since increasing the effort further has a cost, the marginal benefit of increasing the victims' welfare must be balanced against the marginal cost of increasing the injurers' effort. Thus there exists an optimal effort level that the court wants the wealthy to choose.

To summarize, the court wants to tolerate the wealthy injurer's insufficient effort only to some extent. If their effort is below this tolerance, it uses punitive damages to induce them to increase their effort. The existence of a tolerance level matches well with the observation that punitive damages are only levied for outrageous behaviors.

### *4.1. Formalizing the results*

As mentioned above, if the optimal negligence standard $\bar{x}^*$ is greater than the highest standard which wealthy injurers can bear, $\bar{x}(y_2)$, then only poor injurers will abide by the rule as long as $\bar{x}^* < \bar{x}(y_1)$. Wealthy injurers will choose to ignore it and their precautionary effort is $\bar{x}^*\ (y_2)$. In this case, the overall welfare of wealthy injurers and their victims is:

$$W_2(\bar{x}^*) = u(x^*(y_2), y_2) + V(x^*(y_2), y_2^{\mathrm{v}}).$$

Since

$$x^*(y_2) = \arg \max_{x,\, I=1} U(x, y_2),$$

if the precautionary effort of wealthy injurers is increased marginally, to the first order, their welfare is not affected. However, because

$$V_1(x^*(y_2), y_2^{\mathrm{v}}) = P'(x^*(y_2))(1-b)[v(y_2^{\mathrm{v}} - A) - v(y_2^{\mathrm{v}})] > 0,$$

the welfare of their victims increases. Thus if the court can raise the precautionary effort of wealthy injurers without affecting that of poor injurers, then the overall social welfare must be improved. As we will show, punitive damages can achieve this goal.

By how much does the court want wealthy injurers to increase their precautionary effort? We denote $x'$ as:

$$x' = \arg \max_{x \le \bar{x}^*} [U(x, y_2) + V(x, y_2^{\mathrm{v}})]. \tag{12}$$

That is, $x'$ is the precautionary effort which maximizes the joint welfare of wealthy injurers and their victims (subject to the constraint that we cannot ask the wealthy injurers to undertake more efforts than the negligence standard $\overline{x}^*$). Ideally, the court wants to make the wealthy injurers increase their effort from $\overline{x}^*(y_2)$ to $x'$.

We can now formally introduce punitive damages. An injurer's behavior is said to be *outrageous* if, when an accident happens, the precaution he takes turns out to be *less than a certain fraction of the negligence standard*. Let $\epsilon$ denote the tolerance fraction. Since $\overline{x}^*$ is the optimal negligence standard, any precaution less than $(1-\epsilon)\overline{x}^*$ would be considered outrageous.

Suppose the court mandates the following: if the injurer is only negligent, i.e. his precaution level lies between $(1-\epsilon)\overline{x}^*$ and $\overline{x}^*$, he $\overline{x}$ is liable for the compensatory damages, $A$. On the other hand, if the injurer is not only negligent but his behavior is also regarded as outrageous, i.e. his precaution level is less than $(1-\epsilon)\overline{x}^*$, then he is $\overline{x}$ liable for damages $A$ plus punitive damages $B$. The question is how should the court select $\epsilon$ and $B$?

The tolerance level serves to create a downward jump in the wealthy injurers' utility when their effort declines to $(1-\epsilon)\overline{x}^*$, $\overline{x}$ because when it starts to be below $(1-\epsilon)\overline{x}^*$, they immediately face punishment $A + B$ instead of only $A$. When $B$ is large enough, to avoid $B$, wealthy injurers will choose $(1-\epsilon)\overline{x}^*$).[17] If the $\overline{x}$ court chooses the tolerance level $\epsilon$ so that:

$$(1-)\overline{x}^* = x', \tag{13}$$

the joint social welfare of wealthy injurers and their victims is maximized. To make sure that $B$ is large enough, the court needs to punish heavily for any effort lower than $(1-\epsilon)\overline{x}^*$ so that wealthy injurers find this not worthwhile. That is, $B$ has to be large enough so that for all $x < x'$ (see Fig. 2):[18]

$$\begin{aligned}[p(x)bu(y_2 - A - B) + (1 - p(x)b)u(y_2)] - c(x, y_2) \leq [p(x')bu(y_2 - A)\\ + (1 - p(x')b)u(y_2)] - c(x', y_2).\end{aligned} \tag{14}$$

We now summarize the main result of the paper.

**Proposition 5.** *Suppose the optimal negligence standard* $\overline{x}^*$ *is greater than the highest standard the wealthy injurers can bear* $(\overline{x}(y_2))$. *If the effort which maximizes the joint welfare of wealthy injurers and their victims* $(x')$ *is strictly less than* $\overline{x}^*$, *the court can select a tolerance fraction* $\epsilon$ *and make punitive damages B large enough so that the joint social welfare of them is maximized. The pair* $(\epsilon, B)$ *has to satisfy* (13) *and* (14).

Although we impose the use of tolerance and punitive damages to show that the joint welfare of the wealthy injurers and their victims is maximized in Proposition 5, we can

[17] They will not choose any effort higher than $(1-\epsilon)\,\overline{x}^*$. This is because $U(x, y_2)$ is concave in $x$ (when $x < \overline{x}^*$) and reaches its maximum at $x^*(y_2) < \overline{x}^*$. Thus, the farther away it is from $x^*(y_2)$, the lower the utility will be.

[18] This might not be possible if levying the heaviest punishment $y_2 - A$ *is* not enough. In this case, requiring wealthy injurers to choose $x'$ is too strict. Although implementing $x'$ is not possible, imposing additional punishment $y_2 - A$ still serves to make them increase their effort level, though not to $x'$. In this case the joint social welfare of wealthy injurers and their victims is still improved. Considering this extra complication does not affect the result.

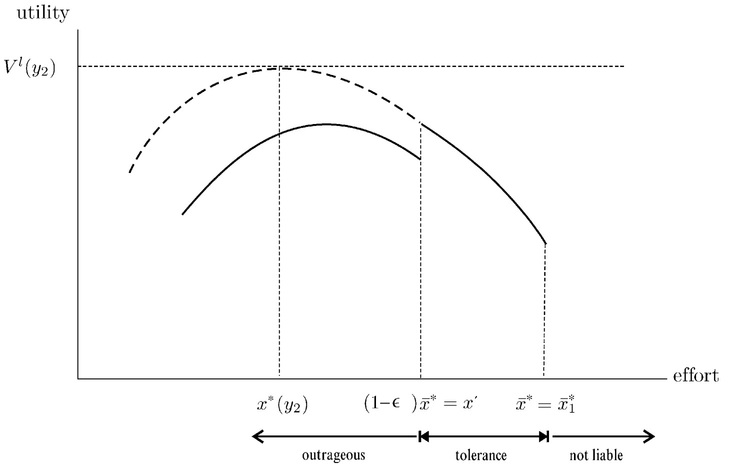


Fig. 2. The use of negligence standard, tolerance, and punitive damages. The negligence standard is set as $\bar{x}^*$. For any effort in the tolerance range, the injurer is liable for the compensatory damages A. For any effort below the tolerance, the injurer is not only negligent but also considered outrageous. Punitive damages B are made large enough so that the wealthy injurer will choose to undertake $(1-\epsilon)\bar{x}^*$, just within the tolerance. The solid line stands for the wealthy injurer's utility where punitive damages are in effect. The discrete jump in his utility when his effort reaches $(1-\epsilon)\bar{x}^*$, makes sure that he will optimally choose to undertake $(1-\epsilon)\bar{x}^*$.

nevertheless interpret the result the other way around. Suppose we instead consider an arbitrary but optimal form of punishment to maximize the joint social welfare of wealthy injurers and their victims conditional on a chosen negligence standard. On the one hand, we do not want the wealthy injurers to undertake any effort above $x'$. Thus for efforts between $x'$ and $\bar{x}^*$, we cannot levy any additional punishment in any optimal design. On the other hand, we want them to choose exactly $x'$. Hence the punishment has to be heavier than $A$ for any effort below $x'$. This implies all forms of optimal deterrence share a common feature: *there has to be additional punishment for any effort below $x'$, and no additional punishment between $x'$and$\bar{x}^*$*. Thus, our interpretation further strengthens the generality of the use of tolerance and punitive damages, since all optimal forms come with the use of tolerance and additional punishment anyway.

A corollary of Proposition 5 is summarized as follows.

**Corollary 6.** *If the socially optimal negligence standard $\bar{x}^*$ is greater than the highest standard which wealthy injurers can bear $(\bar{x}(y_2))$ stated in Proposition 5, the socially optimal negligence standard must maximize the joint social welfare of poor injurers and their victims, $W_1(\bar{x})$. That is, $\bar{x}^* = \bar{x}_1^*$.*

By Corollary 6, when punitive damages are effective, the socially optimal negligence standard serves to maximize the joint welfare of *poor* injurers and their victims only. It is as if it does not apply to wealthy injurers since they will ignore it anyway. This is because, as long as the negligence standard is above $\bar{x}(y_2)$, wealthy injurers will ignore the standard no matter what the standard is. On the other hand, punitive damages and the tolerance fraction serve to maximize the joint welfare of *wealthy* injurers and their victims. It is as if they do not apply to poor injurers since they will never be judged punitively liable anyway.

## 5. Extensions and conclusions

Expressions (10)–(14) allow us to implement some comparative statics regarding the socially optimal negligence standard, the tolerance fraction, and the size of punitive damages. We list the results below and leave details to Chu and Huang (2002). (a) When the wealth of poor injurers is smaller, the optimal negligence standard for poor injurers and their victims is higher, and the tolerance fraction $\epsilon$ is higher. (b) Decreasing the wealth of the victims of poor injurers has a similar effect as in (a). (c)When the wealth of the victims of wealthy injurers is smaller, the optimal effort level for wealthy injurers ($x'$) becomes higher, the tolerance fraction becomes smaller, and the punitive damages become higher. (d) When the wealthy injurers become richer, the optimal effort level $x'$ for them becomes smaller, and the tolerance fraction becomes larger.

In summary, our analysis indicates that the design of punitive damages improves social welfare when the optimal negligence standard is higher than what wealthy injurers can bear. In this situation, the socially optimal negligence standard maximizes the joint welfare of poor injurers and their victims. Additional punishment – in the form of tolerance fraction and punitive damages – serves to bring wealthy injurers' effort more into line. The tolerance fraction exactly reflects the court's optimal effort level of wealthy injurers. Punitive damages are used to provide incentives for wealthy injurers so that it is not worthwhile for them to undertake any effort below the court's "tolerance." In this respect, our theory stays within the *optimal deterrence* framework, and is distinct from the *complete deterrence* hypothesis of Hylton (1998).

According well with the intuitions from the literature on efficient liability, injurers find it worthwhile to take precaution rather than bear the expected liability. Poor injurers choose to abide by the negligence standard. Thus compensatory damages only fall on wealthy injurers. Wealthy injurers choose to undertake the effort level just within the court's tolerance.

## Acknowledgement

Insightful comments from an anonymous referee are gratefully acknowledged.

## References

Abraham, Kenneth S., & Jeffries, John C., Jr. (1989). Punitive damages and rule of law: The role of defendant's wealth. *Journal of Legal Studies*, *18*, 415–425.

Arlen, Jennifer H. (1992). Should defendants' wealth matter? Journal of Legal Studies. *Journal of Legal Studies*, *21*, 413–429.

Biggar, Darryl. (1995). A model of punitive damages in tort. *International Review of Law and Economics*, *15*, 1–24.

Boyd, James, & Ingberman, Daniel E. (1999). Do punitive damages promote deterrence? *International Review of Law and Economics*, *19*, 47–68.

Chu, C. Y. Cyrus, & Huang, Chen-Ying. (2002). On the definition and efficiency of punitive damages. Working Paper (web site http://www.econ.ntu.edu.tw/public/discussionpaper/ntudp2002.htm).

Chu, C. Y. Cyrus Chu, & Neville, Jiang. (1993). Are fines more efficient than imprisonment. *Journal of Public Economics*, *51*, 391–413.

Cooter, Robert D. (1982). Economic analysis of punitive damages. *Southern California Law Review*, *56*, 79–101.

Cooter, Robert D. (1984). Prices and sanctions. *Columbia Law Review*, *84*, 1523–1560.

Cooter, Robert D. (1989). Punitive damages for deterrence: When and how much? *Alabama Law Review*, *40*, 1143–1196.

Daughety, Andrew F., & Jennifer, Reinganum. (1997). Everybody out of the pool: Products liability, punitive damages, and competition. *Journal of Law Economics and Organization*, *13*, 410–432.

Hylton, Keith N. (1998). Punitive damages and the economic theory of penalties. *Georgetown Law Journal*, *87*, 421–467.

Png, Ivan. (1987). Litigation, liability, and the incentives for care. *Journal of Public Economics*, *34*, 61–85.

Polinsky, A. Mitchell, & Shavell, Steven. (1998). Punitive damages: An economic analysis. *Harvard Law Review*, *111*, 870–962.

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

Sunstein, Cass R., Kahneman, Daniel, & Schkade, David. (1998). Assessing punitive damages. *The Yale Law Review*, *7*, 2071–2153.