



Do Punitive Damages Promote Deterrence?

JAMES BOYD

Resources for the Future, Washington, DC, USA

and

DANIEL E. INGBERMAN

LECG, Inc., Emeryville, CA, USA

The paper explores the deterrent effect of punitive damages when capital investment is endogenous to the legal rule. Irrespective of the level of damages, firms can reduce their expected liabilities by either increasing safety (the law's intended goal) or reducing wealth exposed to liability. We show that punitive damages can exacerbate this wealth reduction effect and thereby reduce deterrence. The paper explores policy implications and the conditions under which this deterrence-reducing effect of punitive damages arises. © 1999 by Elsevier Science Inc.

I. Introduction

Punitive damages are intended to increase deterrence when the enforcement of law is otherwise weak or to discourage activities that unambiguously lack social value. However, the practical application of punitive damages law is widely viewed as flawed. Few objective criteria discipline the scale of damages or the types of behavior that trigger punitive liability.¹ The result is uncertainty for potential defendants and a widely held

We thank Timothy Brennan, Don Coursey, Dorsey Ellis, Ronald King, Gary Miller, seminar participants at the University of Pennsylvania and Washington University and the editors and anonymous referees.

¹Both the appropriate scale of damages and the standards that trigger them are loosely defined in the law. See Dorsey Ellis, "Fairness and Efficiency in the Law of Punitive Damages," 56 *Southern California Law Review* (1982) for an analysis that emphasizes the vague guidance given to juries in punitive liability cases. The Supreme Court in *Pacific Mutual Life Insurance Co. v. Haslip* (499 U.S. 1; 111 S. Ct. 1032, 1991) and *TXO Production Corp. v. Alliance Resources* (113 S.Ct. 2711, 1993) has recently upheld the constitutionality of wealth-sensitive damages and broad jury-judge discretion in setting awards.

The uncertainty surrounding awards is also emphasized by Supreme Court Justice O'Connor's dissent in *Pacific Mutual v. Haslip*: "... I would hold that Alabama's common-law punitive damages scheme is void for vagueness ... Without imposing any legislative or common-law limits, Alabama authorizes juries to levy civil fines ranging from zero to tens of millions of dollars. Indeed, a jury would not exceed its discretion under state law by imposing an award of punitive damages that was deliberately calculated to bankrupt the defendant. Unlike compensatory damages, which are tied to an actual injury, there is no objective standard that limits the amount of punitive damages."

suspicion that damages may sometimes be excessive. The possibility of excessive damages, of course, may also lead to overdeterrence. Indeed, recent public policy debates have highlighted overdeterrence as a primary rationale for proposed reforms (such as caps on punitive damage awards).²

This paper explores the deterrent effect of punitive damages. By contrast with most of the existing literature, we let capital investment—an important determinant of deterrence—be made in anticipation of liability. Our analysis challenges several common perceptions regarding the effects of punitive damages law. For instance, it is generally believed that “more punitive” liability increases deterrence, as potential defendants increase safety to reduce risk or to avoid being found punitively liable.³ We show, however, that more punitive liability can in fact lead to less deterrence. Thus, punitive liability law may be in need of reform, not because it leads to excessive safety or product withdrawals, but because it motivates less safety than other forms of liability.

More punishment can lead to less deterrence for a simple reason. Although firms might increase safety expenditures to reduce expected liabilities, they might also reduce the amount of wealth or capital they expose to those liabilities. Wealth reduction externalizes costs by truncating the set of liabilities the firm actually bears and thereby blunts the incentive to invest in safety. In some cases this wealth reduction effect can produce a net decrease in deterrence as damages become more punitive. This is especially likely to be true when punitive damages are conditioned on a defendant’s wealth.

The Literature on Damages, Deterrence, and Wealth

The connection between damages and capital choice has been neglected in analyses of deterrence.⁴ However, the economic literature on tort liability does emphasize the relationship between deterrence and defendants’ wealth. Two conclusions from this literature serve as the starting point for our analysis.

First, the possibility of bankruptcy implies underdeterrence. Because liability is lim-

²A common academic and popular view is that punitive damages “overdeter” potential defendants, leading to excessive investments in preventive measures or the removal of products from the market. This reasoning underlies frequent calls in Congress for punitive liability reform (most recently in the 104th Congresses’ “Contract with America”). See Alfred Cortese and Kathleen Blaner, “The Anti-Competitive Impact of U.S. Product Liability Laws: Are Foreign Businesses Beating Us at Our Own Game,” 9 *Journal of Law and Commerce*, 167 (1989) and Robert Litan and Clifford Winston, *Liability: Perspectives and Policy* (1988) for arguments in favor of capped damage awards.

Also, if the standard that triggers punitive damages is uncertain, overcompliance is likely as potential defendants attempt to insure against being punitively liable. See Richard Craswell and John Calfee, “Deterrence and Uncertain Legal Standards,” 2 *Journal of Law, Economics and Organization* (1986).

³This reasoning underlies the “rule of the reciprocal” in which punitive damages are used to counteract imperfect enforcement probabilities as in Robert Cooter, “Punitive Damages and Deterrence: When and How Much?,” 40 *Alabama Law Review* 1143, (1989). Or see Jason Johnston, “Punitive Liability: A New Paradigm of Efficiency in Tort Law,” 87 *Columbia Law Review* 1385 (1987) who argues that the threat of punitive damages, combined with procedural safeguards, should be used to prevent underdeterrence. Similarly, Ronald Dye, “Auditing Standards, Legal Liability, and Auditor Wealth,” 101 *Journal of Political Economy* 887, (1993) argues that deterrence increases with auditors’ wealth, and deterrence is increased by making care-conditioned punitive damages as large as possible.

⁴See, however, Steven Wiggins and Al Ringleb, “Adverse Selection and Long-Term Hazards: The Choice Between Contract and Mandatory Liability Rules,” 21 *Journal of Legal Studies* 189, (1992) and Ringleb and Wiggins, “Liability and Large-Scale, Long-Term Hazards,” 98 *Journal of Political Economy* 574, (1990) for a theoretical and empirical analysis, respectively, of the effects of liability on capital choice. In their model, potentially insolvent producers can substitute labor for capital to reduce their expected liabilities. They demonstrate empirically a shift over time toward small firm production of risk-generating products, which roughly correlates to the shift from negligence to strict products liability. However, they do not explore the effect of punitive damages on deterrence.

ited by the defendant's ability to pay, a potential defendant knows it is "judgment-proof" for any realizations of tortious losses that exceed its wealth. Thus, when making *ex ante* safety choices, a firm knows that its future liabilities are bounded by its legally recoverable capital value. This externalization of tort costs weakens the motivation to invest in safety. If damages are limited to the compensation of loss, strict liability underdeters those defendants whose wealth falls short of their potential liability.⁵

Second, conditioning damages on care can help deter potentially insolvent defendants. By absolving defendants of liability when they comply with a standard of care, negligence-based rules can lead some undercapitalized firms to higher levels of safety than they would make if they were strictly liable.⁶ Punitive damages law creates a rule similar to ordinary negligence, in that the level of damages is conditioned on care. Typically, punitive damages are imposed on defendants found to be grossly negligent, reckless, or to have engaged in an intentional tort. Firms that comply with these standards are liable for only compensatory damages. As the penalty for noncompliance increases, so too does the incentive to comply.

Overview of Our Results

The above conclusions hinge on the insight that deterrence is a function of the wealth that potential defendants expose to liability. However, these results are derived from the assumption that wealth does not vary with the legal rule. By treating capital investments as a choice variable made in anticipation of liability, our analysis provides new insights into the relationship between the level of damages and their effect on deterrence.

The outline of the analysis is as follows. Section II describes the elements of the model; the Appendix provides a more formal treatment. Section III then analyzes the effects of purely compensatory damages on firms' capital and safety choices. This analysis is essential to understanding the distorting effects of damages on capital investment and safety.

Section IV turns to the effects of punitive damages on capital investments and safety. As we show, more punitive damages can increase the wealth-reducing effect of potential liability. Because firms' incentives to invest in safety increase with the amount of capital they have to lose, increased damages can reduce safety by discouraging firms from investing in capital. Thus, when punitive damages reduce deterrence, they also distort capital investment and reduce productive efficiency.

Section V presents a parametric example of safety and capital investment in response to varying levels of damages. The example is designed to evaluate the effects of existing damages law and proposed tort reforms. For realism, our parameterization employs a description of current punitive liability law with the following properties: punitive damages that are uncertain and increasing in wealth.⁷ We compare capital and safety

⁵For analyses that have explored or employ this reasoning see Steven Shavell, "A Model of the Optimal Use of Liability and Safety Regulation," 15 *Rand Journal of Economics* 271 (1984); Shavell, "The Judgment-Proof Problem," 6 *International Review of Law and Economics* 45 (1986); Alan Schwartz, "Products Liability, Corporate Structure, and Bankruptcy: Toxic Substances and the Remote Risk Relationship," 14 *Journal of Legal Studies* 689 (1985); T. Randolph Beard, "Bankruptcy and Care Choice," 21 *Rand Journal of Economics* 626 (1990); and Lewis Kornhauser and Richard Revesz, "Apportioning Damages Among Potentially Insolvent Actors," 19 *Journal of Legal Studies* 617 (1990).

⁶See Shavell, "The Judgment-Proof Problem," William Landes and Richard Posner, "The Economic Structure of Tort Law," 1987, at 259, and Landes, "Insolvency and Joint Torts: A Comment," 19 *Journal of Legal Studies* 679 (1990).

⁷Wealth-sensitive damages are a time-honored aspect of punitive damages law. The practice is generally justified by

decisions made under this type of damage rule to (1) compensatory damages and (2) capped punitive damages, one of the most prominent reform proposals.

Beyond adding realism to the analysis, the example highlights an important property that is especially relevant in evaluating potential reforms. The net impact of punitive damages on deterrence depends on the productive benefits that firms obtain from capital investment. Firms that place the least value on such capital investment are the least deterred by liability generally. And it is these firms whose capital and safety investments are most likely to be worsened by punitive damages. As a result, punitive damages can worsen the problem they are intended to solve. Instead of furthering a retributive goal, punitive damages may in fact encourage these firms to become more thinly capitalized and, as a consequence, less safe. However, for firms that place more value on capital investment, safety expenditures may be sufficiently responsive to damages that marginal capital costs fall as damages become more punitive. If so, punitive damages can improve capital investments as well as safety choices.

Does this mean that damages should be increasing with wealth? In practice, we suggest not. The goal is to preserve punitive liability's ability to improve the decisions of some firms, while avoiding their propensity to worsen the decisions of others. However, to do this with wealth-sensitive damages, courts would require information that is both firm specific and difficult to measure. This means that practical, normative reforms are unlikely to eliminate the deterrence-reducing effects of punitive liability.⁸

Section V concludes with a discussion of policy implications, including the benefits of bifurcating trials, which separate the determination of liability from the calculation of damages.

II. The Model

This section describes the elements of the model in a relatively informal manner. The Appendix provides a formal presentation.

Capital, Risk, and Safety

The analysis focuses on the effects of liability on the decisions of a profit-maximizing firm. Firms invest in legally recoverable capital (denoted k), which they combine with other inputs to make output. Assume that each unit of the firm's output carries with it a possibility of failure, failure that would in turn lead to harm and potential product liability. A firm can invest in safety, s , which reduces its expected liability for any fixed value of k . Any fraction of the firm's output may fail. This means that the firm faces a range of potential liabilities. We place no additional structure on the effects of safety, except to assume that there are diminishing returns to the effect of safety investments on the risk of product failure.

Assume that capital is obtained from competitive (perfect) capital markets. All else

the claim that the sting of damages is felt less strongly by wealthier defendants, and so wealthier defendants must be punished more severely to achieve an equivalent level of deterrence.

⁸Our analysis therefore takes issue with those who suggest that there is no justification for wealth-sensitive damages. See Kenneth Abraham and John Jeffries, "Punitive Damages and the Rule of Law: The Role of Defendant's Wealth," 18 *Journal of Legal Studies* 415 (1989) who criticize the law's use of defendant wealth generally, and Jennifer Arlen, "Should Defendants' Wealth Matter?," 21 *Journal of Legal Studies* 413 (1992) who concludes that damages should be wealth-sensitive only when defendants are risk averse. By contrast, we show that, in principal, wealth-sensitive damages can improve deterrence. However, we will argue that efficiency-enhancing wealth-sensitive damages are difficult, if not impossible, to implement in practice.

being equal, an increased stock of capital reduces the firm's production costs (with diminishing returns). Because capital is costly, however, no firm seeks to have an unlimited capital stock. The efficient capital investment level, denoted k^* , equates the firm's marginal productive benefits of capital investment to capital's social marginal cost, denoted w .

As we will show, a firm's private marginal capital cost can exceed the social marginal cost. This divergence is due to the possibility of liability and arises because a firm's expected liability may increase as it increases its capital investment. The implication is that a firm's capital investment choice can vary, depending on the specification of damages that it expects to face. Moreover, when capital investment is discouraged by potential liabilities, production costs increase, reflecting the (implicit) substitution of other, less productive inputs for k .

The Choice of Equity Versus Debt and the Priority of Liability Claims

The model assumes the firm's capital investment k is fully recoverable by tort claimants. Given this assumption, it is most natural to think of k as equity capital. Because capital may be lost to liability claimants the equity financier must be compensated. This increases the cost of capital. The results we describe below, however, arise whenever liability adds to the cost of the firm's preferred form of capital—whatever form that may be. For instance, if the firm is instead financed with debt, the possibility of liability claims increases the firm's risk of default on the loan and reduces the amount of principal that will be returned to creditors. As in the case of equity financing, creditors must be compensated for this risk. This will be true as long as liabilities have a legal priority equal to or higher than other debts.

Importantly, however, the law may shield some nonequity forms of capital investment from liability. For example, secured debt can subordinate (take priority over) other obligations or debts. Backed by collateral, secured debt shields creditors from most of the costs associated with default risk. The presence of collateral—and the priority of the debt—reduces the compensation for liability risks that must be offered to this kind of creditor. This can make secured credit an attractive form of financing for potentially liable firms.¹⁰ But secured debt weakens the deterrent value of liability. To see this, consider a firm financed entirely by secured debt. In bankruptcy, the secured collateral will be awarded to its creditors, and not to liability claimants. Because the firm's capital wealth is completely shielded from liability, it externalizes all expected tort liabilities associated with its activities.

Our analysis will describe a liability-induced motivation—enhanced by punitive damages—to reduce absolute capital levels (irrespective of the form of financing) or to substitute toward “unrecoverable” forms of capital financing, such as secured debt. We will show that such substitution away from legally recoverable capital investment can weaken the deterrent effect of punitive liability, so that an increase in punitive damages can reduce deterrence. If, absent liability, firms would choose forms of capital that are inaccessible to tort claimants, this effect will be less important. Empirically, however,

¹⁰For a more detailed analysis of this point and critique of priority rules see Lucian Bebchuk and Jesse Fried, “The Uneasy Case for the Priority of Secured Claims in Bankruptcy,” 105 *Yale Law Journal* 857 (1996).

firms rarely rely exclusively on secured debt.¹¹ The capital reductions and substitutions we describe can, therefore, be expected to arise in a wide variety of circumstances.¹²

A Remark on Activity Levels

For expositional reasons, we focus on wealth reduction as the means to externalize liability costs. It is important to note, however, that the incentive to externalize liability can also manifest itself in a desire to increase output (the activity level). By increasing output, producers can spread their capital exposure over a larger number of units. Note that, holding wealth fixed, a firm's *per unit* expected liability cost falls as output increases. Ultimately, it is this per unit expected liability cost that determines the law's deterrent effect. The most realistic treatment of firm decisions would relate the level of damages to this per unit level of wealth.¹³ For expositional clarity, however, we suppress the choice of output level and use aggregate wealth as the illustrative decision variable.

III. Capital and Safety Investments When Damages are Compensatory

This section describes the effect of compensatory damages on deterrence when liability is strict and capital investments are sensitive to the expectation of liability. Under a broad range of conditions, the prospect of compensatory liability distorts capital investments in a way that leads to suboptimal deterrence.

The essential intuition is as follows. Firms invest in capital to produce or market their products more cheaply. Firms balance this type of benefit against the costs of increased capital. One of these costs is increased exposure to liability. If a capital investment is subject to tort claims, the private cost of that capital will exceed its social or "riskless" cost. To reduce their expected liabilities, firms can increase their safety efforts (the law's intended goal), or they can reduce the amount of wealth they expose to liability. This has two consequences that are important to a welfare analysis of damages. First, welfare maximization requires the minimization of production costs. If capital investment is being chosen in part to limit liability, production costs are not being minimized. Second, wealth reduction externalizes costs and thereby reduces the incentive to invest in safety.

¹¹The lack of available collateral frequently constrains the use of secured credit. Collateral is limited by preexisting, secured claims, such as mortgages. Also, there are situations in which potential liability itself undermines the value of collateral, even if there are no other claims on the collateral. Consider the environmental contamination of real estate. Even if the real estate may be secured, contamination can leave it valueless and thus undesirable as backing for secured debt.

¹²If this were not so, then we would expect to observe a preponderance of firms financed exclusively through secured capital. This is not the empirical reality, however, because many firms are financed with significant portions of equity. Further, suppose that, absent tax or liability considerations, the firm was financed exclusively with secured credit. In that case, liability provides no incentive for deterrence, whatever the specification of damages. So in this special case, punitive liability—trivially—does not improve deterrence.

¹³A firm's choice of wealth reduction versus output expansion is determined by the relationship of capital and output levels to production costs. For a treatment of endogenous capital, output, and safety decisions in the presence of liability see James Boyd and Daniel Ingberman, "The Search for Deep Pockets: Is 'Extended Liability' Expensive Liability," 13 *Journal of Law, Economics and Organization* 232 (1997). In the absence of liability, output is chosen so that the marginal increase in average production costs from producing an additional unit equals the marginal decrease in average capital costs (based on the social opportunity cost w) obtained by spreading capital costs across additional units of output. When the firm is potentially liable, however, the marginal capital cost includes the probability of insolvency. This leads to an expansion of output. The combined effect (reduction in capital, expansion of output) means that the firm's capital/output ratio falls.

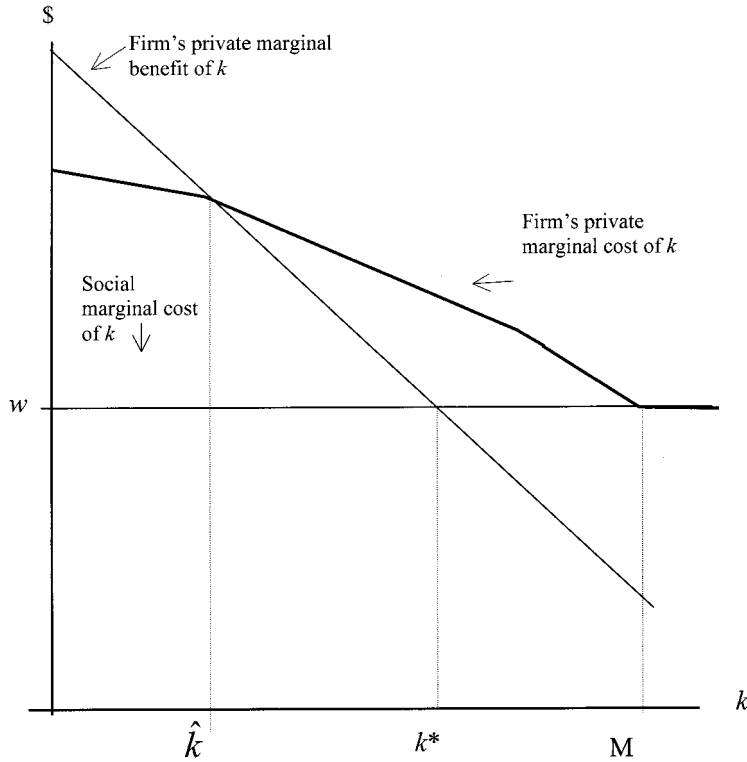


FIG. 1. Capital choice under compensatory liability.

The Marginal Cost of Capital

Consider the way in which financial markets will price the capital used by a potentially liable firm. When capital may be lost to future claims, the capital cost will include a premium to compensate for the risk of loss. When damages are compensatory, the premium can be conveyed simply. Consider a firm contemplating a marginal increase, Δk , to its total capital value, k . The firm expects a future liability L that is of uncertain magnitude. In the event that the realized loss does not exceed the firm's wealth ($L \leq k$), the marginal increase in capital Δk is not lost to liability claims, because the firm's existing capital is sufficient to internalize all claims. However, when damages exceed the firm's wealth ($L > k$), a marginal increase Δk is lost to liability claims. Thus, the liability-based premium to the marginal cost of capital is p , the probability that $L > k$.¹⁴

Liability-Driven Wealth Reduction

When does the risk premium attached to capital create welfare-reducing distortions? To address this question, refer to Figure 1. Let k^* denote the efficient level of capital

¹⁴We will show later that when damages are punitive, the risk premium is not equal to the probability that the firm will be bankrupted, but is bounded below by that probability.

investment. That is, k^* equates the marginal benefits of capital investment MB to w , the social marginal cost of capital. Also, let M equal the firm's maximum possible liability. Using the reasoning above, we know that the private marginal cost of capital must exceed w whenever the firm's wealth falls short of its maximum possible liability ($k < M$). As a result, the firm's private marginal benefits and costs of capital cannot intersect at k^* when $k^* < M$ (as is the case in the figure). Whenever $k^* < M$, the firm makes a capital investment \hat{k} that is strictly less than the efficient level k^* . Under a broad set of circumstances, then, strict compensatory liability has a wealth-reducing effect. In those circumstances, firms are bankrupt with some positive probability and are therefore underdeterred.

PROPOSITION 1: Assume liability is strict and damages compensatory. Then, whenever a firm's maximum liability exceeds its efficient level of capital investment, the firm will (1) underinvest in capital, (2) be bankrupt with positive probability, and (3) underinvest in safety.

The firm will choose the optimal capital investment k^* only if there is no probability of insolvency at that level of investment ($k^* > M$). If the maximum realization of damages M is less than the first-best capital k^* , then k^* is a solution to the firm's capital choice problem.

The degree to which capital reduces production costs varies industry to industry and market to market.¹⁵ In turn, this implies variability in firms' efficient capital scale—i.e., the total capital value that minimizes production costs. Liability-driven distortions in capital investment are more likely in industries where the efficient capital scale is small relative to the social costs of production. When a firm's efficient capital scale is small relative to its potential liability, there is a greater likelihood that liabilities will bankrupt it. In turn, it is more likely that capital's cost involves a significant risk premium and therefore that capital investment will be distorted.

IV. The Effect of Punitive Liability on Deterrence

The previous section derived a simple test to determine whether strict liability with compensatory damages will lead to efficient safety and capital investments. Namely, is the firm's first-best level of capital investment greater than its largest possible liability? If so, compensatory liability will lead to efficient investments. If not, the firm will underinvest in both safety and capital. We now turn to the central question of this paper: Do punitive damages improve deterrence?

To answer the question, we first present a model of safety and capital investment in which liability is strict and total damages are a given multiple of the actual loss. This allows simple analysis of the comparative static effect of more punitive damages on safety and capital investment. We then turn to the imposition of punitive damages in practice and describe why the "fixed multiple" description of punitive damages need not always be the most realistic. Using a richer description of punitive damages, we use a parametric example to illustrate how safety and capital decisions respond to punitive liability. This also illustrates the conditions under which punitive liability reduces, rather than enhances, deterrence.

¹⁵Capital's benefits are a function of the cost savings yielded by investment in capital versus other inputs. For instance, investment in capital may reduce the costs of developing, manufacturing, distributing, and servicing the firm's product.

Damages as a Fixed Multiple of the Loss

Punitive damages are triggered by noncompliance with some standard of care, typically gross negligence or recklessness. Compliant firms—those found to be not grossly negligent or reckless—remain liable, but only for compensatory damages. Compliance cannot be guaranteed, however, when capital investments are endogenous. Punitive liability, therefore, has two general effects. First, it increases the incentive to comply with the relevant standard of care. Second, it affects the decisions of firms that expect to be noncompliant.

Consider the decisions of a firm that expects to be punitively liable. As shown in the Appendix, the firm's maximization of profit by choice of s and k yields two first-order conditions. Totally differentiating these first-order conditions by f (the multiple applied to the actual loss to determine the level of assessed damages) yields the following condition.

$$\frac{d\hat{s}}{df} = \frac{-\pi_{kf}\pi_{ks} + \pi_{kk}\pi_{sf}}{(\pi_{ks})^2 - \pi_{kk}\pi_{ss}}$$

Because our primary interest is in the effect of increased damages on the firm's safety choice (\hat{s}), it is useful to explore this condition in more detail. The interpretation of the terms in the numerator is straightforward. The denominator is assumed to be negative, which guarantees the appropriate curvature to satisfy the second-order conditions for profit maximization.

Π_{kf} describes the effect of increased damages on the net marginal benefit of capital investment. Because higher damages increase the cost of capital, this effect is negative. Π_{ks} describes the effect of increased safety on the marginal benefit of capital investment. Because more safety reduces the cost of capital, this effect is positive. Π_{kk} is negative by assumption, to reflect diminishing marginal returns to capital investment.

Π_{sf} denotes the effect of increased damages on the marginal benefits of safety. This term is of ambiguous sign. It is most intuitive to think of Π_{sf} as being positive, because higher penalties would seem naturally to increase the reward to greater safety. This need not be the case, however. In fact, potential insolvency creates conditions under which Π_{sf} can be negative.¹⁶ To see why, consider a range of possible losses, some of which would leave the firm solvent, some of which would not. Now increase the penalty for those losses. As the penalty increases, the expected cost of losses that do not bankrupt the firm increases. This increases the marginal benefit of safety. However, increasing the penalty also increases the range of losses that leave the firm insolvent. This reduces the marginal benefit of safety. An expanded range of bankrupting losses makes it less likely that an increase in safety will reduce the firm's expected liability. If bankruptcy occurs with or without increased safety, greater safety yields no private benefit.

The effect of punitive damages on wealth—and their ability to reduce deterrence—is most clearly seen if we set $\Pi_{sf} = 0$ (hold constant the effect of increased damages on the marginal benefits of safety). Holding Π_{sf} constant, the sign of Π_{kf} determines the effect

¹⁶This type of effect is at the heart of the literature on "marginal deterrence." For examples, see David Friedman and William Sjöström, "Hanged for a Sheep—The Economics of Marginal Deterrence," 22 *Journal of Legal Studies* 345 (1993) and James Boyd and Daniel Ingberman, "Non-Compensatory Damages and Potential Insolvency," 23 *Journal of Legal Studies* 895 (1994).

of more punitive damages on safety investments. Because higher damages increase the cost of capital, the firm reduces its wealth.¹⁷ This wealth reduction creates a corresponding net reduction in safety. Note also that safety unambiguously falls whenever higher penalties reduce the marginal benefit of greater safety ($\Pi_{sf} < 0$). Thus,

PROPOSITION 2: *Relative to compensatory liability, and when capital investments are made in anticipation of potential liabilities, punitive damages can lead to reduced investment in both capital and safety.*

When higher penalties increase the marginal benefit of greater safety ($\Pi_{sf} > 0$), higher penalties have an ambiguous effect on deterrence. This is due to the tension between punitive damages' positive impact on the benefits of safety and their negative impact on the benefits of capital investment. The example presented in the following section assumes that $\Pi_{sf} > 0$ so that this trade-off can be explored in greater detail.

V. Uncertain, Wealth-Sensitive, and Capped Damages

This section explores the robustness of the previous section's conclusions. First, we offer a brief description of current punitive damages law. This review suggests that, from an *ex ante* perspective, the level of damages is most realistically depicted as simultaneously uncertain and increasing in defendant wealth. Second, we present the results of a parametric example based on a damage rule with those properties. The example confirms the robustness of our conclusions. It also illuminates the relationship between punitive damages' ability to improve deterrence and the marginal benefits of capital investment. Finally, the example is used to compare the desirability of current punitive damages law to two alternatives: compensatory damages and capped punitive damages, a common punitive damages reform proposal.

How Does the Law Define the Level of Damages?

In practice, courts use a wide variety of criteria to calculate the scale of punitive damages, including the defendant's wealth. As a consequence, expected punitive damages should not in general be thought of as a fixed multiple of the actual harm. In the landmark 1991 case *Pacific Mutual v. Haslip*¹⁸ the U.S. Supreme Court approved of standard criteria to determine whether a punitive award is "reasonably related to the goals of deterrence and retribution." The court stated that the following factors could be taken into consideration in determining whether a punitive damage award was excessive or inadequate:

- (a) whether there is a reasonable relationship between the punitive damages award and the harm likely to result from the defendant's conduct as well as the harm that actually has occurred;
- (b) the degree of reprehensibility of the defendant's conduct, the duration of that conduct, the defendant's awareness, any concealment, and the existence and frequency of similar past conduct;
- (c) the profitability to the defendant of the wrongful conduct and the desirability of removing that profit and of having the defendant also sustain a loss;
- (d) the "financial position" of the defendant;

¹⁷Recall that higher damages increase the probability of insolvency, which in turn increases the cost of capital.

¹⁸*TXO Production Corp. v. Alliance Resources* is a subsequent decision in agreement. See note 1 *supra*.

- (e) all the costs of litigation;
- (f) the imposition of criminal sanctions on the defendant for its conduct, these to be taken in mitigation; and
- (g) the existence of other civil awards against the defendant for the same conduct, these also to be taken in mitigation.”

The law does not specify the weight given to each of these factors. Therefore, nearly any punitive damage award is consistent with these guidelines. Thus, firms may plausibly anticipate damages that are essentially randomly determined somewhere between the actual loss and k .¹⁹ Although state reforms have in some cases provided more specific guidelines for the damage calculation, by and large the pattern of state jury instructions does little to reduce uncertainty.²⁰

The parametric example presented below compares three damage “rules,” *compensatory damages*, *uncapped punitive damages*, and *capped punitive damages*. Compensatory damages are assumed to be certain and equal to the actual loss. The uncapped damages rule is our representation of the range of damage calculations permitted in *Pacific Mutual*. We model the uncapped damages rule, as a random draw from a distribution bounded below by the actual loss and above by the defendant’s wealth. The capped damages rule is a common reform proposal. To limit the uncertainty associated with current punitive damages law, proponents urge a cap on punitive damages equal to three times the actual loss. Thus, we model the capped rule as a random draw between the actual loss and three times the actual loss (or the firm’s wealth, if it is less).

Many states explicitly permit the scale of punitive damages to be conditioned on the defendants’ wealth.²¹ Beyond reflecting the underlying uncertainty in the damage calculation, both the capped and uncapped rules are consistent with this feature of law, as well. That is, the expected damage is increasing in defendant wealth under both rules.

A Parametric Example

The Appendix describes a parametric model of a firm’s optimal safety and capital choices under the three damage rules described above. Two figures illustrate how more punitive liability can have possibly counterintuitive implications for the efficiency of safety and capital decisions.

Figure 2 depicts the privately optimal safety investment as a function of the firm’s wealth k and the damage rule in place.²² This figure conveys the common, intuitive, but incomplete view of punitive damages’ effect on precaution: “more punishment” induces better deterrence. Because of the “marginal deterrence” problem, however, even this feature is a consequence of our parametric assumptions. The assumptions imply that for

¹⁹Justice O’Connor’s dissent in *Pacific Mutual*, note 1 *supra*, is consistent with this interpretation.

²⁰See Ellis, *supra* note 1, or the court in *Gertz v. Robert Welch, Inc.* 418 U.S. 323, 360 (1974): “In most jurisdictions jury discretion over the amounts awarded [as punitive damages] is limited only by the gentle rule that they not be excessive. Consequently, juries assess punitive damages in wholly unpredictable amounts bearing no necessary relation to the actual harm caused.”

²¹Thirty-eight states permit the admissibility of evidence of wealth or financial status of the defendant when other evidence in the case supports the submission of a punitive damages claim to the jury. See Michael Rustad, “In Defense of Punitive Damages in Products Liability: Testing Tort Anecdotes with Empirical Data,” 78 *Iowa Law Review* 1 (1992), and James D. Ghiardi and John J. Kircher, “Punitive Damages Law and Practice” @ 5.36 (4th ed. Supp. 1990).

²²In each figure, the solid curve corresponds to compensatory damages, the dotted curve to capped punitive liability, and the dashed curve corresponds to uncapped punitive liability.

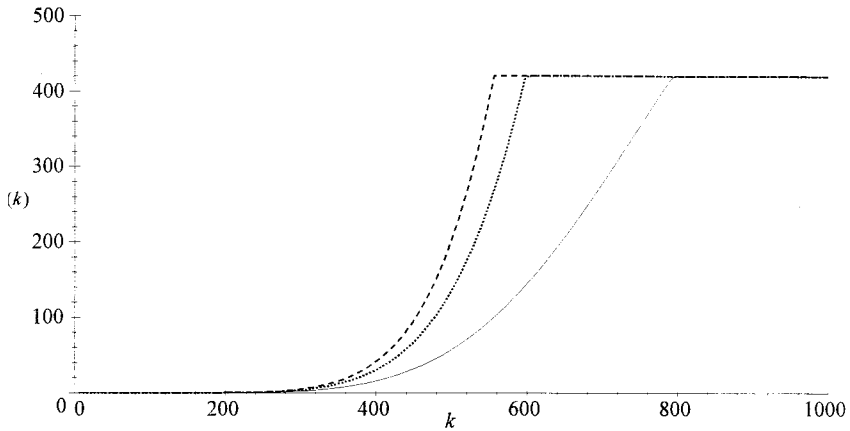


FIG. 2. Safety choices.

a *given* level of wealth, the uncapped damages rule leads to greater safety than the capped damages rule, and both lead to more safety than compensatory damages.

As shown in Figure 2, the rules induce relatively insignificant differences in safety expenditure when the firm has little wealth. This is true for two reasons. First, greater wealth amplifies the effect of higher damages. Second, expected damages under the uncapped and capped rules are increasing in wealth.

However, Figure 2 by itself does not describe the effect of damages on deterrence, because the firm's choice of wealth is sensitive to the level of damages. To complete the analysis, it is necessary to compute the levels of wealth chosen under the damage rules. The relevant comparison is between the safety choices implied by those *different* levels of wealth.

Figure 3 depicts the marginal cost of capital as a function of wealth (in a manner similar to Figure 1). When wealth is low, the effect of damages on capital costs is as expected. Higher damages imply higher marginal capital costs. But note that punitive

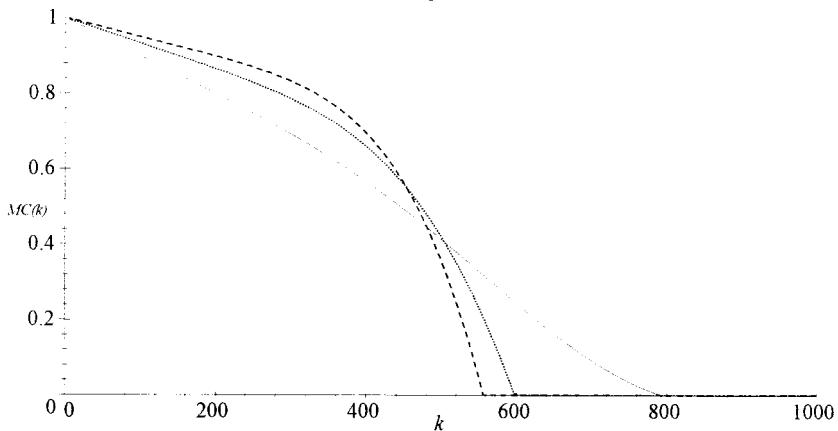


FIG. 3. Marginal costs of capital.

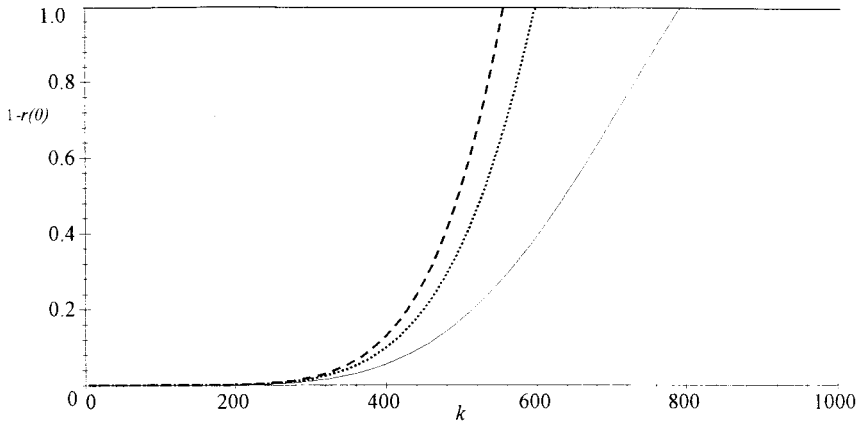


FIG. 4. Probabilities of loss given profit-maximizing safety choices.

damages do not always imply higher marginal capital costs. In fact, marginal capital costs are lowest under the uncapped damages rule for some higher levels of wealth. To understand why, note that the marginal cost of capital depends not only on k and the level of damages, but also on the (simultaneously determined) choice of safety. More punishment can stimulate the firm to invest relatively heavily in safety. When this occurs, the marginal cost of capital is driven down because it reduces the probability of insolvency.

This ambiguity was noted earlier when we considered the sign of $d\hat{s}/df$. That analysis revealed a tension between punitive damages' tendency to increase capital costs by increasing the cost of a given loss and their tendency to reduce capital costs by stimulating greater investment in safety. Figure 3 depicts that tension clearly. At low wealth levels the cost-increasing effect of punitive damages dominates.

The net effect of punitive damages on safety fundamentally depends on the firm's demand for capital. Consider two different types of firms, distinguished by their marginal benefits of capital, low (MB_L) and high (MB_H). Figure 4 is like Figure 2, with one difference. Instead of showing safety expenditures as a function of k , Figure 2 illustrates the probability that a loss occurs, given the firm's safety expenditures (in this parameterization, that probability is a direct measure of the expected social loss associated with the firm's activities). Figure 5 reproduces Figure 3, with the addition of the two marginal benefit curves MB_L and MB_H .

Note that when the marginal benefits of k are low, higher damages reduce the firm's choice of capital investment. Translating these k choices into safety levels, as in Figure 2, it is therefore possible that punitive damages will reduce safety and thereby deterrence. (At lower k 's safety is relatively unresponsive to changes in damages and relatively responsive to changes in k . As a result, unless MB_L is nearly vertical, safety falls as damages increase.) Thus, when firms place relatively little value on capital investment, more punishment leads to less deterrence. At best, if deterrence increases, then it does so at the cost of distortions in capital investment.

The conclusion is different when firms value capital more highly. For a firm with marginal benefits of capital MB_H , punitive damages *increase* capital investment by stimulating a large increase in safety. For this type of firm, future liabilities are minimized via greater investment in safety, rather than by reductions in liability exposure.

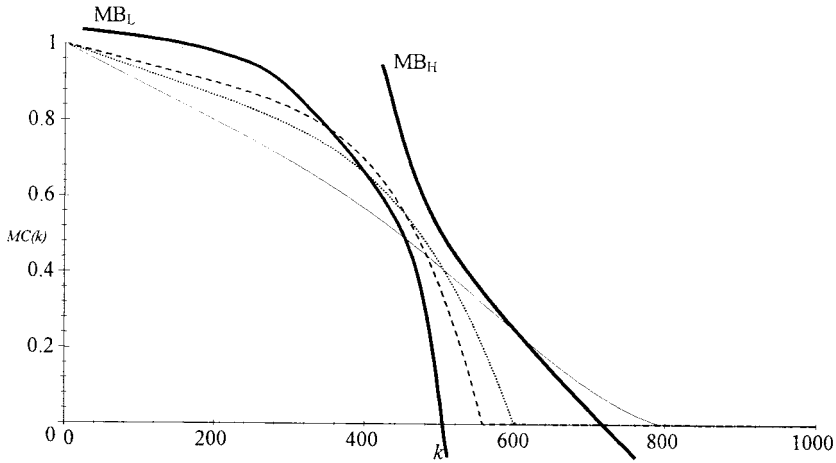


FIG. 5. Profit-maximizing capital choices.

Thus, for firms that derive enough value from capital investment, punitive liability not only increases deterrence, but also reduces liability-driven distortions in capital investment.

Clearly, the analysis reveals that the optimal scale of damages hinges on the demand for capital, a firm-specific and largely unobservable attribute. This suggests that the overall effect of punitive liability—and consequently, the desirability of reform—is uncertain. However, the parameterization makes clear that punitive damages reduce the safety of those firms who are least deterred by liability generally (i.e., firms with weak demand for capital). Thus, punitive damages can exacerbate, rather than mitigate, the inefficiency they ostensibly address: the social costs created by significantly underdeterred firms.

Is it practical for courts to acknowledge these contrasting effects of punitive liability and tailor damages to account for it? Although conceptually possible, such a rule seems to be impractical. To apply punitive damages only when they improve deterrence, courts would need extensive information about the way in which firms' safety expenditures respond to higher penalties and increased capital. Even in the unlikely event that courts could access such information, the resulting legal rule would likely apply punitive damages in firm- or industry-specific ways. The resulting system of liability would effectively apply different rules to different types of firms and industries. We doubt that such a system would be politically sustainable.

VI. Conclusions and Policy Implications

This paper has explored the deterrent effect of punitive damages when capital investment is endogenous to the legal rule. Particularly for those firms least deterred by liability generally, punitive damages can exacerbate wealth reductions induced by the threat of liability and thereby reduce deterrence. The analysis can also be used to weigh the benefits of proposed reforms.

First, consider questions regarding the scale of damages. For example, would capping punitive damages at three times the actual loss increase efficiency? More generally, do

punitive damages distort capital and safety decisions so much that compensatory liability is preferred?

One interpretation of the paper's results is that the normative comparison of different damage levels is inherently ambiguous. Punitive damages improve the decisions of some firms and worsen the decisions of others. Punitive damages can improve the safety and capital investments of firms that place a relatively high value on capital. Do these benefits outweigh the negative impact of punitive damages on the decisions of firms that place relatively low value on capital? This is primarily an empirical question, and one that requires detailed knowledge of potential defendants' marginal benefits of capital investment and the ability to avoid losses through safety investment. It is unlikely that the information needed to weigh the overall desirability of punitive damages could be available to either courts or econometricians. Further, the welfare implications of punitive liability also depend on its effects on capital investment. For instance, more punitive liability can increase the cost of legally recoverable capital relative to the cost of other productive inputs. After an earlier discussion, this might mean that firms will seek secured financing that is otherwise more costly than other forms of capital and that is shielded from tort claims.

However, in one important respect the results of our analysis are not ambiguous. Specifically, the firms who are least deterred by liability generally—i.e., those firms whose safety incentives can be most improved—are also the most likely to have their decisions worsened by punitive damages. Thus, punitive damages reduce deterrence in precisely those cases where improvement is most needed. Punitive liability is an unlikely "solution" if the "problem" is the decisions of firms that are significantly underdeterred by liability.

As a second application of the model, consider the issue of whether or not the standards of care that trigger punitive liability should be wealth sensitive. As described earlier, once a defendant's actions are found to warrant punitive liability, the law allows the size of damages to be conditioned on wealth. However, because standards are not objectively defined and defendant wealth is introduced to determine the scale of damages, there is the practical possibility that juries will use defendant wealth to determine the standard of care that triggers punitive damages. Several states have instituted procedural safeguards (specifically, bifurcated punitive damage trials) to ensure that the standard that determines punitive liability is not wealth sensitive.²³

We have shown that punitive damages increase the marginal cost of capital when wealth is small. Marginal capital costs are further increased when the scale of the damage is increasing in the defendant's wealth. When the standard that triggers punitive liability is also increasing in wealth, the marginal cost of capital increases even more. In this case, more wealth means being held to a higher standard. Thus, although wealth-sensitive standards can improve the decisions of some firms, they share the same property as wealth-sensitive damages: Both exacerbate the capital distortions and lack of safety exhibited by the firms least deterred under compensatory liability. We therefore conclude that wealth-sensitive standards are not likely to be an effective means of deterring this class of potential defendant. Procedural reforms, such as bifurcated trials,

²³Bifurcated trials split the trial into two separate proceedings, one to establish whether or not liability is to be punitive and, if it is, another to establish the size of damages. This procedure is an attempt to disengage the determination of whether or not a defendant's behavior warrants punitive damages from its characteristics, in particular the depth of its pockets. Several states require bifurcation by statute. See Ga. Code Ann. §51-12-5.1(d) (Supp. 1988), Mont. Code Ann. §27-1-221(7)(a) (1987), N.J. Stat. Ann. §2A:58C-5(b) (1987).

that limit the conditioning of standards on wealth, are therefore a desirable policy intervention.

Appendix

Proof of Proposition 2

We now demonstrate that increased f can reduce deterrence. Our basic assumptions are as follows. Equity capital investment reduces production costs, but can increase expected liabilities. Safety investments can reduce expected liabilities. There are diminishing returns in both safety and capital.

The possibility that increasingly punitive damages can reduce deterrence can be seen at a general level, without imposing any assumptions on the market structure in which the firm operates or any requirement of long-run equilibrium.

The firm's problem is to

$$\max_{k,s} \pi(k, s, f). \quad (1)$$

The first-order conditions for problem (1) are

$$\frac{\partial \pi}{\partial k} = \pi_k = 0, \quad (2)$$

$$\frac{\partial \pi}{\partial s} = \pi_s = 0. \quad (3)$$

We are assuming that associated second-order conditions hold. Thus, the first-order conditions [equations (2) and (3)] define the firm's privately optimal safety investment, $\hat{s}(\cdot)$, and its privately optimal capital investment, $\hat{k}(\cdot)$, as identities:

$$\pi_k(\hat{s}, \hat{k}, f) = 0, \quad (4)$$

$$\pi_s(\hat{s}, \hat{k}, f) = 0. \quad (5)$$

We are interested in the effect of increased f on $\hat{s}(\cdot)$. Totally differentiating the identities defining \hat{s} and \hat{k} [equations (4) and (5)] yields

$$\pi_{ks}d\hat{s} + \pi_{kk}d\hat{k} + \pi_{kf}df = 0, \quad (6)$$

$$\pi_{ss}d\hat{s} + \pi_{sk}d\hat{k} + \pi_{sf}df = 0. \quad (7)$$

Dividing equations (6) and (7) by df yields

$$\pi_{ks} \frac{d\hat{s}}{df} + \pi_{kk} \frac{d\hat{k}}{df} + \pi_{kf} = 0, \quad (8)$$

$$\pi_{ss} \frac{d\hat{s}}{df} + \pi_{sk} \frac{d\hat{k}}{df} + \pi_{sf} = 0. \quad (9)$$

Solving equation (8) for $d\hat{s}/df$ yields

$$\frac{d\hat{s}}{df} = \frac{-\pi_{kf} - \pi_{kk} \frac{d\hat{k}}{df}}{\pi_{ks}}, \quad (10)$$

and solving equation (9) for $d\hat{k}/df$ yields

$$\frac{d\hat{k}}{df} = \frac{-\pi_{sf} - \pi_{ss} \frac{d\hat{s}}{df}}{\pi_{sk}}. \quad (11)$$

Substituting the left-hand side of equation (11) into equation (10) yields

$$\frac{d\hat{s}}{df} = \frac{-\pi_{kf} - \pi_{kk} \left(\frac{-\pi_{sf} - \pi_{ss} \frac{d\hat{s}}{df}}{\pi_{sk}} \right)}{\pi_{ks}} \quad (12)$$

$$= -\frac{\pi_{kf}}{\pi_{ks}} + \frac{\pi_{kk}\pi_{sf}}{(\pi_{ks})^2} + \left(\frac{\pi_{kk}\pi_{ss}}{(\pi_{ks})^2} \right) \left(\frac{d\hat{s}}{df} \right). \quad (13)$$

Thus

$$\frac{d\hat{s}}{df} = \frac{-\frac{\pi_{kf}}{\pi_{ks}} + \frac{\pi_{kk}\pi_{sf}}{(\pi_{ks})^2}}{\left(1 - \frac{\pi_{kk}\pi_{ss}}{(\pi_{ks})^2} \right)}. \quad (14)$$

Multiplying the numerator and denominator of equation (14) by $(\pi_{ks})^2$ then yields

$$\frac{d\hat{s}}{df} = \frac{-\pi_{kf}\pi_{ks} + \pi_{kk}\pi_{sf}}{(\pi_{ks})^2 - \pi_{kk}\pi_{ss}}. \quad (15)$$

Observe that the denominator of equation (15) is negative by second-order conditions. The terms in the numerator have the following signs:

- π_{kf} $-$ Higher f reduces the net marginal benefits of k .
- π_{ks} $+$ Higher s increases the net marginal benefits of k .
- π_{kk} $-$ k provides diminishing returns in production.
- π_{sf} \pm Higher f can either increase or decrease the net marginal benefits of s .

The sign of π_{sf} depends on the possibility of the “marginal deterrence” problem. As liability becomes more punitive, a larger range of losses, b , will lead to the firm’s insolvency. Because all b ’s that lead to insolvency are equally costly to the firm, a firm never invests in safety to reduce the realization of b among those b ’s that lead to insolvency. Thus, when marginal deterrence is a problem, increasingly punitive liability can lead to less deterrence, even when capital investments are exogenously fixed.

Suppose, therefore, that marginal deterrence is not a problem. Then increasingly

punitive liability increases the net marginal benefits of safety, all else equal (i.e., $\pi_{sf} > 0$). As shown by Boyd and Ingberman (1993), this is true whenever safety investments influence only the probability that a loss occurs, and not its scale, as long as the firm is not insolvent for every realization of loss.

Thus, when marginal deterrence is ruled out, then—all else being equal—safety responds positively to increasingly punitive liability or increased capital exposure, and capital responds positively to increased safety. Nonetheless, it is still possible that increased f reduces deterrence.

To see this, observe that the two products in the numerator of equation (15) have opposite signs. The first term says that, all else being equal, as f increases, the incentive to invest in capital is reduced. This incentive to reduce capital exposed to liability reduces the net marginal benefits of safety investment, leading to an incentive to reduce safety. On the other hand, holding capital fixed, increased f increases the incentive to invest in safety. And, all else being equal, the increase in safety increases the net marginal benefits of capital investment.

The relative magnitudes of these two effects determine the net effect of increased f on safety investment.

First, suppose that, all else being equal, the safety response to increased f is relatively small. Then punitive liability reduces the net marginal benefits of capital investment created by liability. This reduction in capital exposure can lead to less deterrence overall: Although safety is increased—slightly—for every k , the privately optimal k can be significantly reduced by increasingly punitive liability.

Otherwise, suppose that, all else being equal, safety responds significantly to increased f . Then punitive liability can increase the net marginal benefits of capital investment, by significantly reducing liability risk at every level of k . If capital investment is thereby increased, then—at least when marginal deterrence is not a problem—increasingly punitive liability can increase deterrence while it reduces the distortion in capital investment created by liability.

Parametric Example

We make the following assumptions.

1. Safety expenditures influence the probability but not the scale of a loss b . If a loss does occur, then b is distributed uniformly over the interval $[0, M]$.
2. The loss is preventable. The probability that no loss occurs is $1 - r(s) = \min\{1, cs^{1/\alpha}\}$, where c and α are positive constants.

Analysis for Arbitrary Specification of Damages.

EXPECTED LIABILITIES AS A FUNCTION OF THE DAMAGE RULE. Observe that a safety expenditure $\bar{s} = (1/c)^\alpha$ will ensure that no loss occurs. The firm's expected liability, given that a loss occurs, is independent of its safety expenditure. Given that a loss does occur, the firm's expected liability is determined by the firm's capital value, the damage rule that is in place, and the assumption that b is distributed uniformly when a loss occurs.

Given this parametric structure, the firm's expected liabilities $L(k, s, f)$ can be written in a particularly simple form. $L(s, k, f)$ is the product of the probability that a loss occurs, given the firm's safety expenditures, and the firm's expected liability, given that a loss occurs. Let $E(k, M, f)$ denote the firm's expected liabilities, given that a loss occurs, and given the damage rule in place. (For example, as shown below, with

compensatory liability, $E(k, M, f) = [k - k^2/2M]$.) Then the firm's expected liabilities, given the damage rule in place f and the firm's safety and capital choices, is given by

$$L(k, s, f) = r(s) E(k, M, f). \tag{16}$$

SAFETY CHOICES. For any damage rule and any capital investment k , the firm chooses safety to minimize its combined expected liabilities and expenditures on safety. That is, given any k , the firm safety decision solves

$$\min_s (s + r(s) E(k, M, f)). \tag{17}$$

The solution of this problem is

$$\hat{s}(k, f) = \min \left\{ \bar{s}, \left(\frac{\alpha}{cE(k, M, f)} \right)^{\alpha/1-\alpha} \right\}. \tag{18}$$

CAPITAL INVESTMENTS. For simplicity, we can assume the social marginal cost of capital $w = 0$. Then the firm's private marginal cost of capital is found by partially differentiating the firm's expected liabilities by k and evaluating at $\hat{s}(k, f)$:

$$MC(k, f) = r(\hat{s}(k, f)) \left(\frac{\partial E(k, m, f)}{\partial k} \right). \tag{19}$$

Given any f , the firm's capital investment decision \hat{k} equates $MC(k, f)$ to the firm's marginal benefits of capital, $MB(k)$.

Compensatory Damages. With compensatory damages and $s < \bar{s}$, the firm's expected liability is

$$L(s, k) = \left(k - \frac{k^2}{2M} \right) (1 - \min\{1, cs^{1/\alpha}\}), \tag{20}$$

which is the product of the probability that a loss occurs $(1 - \min\{1, cs^{1/\alpha}\})$ and the firm's expected liability, given that a loss occurs $(k - (k^2/2M) = \int_0^k (1/M) b db + k \int_k^M (1/M) db)$.

Given any capital investment k , the firm chooses safety to minimize its combined expenditures on safety and capital. With compensatory damages, then, the firm's problem is to

$$\min \left(s + \left(k - \frac{k^2}{2M} \right) (1 - \min\{1, cs^{1/\alpha}\}) \right). \tag{21}$$

The solution of this problem is

$$\hat{s}(k) = \min \left\{ \bar{s}, \left(\frac{\alpha}{c \left[k - \frac{k^2}{2M} \right]} \right)^{\alpha/1-\alpha} \right\}. \tag{22}$$

Assume $w = 0$. Then the marginal cost of capital is found by partially differentiating the firm's expected liabilities by k and evaluating at $\hat{s}(k)$:

$$MC(k) = r(\hat{s}(k)) \left(1 - \frac{k}{M}\right) \quad (23)$$

$$= \left(1 - \min \left\{ 1, c \left(\left(\frac{\alpha}{c \left(k - \frac{k^2}{2M} \right)} \right)^{1/1-\alpha} \right) \right\} \right) \left(1 - \frac{k}{M}\right). \quad (24)$$

Observe that with compensatory damages, the marginal cost of capital is simply the probability of insolvency. To see this, note that $r(\hat{s}(\hat{k}))$ is the probability that some loss occurs, given \hat{k} and \hat{s} . Given that a loss occurs, $(1 - k/M)$ is the probability that such a loss leads to insolvency. The product of these terms is thus the probability that the firm will be made insolvent from its liabilities. Intuitively, this follows from the fact that with compensatory damages, the firm anticipates damages $\min\{b, k\}$ when loss b occurs. Thus, increases in k do not change the firm's expected liabilities for losses $b < k$. As a result, the marginal change in expected liability from a small increase in k is just the probability that the incremental capital investment is lost to liability assessments. With compensatory damages, this can only occur when $b \geq k$.

Punitive Liability. Because the loss is preventable, we assume that punitive liability can apply whenever a loss does, in fact, occur. This means that firms' expected liabilities are continuous functions of their safety decisions; in particular, in contrast to some analyses of liability conditioned on safety choices, expected liabilities do not "jump" downward when a threshold of compliance is reached. Therefore, the firms' safety decisions are continuous functions of their capital values for all the damage rules we consider.

UNCAPPED PUNITIVE LIABILITY. With uncapped punitive liability, any realization of loss leads to the following assessed damages. Compensatory damages of b are augmented with a punitive award. If $b < k$, the firm's total damages are drawn from a uniform distribution defined over $[b, k]$. If $b > k$, then the firm is assessed damages k . Thus, for realizations $b < k$, the expected fine equals the midpoint between the firm's wealth and the realized social loss. If the realized b exceeds the firm's wealth, then the firm loses k with certainty. For simplicity, we use the notation $f = u$ to denote uncapped punitive liability. Given this description, the firm's expected liability costs are

$$L(s, k, u) = r(s) E(s, k, u) \quad (25)$$

$$= r(s) \left(\int_0^k \left(\int_b^k \frac{x}{k-b} dx \right) \left(\frac{1}{M} \right) db + \int_k^M \left(\frac{k}{M} \right) db \right). \quad (26)$$

The first integral inside the brace measures the firm's expected liabilities for realizations of loss $b < k$, while the second integral reflects liabilities for realizations of loss $b \geq k$.

Evaluating the integrals yields

$$E(s, k, u) = \left(\int_0^k \left(\int_b^k \frac{x}{k-b} dx \right) \left(\frac{1}{M} \right) db + \int_k^M \left(\frac{k}{M} \right) db \right) \quad (27)$$

$$= k - \frac{k^2}{4M} \quad (28)$$

so

$$L(s, k, u) = r(s) \left(k - \frac{k^2}{4M} \right). \quad (29)$$

Then from equation (18) we have the firm's safety choice, given its k and uncapped liability:

$$\hat{s}(k, u) = \min \left\{ \bar{s}, \left(\frac{\alpha}{c \left[k - \frac{k^2}{4M} \right]} \right)^{\alpha/1-\alpha} \right\}. \quad (30)$$

Similarly, using equation (19) we obtain the firm's marginal cost of capital under uncapped liability:

$$MC(k, u) = r(\hat{s}(k, u)) \left(1 - \frac{k}{2M} \right). \quad (31)$$

Interestingly, the firm's probability of insolvency has the same form under uncapped punitive liability as under compensatory damages. The probability of insolvency in both cases is simply $r(\hat{s})(1 - k/M)$. In contrast to the compensatory case, though, uncapped punitive liability leads the firm's marginal cost of capital to strictly exceed its probability of insolvency. The reason is that for any $b < k$, the firm's liability with compensatory damages is simply b . But under uncapped punitive liability a realization of loss $b < k$ leads to an expected damage assessment of $(b + k)/2 = \int_b^k [x/(b - k)] dx$. That is, uncapped punitive damages cause expected damage assessments for any b to increase with wealth k . This means that a small increase in capital investment increases the firm's expected liability for b 's that do not bankrupt the firm as well as those b 's that lead to bankruptcy. Therefore, uncapped punitive liability leads the firm's private marginal cost of capital to strictly its probability of insolvency.

CAPPED PUNITIVE LIABILITY. With capped punitive liability, total assessed damages cannot exceed $3b$, i.e., three times the actual loss. We continue to assume that the firm's actual assessed damage is random within this interval. That is, if $b < k$, the firm's total damages are drawn from a uniform distribution defined over $[b, \min\{3b, k\}]$. Otherwise, if $b > k$, then the firm is assessed damages k .

For simplicity, we use the notation $f = m$ to denote capped punitive liability. Given this description, the firm's expected liability costs are

$$L(s, k, m) = r(s) E(s, k, m) \quad (32)$$

where $E(s, k, m)$ is given by

$$\left(\int_0^{k/3} \left(\int_b^{3b} \frac{x}{2b} dx \right) \left(\frac{1}{M} \right) db + \int_{k/3}^k \left(\int_b^k \frac{x}{k-b} dx \right) \left(\frac{1}{M} \right) db + \int_k^M \left(\frac{k}{M} \right) db \right). \quad (33)$$

The first integral inside the brace measures the firm's expected liabilities for realizations of loss $b < k/3$ (so that the triple damage award does not bankrupt the firm), the second integral reflects liabilities for realizations of loss $k \geq b \geq k/3$ (where the triple damage award does bankrupt the firm but the compensatory portion does not), and the third integral represents the liabilities for realizations of loss $b > k$ (where compensatory damages bankrupt the firm).

Evaluating the integrals yields

$$E(s, k, m) = k - \frac{k^2}{3M} \quad (34)$$

so

$$L(s, k, u) = r(s) \left(k - \frac{k^2}{3M} \right). \quad (35)$$

Following the same steps as in the uncapped punitive liability case, we have the firm's safety choice, given its k and capped liability:

$$\hat{s}(k, u) = \min \left\{ \bar{s}, \left(\frac{\alpha}{c \left[k - \frac{k^2}{3M} \right]} \right)^{\alpha/1-\alpha} \right\}. \quad (36)$$

Similarly, the firm's marginal cost of capital under capped liability is

$$MC(k, m) = r(\hat{s}(k, m)) \left(1 - \frac{2k}{3M} \right). \quad (37)$$

Numerical Calculations. All figures use the following parameter values: $\alpha = 1.14$, $c = 0.005$, $M = 1000$.