Punitive Damages for Deterrence: When and How Much?

Robert Cooter

Available at: https://works.bepress.com/robert_cooter/91/
PUNITIVE DAMAGES FOR DETERRENCE: WHEN AND HOW MUCH?

Robert D. Cooter
PUNITIVE DAMAGES FOR DETERRENCE: WHEN AND HOW MUCH?

Robert D. Cooter*

Abstract: An economic model of deterrence is used to derive two prescriptions for punitive damages that apply to economically motivated defendants, including businesses. To recover punitive damages, a plaintiff should have to prove a gross shortfall from the legal standard by the defendant that would be profitable if liability were limited to compensatory damages. In a word, "incentive inadequacy" should be proven. A key element in the proof is establishing the existence and extent of enforcement error. If the court finds that punitive damages are to be awarded, the punitive multiple should be set equal to the reciprocal of the enforcement error. These rules, if adopted, would provide incentives for efficient deterrence and would have significant consequences for current legal practice, including making the ratio rule more definite, lowering the ratios actually allowed, and largely eliminating a defendant's net worth as a consideration for setting punitive damages.

TABLE OF CONTENTS

INTRODUCTION .......................................................... 1144
1. ECONOMIC MODEL OF PUNITIVE DAMAGES FOR DETERRENCE ........................................... 1149
   A. Simple Model of Binary Choice .................................................. 1149
   B. Simple Model of Continuous Choice .............................................. 1153
   C. Uncertain Legal Standards .......................................................... 1159
   D. Graphical Exposition of the Model ............................................... 1162
   E. Costly Dispute Resolution ........................................................... 1166
   F. Limitations of the Model ............................................................... 1169
      1. Large or Biased Error in Standards ............................................. 1169

* Professor of Law, University of California at Berkeley; Olin Visiting Professor of Law, University of Virginia.
When should punitive damages be awarded? How large should they be? Court practice is currently lawless in the sense that pre-

1. This Article was written for the Alabama Law Review symposium and conference on punitive damages, April 14, 1989, and for presentation to the John M. Olin School of Business, Washington University in St. Louis, 12 April, 1989.

The theory developed in this article builds upon my earlier article, Economic Analysis of Punitive Damages, 56 S. Cal. L. Rev. 79 (1982), which concerned when punitive damages should be awarded for purposes of deterrence, and some mathematical notes from my recent book, R. COOPER & T. ULEN, LAW AND ECONOMICS 363-97 (1988), which concern how large punitive damages should be. The pioneering article on the economic analysis of punitive damages is Ellis, Fairness and Efficiency in the Law of Punitive Damages, 56 S. Cal. L.
dicting punitive damages, even within a wide margin of error, is impossible in particular cases from knowledge of the law and a description of the facts. Juries are not receiving sufficiently definite instructions about punitive damages to keep their findings within reasonable bounds. To illustrate, the ratio of punitive damages to compensatory damages awarded by a jury in recent cases varies from as low as 1/10 to as high as 300,000/1, or even higher. As for absolute size, a jury award exceeding $11 billion against Texaco and in favor of Pennzoil, which included $3 billion in punitive damages, surprised the stock market sufficiently to cause Texaco’s stock to fall in value by significantly more than Pennzoil’s rose. Unpredictable damages are neither fair nor efficient. These

Rev. 1, 25-26 (1982) [hereinafter Ellis, Fairness and Efficiency]. I am grateful for comments from members of the University of Alabama Law School’s symposium and conference on punitive damages (especially Mark Grady), and the law and economics seminar of the John M. Olin School of Business, Washington University in St. Louis.

2. This problem is not new:

(P)rovisions of Magna Carta required that there be a reasonable, proportional, and sensible relationship between punishment and offense, and that the penalty exacted should not destroy the offender’s means of making a living in his trade. The Eighth Amendment to the United States Constitution, which prohibits excessive fines, was based on those provisions of Magna Carta.


3. In Roy Export Co., etc. v. Columbia Broadcasting Sys., Inc., 503 F. Supp. 1137 (S.D.N.Y. 1980), compensatory damages were $1 and punitive were $300,000 for the plaintiff’s common law copyright infringement claim. In Auburn Harpwell Asa’n v. Day, 438 A.2d 234 (Ms. 1981), a ratio of 60/1 was deemed reasonable. Consider that as compensatory damages approach zero and punitive damages remain at any positive number, their ratio approaches infinity. Of sixteen cases reported by Landes and Posner, the ratios of punitive damages to compensatory damages range from 10/1 to 21/1; of the six of these values reported in the tables that were not changed on appeal, the range is from 3.3/1 to 50/1. W. Landes & R. Posner, the Economic Structure of Tort Law 304-06 (1987) (Note that the ratios reported in the tables in Landes and Posner are actual to punitives; thus, they are the reciprocals of the ratios reported in this footnote.)

goals require, at a minimum, making punitive damages more predictable by making the law more exact.

To answer the questions, "When?" and "How much?", law must be shaped in light of underlying policy objectives and concepts of fairness. A consensus among legal scholars holds that deterrence, along with punishment, is the goal of punitive damages.\textsuperscript{a} Deterrence is, furthermore, a sufficiently well-defined goal to

Supreme Court ruled on Texaco's petition for writ of certiorari. See Texaco, Inc. v. Pennzoil Co., 748 S.W.2d 631 (Tex. Ct. App. 1988). The total value of the jury award, which, with prejudgment interest, exceeded $11 billion, was computed by Steve Bundy to be more than five times the largest previous judgment litigated on appeal. Bundy, Commentary on "Understanding Pennzoil v. Texaco": Rational Bargasging and Agency Problems, 75 Va. L. Rev. 335, 346 n.55 (1989) (citing MCI Communications Corp. v. American Tel. & Tel. Co., 708 F.2d 1081 (7th Cir.), cert. denied, 464 U.S. 891 (1983) (judgment set aside on appeal)).

The curious fact investigated by Cutler and Summers was why the combined value of the stock of Texaco and Pennzoil should have fallen by $3.4 billion dollars. Cutler & Summers, supra, at 158. Mnookin and Wilson try to explain this fact, and the fact that Pennzoil settled for $3 billion, as a rational bargaining outcome produced in part by the threat of bankruptcy and in part by a conflict of interest between Texaco directors and its shareholders. Mnookin & Wilson, supra, at 299-300.

5. E.g., K. Randem, supra note 2, at 23-24 ("In almost all jurisdictions of the United States where punitive damages are available, their stated purpose is nonremunerative. Their goal is to punish the wrongdoer and to deter others from similar conduct."); PUNITIVE DAMAGES: A GUIDE, supra note 2, at 15 ("In this book, the term 'punitive damages' is used to refer to those damages that are awarded to punish a party or to deter others in that party's position from similar conduct."); Sales & Cole, PUNITIVE DAMAGES: A Relic That Has Outlived Its Origins, 37 Vand. L. Rev. 1117, 1124 (1984) ("The first American enunciation of the theory of punitive damages occurred in 1791 [in Coryell v. Colbaugh, 1 N.J.L. 77 (1791)]. . . . The Coryell court also stated that damages should be 'such a sum as would mark their disapprobation, and be an example to others.'"); Note, Formulating Standards for Awards of Punitive Damages in The Borderland of Contract and Tort, 74 Cal. L. Rev. 2033, 2053 (1986) (hereinafter Note, Formulating Standards) ("Neither punishment nor compensation offers a complete rationale for awarding punitive damages in borderland cases [between torts and contract]. Deterrence provides the only completely satisfactory rationale for the imposition of punitive damages in these cases.")

In an interesting paper, Haddock, McChesney, and Spiegel argue that in thin markets where prices are negotiable, a party who expropriates another's asset and pays compensatory damages will extract the full surplus from exchange. There is, consequently, an incentive to convert property rights to liability rights if damages are based upon compensating the victim. To deter expropriation and redirect activities towards market exchange, compensation must be replaced with restitution as the principle for determining damages. See Haddock, McChesney & Spiegel, An Ordinary Economic Rationale for Extraordinary Legal Sanctions (unpublished manuscript 1989).

Epstein, however, dissents from the consensus and argues that retribution should be the goal to the exclusion of deterrence:

In the end the strongest, indeed the only, justification for punitive damages is, as the name itself suggests, as a substitute for the direct criminal sanctions that might be brought against defendants. As the criminal law functions best when it operates on
give definite form to law. It is surprising, therefore, to find an absence of useful prescriptions for using punitive damages to deter injurers. This Article develops such prescriptions from an economic model that predicts responses to changing legal incentives.

A recent study disclosed that eight of America's fifty states allow the award of punitive damages for gross negligence, while thirty-seven require intentional wrongdoing or a similar measure of moral culpability beyond negligence.\textsuperscript{*} Intentional wrongdoing and the like is distinguished by type, such as malice, intentional harm, wanton and reckless behavior, callous disregard for safety of others, deliberate exposure to undue risk, and so forth. The strength of the economic analysis of law consists, not in refining such distinctions, but in relating them to predictions about behavior. Definite predictions can be made about business defendants who are profit oriented, or more generally, about any rational deci-

---


6. A recent survey of the 47 states that allow recovery of punitive damages divided the conduct requirement into four general categories, as follows:

These general categories focus on the least egregious conduct required to be proved in order to obtain an award of punitive damages. From the most difficult, to the least difficult to prove, these general categories are 1) malice, 2) conduct exceeding gross negligence but not requiring malice, which is referred to in this book as strict conduct, 3) gross negligence, and 4) various statutory requirements. In the first category, fourteen states appear to require malice or a similar state of mind, and, in these states, acts of omission would not be sufficient to recover punitive damages. In the second category, twenty-three states require conduct exceeding gross negligence, but proof of malice is not required. In the third category, eight states require a showing of gross negligence. In the fourth category are the remaining two states that have enacted various statutes that govern the conduct which must be proven in different causes of action to obtain a recovery of punitive damages. As noted above, three states, Nebraska, New Hampshire, and Washington, do not allow the recovery of punitive damages.

PUNITIVE DAMAGES: A GUIDE, supra note 2, at 16-17; see also K. Redden, supra note 2, at 23 ("The manner or intention with which an act is committed is crucial to punitive damages."); Sales & Cole, supra note 5, at 1131 ("Those jurisdictions and commentators who recognize the validity of the punitive damage doctrine envision recovery when the defendant's actions either are or are deemed to be intentional.").

sionmaker motivated by private economic gain. Such a decisionmaker will not violate a legal standard intentionally unless the gain from doing so is, at a minimum, larger than the expected costs of liability.

Inadequate incentives for conforming to a legal standard usually result from imperfect enforcement of the law. A small imperfection, however, will not make deviation from the legal standard profitable. Instead, a large enforcement error is needed, which in turn makes the injurer's most profitable behavior fall far short of the legal standard. Consequently, plaintiffs should have to prove the fact and profitability of a gross shortfall from the legal standard in the face of compensatory damages before obtaining punitive damages from business defendants. I call this the "gross shortfall" standard.

In the absence of punitive damages, enforcement errors enable injurers to externalize a portion of expected social costs that they cause. Punitive damages should be set for the sake of deterrence at a level that eliminates the advantage of noncompliance and forces potential injurers to internalize the expected social costs of their actions. For example, if 1/2 of tort victims who are entitled to recover actually bring suit, and those who do bring suit recover in full, then, under certain conditions, 1/2 of the social costs caused by injurers will be externalized. In these circumstances, the ratio of total damages to compensatory damages paid by intentional wrongdoers should equal 2/1 for the sake of deterrence. In general, the punitive multiple should equal the reciprocal of the enforcement error for the sake of deterrence, which I call the "rule of the reciprocal."  

The gross shortfall standard answers for business defendants the question, "When should punitive damages be awarded for the sake of deterrence?" The rule of the reciprocal answers the question, "How much should be awarded as punishment?" A law of punitive damages that incorporated these proposed rules would

---

7. This principle is understood by writers who take the economic approach to torts. See, e.g., Ellis, Fairness and Efficiency, supra note 1, at 48-49. A numerical example of the reciprocal rule appears in W. Landes & R. Posner, supra note 3, at 160 n.12, and S. Shavell, Economic Analysis of Accident Law 131 example 6.2 (1987).

8. The reciprocal of 1/2 equals 2/1, the reciprocal of 1/4 equals 4/1, etc. Note also that if \( R \) is the ratio of punitive to compensatory damages, then the "punitive multiple" as the term is used in this Article equals \( R + 1 \).
serve the goal of deterrence, at least in principle. The Article proceeds to develop the economic model in Part I, and its legal implications are explored in Part II.

I. ECONOMIC MODEL OF PUNITIVE DAMAGES FOR DETERRENCE

The economics of punitive damages for deterrence are first explained in a simple model of binary choice. Next, the model is reconsidered when choice is continuous. Finally, the continuous model is reconsidered when legal standards are uncertain.

A. Simple Model of Binary Choice

Suppose that a person can either take precaution, which completely prevents accidents, or not take precaution, in which case the probability of an accident is predictable, say \( \frac{1}{2} \). Precaution is costly to the potential injurer, say 110, but an accident is even more costly to its victims, say 600. The expected cost of the harm thus equals \( \frac{1}{2} \times 600 = 200 \). Since injurers can save victims 200 in expected harm at a cost of 110, efficiency requires them to take precaution. It can be argued, furthermore, that injurers should take precaution in these circumstances because fairness requires them to give similar weight to their own costs and the costs they impose upon others. This principle is, arguably, a social norm that the law expects reasonable people to respect.

Suppose that injurers are held liable for only fifty percent of the accidents caused by their failure to take precaution. Victims may fail to recover in the other fifty percent of cases for a variety of reasons—they do not know that they were injured, they do not know who injured them, they know who injured them but cannot prove it, or they cannot prove that the injurer violated the legal standard. The enforcement error in this example is thus one-half. The injurer's expected cost of liability from not taking precaution equals the expected social harm multiplied by the enforcement error, or \( 200 \times (\frac{1}{2}) = 100 \). Suppose, furthermore, that the injurer's reputation is not tarnished, the demand for injurer's products is not reduced, nor is there any cost except liability that is born by the injurer as a result of accidents. So the law requires the injurer to pay 110 to avoid expected liability of 100.
A reasonable person would, presumably, conform to the legal standard even though doing so costs a little more than the liability it avoids. A reasonable person who fails to conform to the legal standard probably makes a mistake. For example, the injurer might not conform because she mistakenly thinks that accident victims suffer costs of 300, rather than 600. Her underestimation will be corrected when courts impose compensatory damages. The decision of the court will make the law’s requirement emphatically clear to injurers, even without punitive damages, which is all that is normally required to induce reasonable people to conform to the law.

The legal idea of a reasonable person, however, is not the same as the economic ideal of a purely self-interested decisionmaker. A self-interested decisionmaker will violate the legal standard intentionally when conforming avoids expected liability of 100 at a cost of 110, assuming that liability is the only cost of violating the standard. Deterring the purely self-interested decisionmaker requires supplementing compensatory damages with punitive damages to overcome the enforcement error.

When compensatory damages are augmented by punitive damages, the punitive multiple, by definition, equals the ratio of total damages to compensatory damages. Suppose that the punitive multiple exactly offsets the enforcement error. The punitive damages thus bring expected liability into line with expected social harm, as required for social efficiency. To illustrate, the enforcement error in the preceding example equals \( \frac{1}{2} \), so the punitive multiple should equal 2/1. The injurer’s expected liability, including compensatory and punitive damages, equals \( (\frac{1}{2}) \times \frac{2}{1} \times 200 = 200 \), which also equals the expected cost of social harm.

The purely self-interested decisionmaker, knowing that she faces punitive liability for intentionally violating the legal standard, now finds that conforming to the legal standard costs 110, whereas not conforming costs 200. In general, if the punitive multiple equals the reciprocal of the enforcement error, the one exactly offsets the other, and the injurer’s expected liability equals the expected social harm. The decisionmaker thus sweeps the full social costs of not taking precaution into her calculus of self-interest, which compels her to take precaution when required for efficient deterrence.
The rule of the reciprocal can be explained with the help of Judge Hand's formula for determining liability in tort cases. The Hand Rule finds negligent the injurer who failed to take precaution if the burden $b$ of precaution is less than the product of the probability of an accident $p$ and the loss $L$ resulting from it, but not otherwise:

\[
\begin{align*}
    b < pL & \Rightarrow \text{liable} \\
    b \geq pL & \Rightarrow \text{not liable.}
\end{align*}
\]

Self-interest compels potential injurers to satisfy the negligence standard whenever precaution costs less than the resulting fall in expected liability. So long as the Hand Rule standard of negligence is enforced without error, potential injurers will find that conforming to it is the cheapest decision for them.

Suppose, however, that enforcement error reduces expected liability for compensatory damages, leaving liability only a fraction $q$ of expected social costs. The purely self-interested decisionmaker, whose only cost of nonconforming is the expected liability, decides whether or not to take precaution by comparing precaution costs $b$ and expected liability $pLq$ to see which is smaller. If the fraction $q$ is small enough, the decisionmaker will minimize costs by intentionally violating the standard of negligence defined by the Hand Rule. To correct this error, a punitive multiple $m$ can be applied to damages, where $m = 1/q$. Now the potential injurer decides whether or not to take precaution by comparing $b$ and $pLqm$, where $qm = 1$. Since $m$ and $q$ cancel each other, this is equivalent to comparing $b$ and $pL$, as required to induce conforming to the Hand Rule standard of negligence.

The enforcement error sometimes can be inferred when it cannot be observed directly. A self-interested decisionmaker is on the margin of indifference between conforming to the standard and violating it when the cost of the former equals the expected cost of the later. It is easy to show that the enforcement error for the marginal decisionmaker will equal the ratio of costs of precaution to the expected harm resulting from its absence. This rule of thumb can be used to compute the punitive multiple under certain condi-

---

10. This fact is easy to see with the help of some notation. Let $b$ equal the cost of precaution, let $q$ equal the enforcement error, and let $d$ equal compensatory damages. For the marginal injurer in the absence of punitive damages, $b = q \cdot d$, which implies $q = b/d$. 
tions when the enforcement error cannot be proved directly to the court.

To illustrate, retain the assumption in the preceding example that injurer's precaution avoids accidents for which the expected social harm is 200 and the enforcement error is \( \frac{1}{2} \), so expected liability for compensatory damages equals 100. Now modify the preceding example to assume that the costs of precaution are 100, instead of 110. A self-interested decisionmaker faced with spending 100 to avoid expected liability costs of 100 is on the margin of indifference, as indicated by the equation \( 100 = 200 \times \frac{1}{2} \). This equation can be rewritten as \( \frac{1}{2} = 100/200 \), whose interpretation is that enforcement error equals the ratio of costs of precaution to the expected harm from not taking precaution.

The preceding numerical example assumes that injurers are held fully liable for half of the injuries that they cause. The same results can be obtained by assuming instead that injurers are held liable for half of the harm suffered in all the injuries that they cause.\(^{11}\) The rule of the reciprocal is correct regardless of whether the enforcement error arises from compensation that is too little or too infrequent. The enforcement error in either case is \( \frac{1}{2} \), and the punitive multiple should be 2/1 for the sake of deterrence in both cases.\(^{12}\)

The example discussed so far concerns failure to take precaution, which exposes others to excessive risk. Another type of wrongdoing causes certain harm, not just the risk of it. Punitive

---

11. There is a range of views on the adequacy of compensation under current law. E.g., Sales & Cole, supra note 8, at 118 ("The quest to bestow increasing compensation no longer can justify punitive damage awards because actual damages currently recoverable compensate plaintiffs more than adequately for every conceivable element of physical, emotional, or imagined injury."); Sebert, Punitive and Nonpecuniary Damages in Actions Based Upon Contract: Toward Achieving the Objective of Full Compensation, 33 UCLA L. Rev. 1855, p. 1570 (1986) ("I argue for a broader and more overt recognition of both nonpecuniary loss and punitive or supracompensatory damages in contract.") (footnotes omitted).

As a general rule, exemplary damages may not be recovered in an action based on breach of contract. The two historical exceptions cited are contracts to marry and public service contracts. A recent article advocates allowing punitive damages if there is fraudulent breach of contract duty or fiduciary duty, or breach of implied covenant of good faith and fair dealing. Note, Formulating Standards, supra note 5; see also Note, Tort Remedies for Breach of Contract: The Expansion of Tortious Breach of the Implied Covenant of Good Faith and Fair Dealing Into the Commercial Realm, 86 Colum. L. Rev. 377 (1986).

12. It is preferable, however, to correct this kind of enforcement error, not by awarding punitive damages, but by expanding compensatory damages until they are fully compensatory.
damages are often awarded when the actor intentionally harms others for personal gain. For example, a survey by the American Bar Foundation found that local courts in Chicago most often award punitive damages in cases of fraud.\textsuperscript{13} Fraud sometimes results from an omission. To illustrate, an accountant may commit fraud by failing to disclose information as required by law. Failing to disclose information is similar to failing to take precaution. Alternatively, fraud may consist in a commission, as when a dishonest bookkeeper diverts funds and disguises this fact. Even with commissions, the theory outlined above fits closely. Fraud is an intentional wrong undertaken for personal gain, which meets the test invoked above for imposing punitive damages. Furthermore, the victim of fraud may not know it, or the victim may not know who did it, or the victim may not be able to prove who did it. For these reasons, there will be enforcement error.\textsuperscript{14} Commissions and omissions are not very different from a technical standpoint, so a model of punitive damages that fits one also fits the other.\textsuperscript{15}

\textbf{B. Simple Model of Continuous Choice}

Although some forms of precaution are binary, others present a range of choice that is more nearly continuous. Examples are driving speed, frequency for testing railroad tracks, sampling ratios when searching for defective products, expenditure on quality control in manufacturing, expenditure on improved safety of product designs, and severity of warning about product hazards. This section of the Article extends the propositions developed for binary

\begin{itemize}
  \item \textsuperscript{13} S. Daniels, Empirical Patterns in Punitive Damage Cases: A Description of Incidence Rates and Awards (American Bar Foundation Working Paper 8705, 1987).
  \item \textsuperscript{14} Whether the punitive multiple in actual cases of fraud approximately equals the enforcement error is a question that needs to be investigated.
  \item \textsuperscript{15} This is easy to see with the help of some notation. Let $b$ equal the cost of precaution, let $p$ equal the probability of an accident, let $L$ equal the social harm from an accident, and let $q$ equal the enforcement error. The expected social harm is $pL$. Social efficiency requires the injurer to take precaution if $pL > b$. If injurer's liability is limited to compensatory damages, her expected liability equals $pLq$. Cost internalization is achieved by setting the punitive multiple $m$ equal to $1/q$, so that the injurer's expected liability equals expected social harm, $pL = m q p L$. The marginal injurer under a regime of compensatory damages is the person for whom $b = pLq$, which implies $q = b/pL$, and thus $m = pL/b$. These are the propositions about the simple binary model developed in the text.
  
  For commissions, retain the preceding notation, except let $g$ indicate the gain (e.g. the money stolen by fraud), and replace $b$ with $g$ in all the preceding propositions.
\end{itemize}
choice to continuous choice, beginning with a background rule of negligence.

An injurer who conforms to the legal standard imposed by a negligence rule escapes liability, whereas an injurer whose care falls just below the legal standard is fully liable for the victim’s accident costs. The injurer’s expected costs typically jump at the point where his care dips below the legal standard. Costs jump abruptly because the injurer has no liability when his precaution exactly equals the legal standard, whereas he has full liability for compensatory damages when his precaution falls just below the legal standard.

The jump in costs at the legal standard can be understood by extending a famous example. Suppose that a railroad locomotive emits sparks that sometimes set fire to fields bordering the track. Suppose further that the legal standard requires installing a spark arrester, which cuts in half the emission of sparks and the resulting fires. Under a negligence rule, the railroad company that did not install spark arresters would be liable for the fires that would have been avoided by installing a spark arrester. In reality, however, it is difficult or impossible to distinguish between fires that a spark arrester would have prevented and fires that would have been caused by sparks passing through a spark arrester. In practice, the negligent railroad is likely to be held liable for all or nearly all of the fires caused by its sparks.

Although a negligent railroad will probably be liable for one hundred percent of the damage from fires set by sparks emitted from its trains, a railroad that installed spark arresters would escape all liability under a negligence rule. By installing spark arresters, the railroad avoids bearing the cost of the fires prevented by the spark arresters as well as the fires not prevented by it.


17. This example is taken from Coase, The Problem of Social Cost, 3 J. LAW & ECON. 1, 29-35 (1960) (example taken from A. C. Pigou, THE ECONOMICS OF WELFARE (4th ed. 1932)).

18. For evidence from cases that this generalization is correct, see Grady, Punitive Damages and Subjective States of Mind: A Positive Economic Theory, 40 ALA. L. REV. 1197 (1988).
fact that conforming to the legal standard enables the injurer to escape the cost of accidents that do occur, not just the cost of accidents prevented by his care, causes the jump in injurer's expected costs at the legal standard.¹⁹

The problem raised by this example can be stated abstractly. Reasonable care does not ordinarily eliminate accidents completely. Assume that the probability of an accident is still positive when care exactly equals the reasonable level, say the probability is .02. If care falls a little below the legal standard, the probability of an accident rises, say to .03. The failure to take reasonable care causes the probability of the accident to rise by .01. However, it may be impossible for most or all plaintiffs to show whether a particular accident was one of the .01 that would have been avoided by reasonable care.

In the example, the negligent railroad is liable for some harm caused by its activities, even though the harm would not have been prevented by conforming to the legal standard. Similarly, an injurer whose precaution falls below the legal standard may be held liable for harm not even caused by its activities.²⁰ To illustrate, suppose that a defective design causes contamination of food by a carcinogenic chemical. Some people who ate the contaminated food may develop cancers for entirely unrelated reasons. Given a high level of uncertainty about the origins of cancer, a court may allow recovery by everyone who ate the contaminated food and subsequently developed the types of cancer that it causes.

Imperfection in the court's causal attribution may be due to the fact that some of the needed information is difficult or impossible to obtain, as suggested in the preceding examples. On the other hand, the problem may not be insufficient information so much as a connection that is probabilistic by its very nature.²¹ The difference being discussed is between subjective uncertainty and objective probability. To illustrate, suppose that exposure to a certain chemical increases the probability of cancer but does not raise the probability to one for anyone. When a suit arises, it may be

¹⁹. The rule of law is, in effect, that a non-negligent injurer is not liable, and a negligent injurer is strictly liable for accidents caused by his activity, whether or not they were caused by his negligence.

²⁰. I am grateful to George Priest for this point.

²¹. This possibility assumes that probabilistic connections can be objective facts about the world, not just subjective reports of our uncertainty about it.
impossible even in principle to decide whether this accident would have been avoided by taking efficient care, regardless of how much information is available to the court. It would be just as misleading in such a case to say that exposure caused a person's cancer as to say that it did not cause the cancer.

From the viewpoint of legal policy, it does not matter whether the imperfection in causal attribution is inherent or due to incomplete information. In either case, plaintiffs will not recover if required by the court to prove that reasonable care would have prevented the accident. The court may respond by shifting the burden of proof and requiring the defendant to prove that his negligence did not cause the injury. If the court responds in this way, the injured's costs will jump at the level of care corresponding to the legal standard.

There are other facts about legal standards besides imperfect causal attribution that make costs jump. To illustrate, consider a class of injuries in which disputes arise not over liability, but over the extent of damages. An injurer whose behavior falls below the legal standard must bear the cost of litigating damages, whereas an injurer who conforms to the legal standard has no claims brought against him by assumption. Thus the injurer's litigation costs jump at the legal standard of care. Alternatively, an injurer whose precaution is found by courts to fall below the legal standard will be declared a wrongdoer. The resulting adverse publicity may be worse than if the court vindicated the defendant. Reputation effects can thus cause a jump in costs at the legal standard. In addition, many business executives care about reputation apart from its effects upon profits.

These examples show that a multiplicity of causes contribute to the clustering of costs at the legal standard. In economic jargon, violating a legal standard has an element of fixed costs that produce a discontinuity in the injurer's cost function at the legal standard.22 By conforming exactly to the legal standard, injurers

22. For the view that the discontinuity is smaller and less significant than I consider it to be, see Mark Grady's comment on this Article in this issue, supra note 18. Also see the following articles by him: Untaken Precautions, 18 J. LEGAL STUD. 139 (1989); Discontinuities and Information Burdens: A Review of The Economic Structure of Tort Law by William M. Landes and Richard A. Posner, 56 GEO. WASH. L. REV. 658 (1988); Common Law Control of Strategic Behavior: Railroad Spots and the Farmer, 17 J. LEGAL STUD. 15 (1988); Why Are People Negligent? Technology, Nondurable Precautions, and the Medical Malpractice Explosion, 82 NW. U.L. REV. 293 (1988).
typically keep their costs at the lowest possible level. The typical injurer, furthermore, is not balancing costs and benefits of additional precaution at this point. Rather, the benefit from taking a little less precaution is much less than the cost. A decisionmaker whose care falls below the legal standard in these circumstances, having failed to minimize costs, must have made a mistake, possibly due to imperfect information. Enforcing the law and imposing compensatory damages should bring the mistake to the notice of the decisionmaker and cause him to correct his ways in the future. There is, consequently, no need for punitive damages for purposes of deterrence.

Suppose, however, that the liability rule or its enforcement is imperfect. For example, suppose that some accident victims do not recover from negligent injurers, or they recover but not fully. It might be supposed that purely self-interested injurers will reduce their precaution by a small amount in response to a small enforcement error, but this supposition is mistaken. Recall that an injurer’s costs jump abruptly when his care falls below the legal standard. A small enforcement error is not enough to overcome this jump in costs. To induce the cost minimizing injurer to violate the legal standard, the enforcement error must be large enough to overcome the abrupt jump in costs. If the enforcement error is large enough to overcome the jump in costs, the purely self-interested decisionmaker will choose a level of precaution less than the legal standard. It typically pays him to reduce precaution, not a little, but a lot. To see why, consider that the disadvantage of violating the standard is the abrupt assumption of liability, whereas the advantage is the savings from less expenditure on precaution. To justify assuming liability, the savings in precaution must be substantial, which requires cutting back a lot on precautionary expenditures. That is why precaution falls a lot, not a little.

Behavior that falls far short of the legal standard probably does so intentionally, because of a deficiency in legal incentives. This deficiency can be corrected by punitive damages that overcome the enforcement error. Thus a gross shortfall from the legal standard is the criterion for imposing punitive damages on businesses and other economic decisionmakers. Owen distinguished cases according to the types of defendants, states of mind, and the
nature of the misdeed. This Article suggests that for a certain type of defendant the nature of the misdeed may be sufficient, without inquiring into state of mind.

The gross shortfall is caused by a large enforcement error that makes expected liability substantially less than the social cost of accidents. If punitive damages equal the reciprocal of the enforcement error, the effects of the latter will be offset and the incentive to conform to the law will be restored.

The importance of enforcement error in an economic analysis of punitive damages was recognized by Ellis. In criticizing Ellis, Owen noted that there is a positive probability that the injurer will escape liability in almost any tort case, so there is always enforcement error. The logical implication is that modest punitive damages would be appropriate in almost every tort case. This argument fails to take account of the jump in cost at the legal standard and its effect upon intentional violation of a legal standard. Even with enforcement error, most decisionmakers find that conforming to the legal standard is cheaper than violating it.


24. As before, let \( x \) denote injurer's precaution, let \( b \) denote the unit cost of precaution, let \( p(x) \) denote the probability of an accident, let \( L \) denote the social harm, let \( q \) denote the enforcement error, and let \( m \) denote the punitive multiple. The injurer's expected costs in the zone of liability can be written

\[
EC = bx + p(x)Lq\mu.
\]

Social costs \( SC \) are the same as injurer's expected costs but for enforcement error \( q \) and punitive multiple \( m \):

\[
SC = bx + p(x)L.
\]

If the punitive multiple equals the reciprocal of the enforcement error, \( m = 1/q \), the injurer's expected costs reduce to the social costs.

25. There are three situations in which punitive damages might produce efficient outcomes: (1) when the \textit{ex ante} probability of Mal being held liable for the loss is less than the actual probability of the harm; (2) when the compensatory damages for which Mal is held liable are less than the actual amount of the loss; and (3) when the subjective cost of avoidance (the cost as Mal appraises it) is greater than the cost recognized by the law.

Ellis, Fairness and Efficiency, supra note 1, at 25.

Some injurers are better positioned than others to avoid liability, so the enforcement error may be different from one injurer to another. For most injurers, enforcement error is presumably not large enough to make violating the legal standard profitable. The institution of punitive damages will have little or no effect on their behavior. However, some potential injurers will be at the margin of indifference between conforming and grossly violating the legal standard. Punitive damages will affect their precaution a lot. Insofar as legal standards are certain, the institution of punitive damages generally causes a few decisionmakers to increase their precaution a lot and has little effect on most actor's precaution. The response, in economic jargon, is on the extensive margin not the intensive margin.

As in the case of binary precaution, the enforcement error can be inferred sometimes when it cannot be observed. For the injurer on the margin between conforming to the legal standard and violating it, the enforcement error equals the ratio of the additional costs of precaution needed to satisfy the legal standard and the expected cost of the harm from accidents when the standard is violated.\(^{27}\) Furthermore, as with binary precaution, the model of continuous precaution is easily extended to continuous acts of commission, like fraud.\(^{28}\)

C. Uncertain Legal Standards

This argument retains much of its force, although not its simplicity, when some implicit assumptions are relaxed. An implicit assumption in the preceding analysis is that injurers know with certainty the legal standard that courts will apply to them. A more

27. Using the same notation as in the preceding footnote, the injurer who conforms to the legal standard faces expected costs of \(b x^*\), whereas the nonconforming injurer who chooses precaution \(x^*\), where \(x^* > x^*\), faces expected costs of \(b x^* + p(x^*)/L\). These costs are equal for the marginal injurer, so solve the resulting equation to obtain

\[ q = (x^* \propto x^*)/b/p. \]

28. This model is easily adapted to accommodate acts that are harmful, not just risky. Intentionally harming someone else is usually a tort and often a crime. The actor who intentionally harms another exposes himself to the risk of civil liability and, possibly, criminal prosecution. There is a threshold that separates the forbidden, harmful activity from other activities. The injurer's expected costs consequently jump abruptly at this threshold, just as with the negligence model. Thus the simple model of continuous choice, which covers precaution, also extends to fraud. The discontinuity analysis is extended to crimes more generally in Cooter, Prices and Sanctions, 84 Colum. L. Rev. 1823 (1984).
realistic assumption is that uncertainty surrounds the legal standard of care. For example, the requirement that precaution should be “reasonable under the circumstances” is a distinction without a bright-line boundary, like the difference between your face and the back of your head. As for the facts, their determination is affected by the gullibility of juries. Given uncertainty, the injurer’s expected liability does not jump abruptly at the legal standard as in the simple model. Rather, the more precaution the injurer takes, the more likely courts are to find that he satisfied the legal standard.

Injurers typically respond to moderate uncertainty about legal standards by increasing their precaution, so that they moderately exceed the legal standard, as proved in the technical appendix. By exceeding the legal standard, injurers give courts a margin of error within which to find “no liability.” Since uncertain standards typically cause injurers to exceed the legal standard as they perceive it, a court finding that the defendant’s precaution fell moderately short of it should be attributed to mistake, either on the part of the defendant or the court. Punitive damages are inappropriate in such circumstances for purposes of deterrence; indeed, compensatory damages already provide incentives for overdeterrence.

Suppose, however, that enforcement errors of the kinds discussed in the simple model are present (some victims do not sue or victorious plaintiffs do not recover fully). Some potential injurers who are well positioned to evade liability may find that violation of the legal standard is profitable. Indeed, those potential injurers who are on the margin of profitability between conforming and violating when the legal standard is certain will tip in favor of violating it when the legal standard becomes moderately uncer-

29. See infra Technical Appendix. The technical appendix proves that, under certain conditions, a mean-preserving spread in the distribution of court errors causes injurers to increase their precaution. In general, however, it is possible for uncertainty to cause either too much or too little precaution. See Craswell & Calfee, Deterrence and Uncertain Legal Standards, 2 J. L. Econ. & ORGANIZATION 279 (1986); Johnston, Bayesian Fact-Finding and Efficiency: Toward an Economic Theory of Liability Under Uncertainty, 61 S. CAL. L. REV. 137 (1987), Part II; Calfee & Craswell, Some Effects of Uncertainty on Compliance With Legal Standards, 70 VA. L. REV. 965 (1984). Empirical research on this question is badly needed. Viscusi has shown that consumers and workers tend to overreact to risk. See Viscusi, Compensating Workplace Toxic Torts, in NEW DIRECTIONS IN LIABILITY LAW 126 (The Academy of Political Science 1988).
tarn. And their violation must be gross in order to be profitable. The gross shortfall standard is, consequently, a good criterion for punitive damages in the presence of moderate, random errors in applying legal standards. Furthermore, the argument that the punitive multiple should equal the reciprocal of the enforcement error still applies in the presence of uncertain legal standard for the same reasons given before (although the argument has a new wrinkle).
D. Graphical Exposition of the Model

The model of continuous choice can be elucidated with the help of graphs, which the reader with an aversion to technique may skip. The potential injurer’s level of precaution is measured on the horizontal axis in Figure 1; the social cost of accidents is measured on the vertical axis. The amount of precaution, denoted $x$, multiplied by the cost per unit of precaution, denoted $b$, equals the total cost of precaution $bx$. The probability of an accident, denoted $p(x)$, multiplied by the social loss from an accident, denoted $L$, equals the expected social loss $p(x)L$. The social cost of accidents are taken to be the sum of the costs of precaution and the expected harm from accidents for purposes of this simple model, or $bx + p(x)L$. The curve in Figure 1 initially slopes down, which indicates that, as precaution increases, its cost is less than the resulting savings from fewer accidents. As more precaution is taken, however, diminishing returns set in; a point is reached where the cost of further precaution exceeds the additional savings from fewer accidents, so the curve slopes up. There is, consequently, a level of precaution, denoted by $x^*$, where social costs are minimized.

**Fig. 1: Social Cost of Accidents**

![Diagram showing social cost of accidents with variables and equations]

-Injurer’s Precaution

$x =$ injurier’s precaution  
$p =$ probability of an accident  
$b =$ cost per unit of  
$L =$ cost of social harm from an accident

punitve, supercompensatory damages can induce the potential defendant to choose the socially optimal level of care.” *Id.* at 1395.

32. The representation of social costs shown here was originally developed by Brown in his pioneering paper. See Brown, supra note 16. The analysis of the discontinuity was developed by Cooter. See Cooter articles cited supra note 16.
Efficiency considerations suggest that the legal standard of care should equal the level of precaution \( x^* \) that minimizes social cost. (Indeed, the Hand Rule guarantees it.)\(^{33}\) To implement the legal standard \( x^* \), a negligence rule divides Figure I into a permitted zone and a forbidden zone, as depicted in Figure 2. As long as a potential injurer stays in the permitted zone where \( x \geq x^* \), he only bears the cost of his own precaution, which is indicated by the straight line \( bx \) through the origin of the graph. If his behavior falls into the forbidden zone where \( x < x^* \), however, he is legally liable. In the event that there is no enforcement error, his expected liability in the forbidden zone equals the expected social cost of the harm caused by accidents. Thus the injurer's costs in the forbidden zone, assuming no enforcement error, are indicated by the solid curved line in Figure 2, whereas his costs in the permitted zone are indicated by the solid straight line. Notice that the potential injurer's costs are indicated in Figure 2 by the solid lines that consist of two distinct pieces.

**Fig. 2: Negligence Rule**

The potential injurer's costs jump abruptly at the border between the forbidden and permitted zones in Figure 2, when he passes from one solid line to the other. The lowest point on the

---

solid lines, where the injurer’s costs are minimized, is achieved where the injurer’s precaution equals the legal standard $x^*$. The marginal costs and benefits of a small reduction in precaution are not equal at this point. In technical jargon, the injurer is at a corner of his cost function.

If some victims do not recover fully, the injurer’s cost curve in the forbidden zone of Figure 2 shifts down, as depicted in Figure 3. It is easy to prove that the lowest point on this curve moves to the left. For a small error, the injurer’s lowest costs in the forbidden zone, which occur at precaution level $x^*$, are still greater than his costs at $x^*$ in the permitted zone. For a large error, however, his lowest costs in the forbidden zone, which occur at precaution level $x^-$, are lower than his costs at $x^*$ in the permitted zone. Notice, also, that $x^-$ is a large distance from $x^*$. So a small enforcement error does not cause the rational injurer’s precaution to fall below the legal standard, whereas a large enforcement error causes the rational injurer’s precaution to fall far below the legal standard.

A punitive multiple equal to the reciprocal of the enforcement error restores the injurer’s expected cost curve to the expected social cost curve in Figure 3, as required for cost internalization.

**Fig. 3: Enforcement Error**

34. As in the preceding footnote, the injurer’s expected costs in the zone of liability are written

$$EC(x, q) = bx + p(x)Lq.$$ 

In the absence of punitive damages, $m = 1$. Assuming an internal solution, the cost minimizing $x$ satisfies the following equation for all $q$:

$$0 = EC_i = b + p'(x)Lq.$$
Figure 4 modifies Figure 2 to show the effect of moderate, unbiased error in applying legal standards. As the injurer's precaution increased, the probability of the court finding "no liability" increases, so the injurer's expected costs dip below the "cost when liable" curve and approach the "cost when not liable" line. These changes in probability are fastest near the true legal standard \( x^* \), where judgments require the most discernment. The jump in injurer's expected costs at \( x^* \) in Figure 2 has been smoothed out in Figure 4. The cost-minimizing level of precaution, \( x^* \), exceeds the legal standard \( x^* \), which provides a margin for court error.

**Fig. 4: Moderately Uncertain Standard**

![Graph](image)

Injurer's Precaution

Intentional violation of the legal standard does not pay when costs are depicted as in Figure 4. By changing the graph a situation

Fully differentiate the preceding equation yields

\[
\frac{dx}{dq} = (-EC_{11})^{-1} p'(x)L
\]

\( EC_{11} > 0 \) by the second order condition and \( p' < 0 \) by the assumption that precaution reduces accidents. Thus injurer's optimal precaution increases when \( q \) increases, which indicates a decrease in the enforcement error.

35. A court error not discussed here is setting the legal standard of care too high or too low. That is not, however, an enforcement error, so it is not discussed in the Article. For a discussion in the same framework as this Article, see Cooter, *Prices and Sanctions*, supra note 29.

36. In technical terms, the discontinuity is gone but the non-convexity persists.
could be depicted in which violating the legal standard pays, provided that liability is limited to compensatory damages. As enforcement error becomes larger, liability is more and more a matter of luck, unrelated to level of precaution, which gives no incentive for care. It can be shown that as the error in applying the legal standard increases ("mean-preserving spread in the error distribution"), a point is reached where the injurer's cost-minimizing level of precaution falls below the legal standard.37

E. Costly Dispute Resolution

The costs of resolving disputes, which have not been discussed so far, create additional complexities for the model without altering its conclusions. Disputes can be divided into three stages for purposes of analysis: asserting a claim, settlement bargaining, and trial. Each stage in the process is controlled in part by the parties' expectations about what will happen in the subsequent stages. Thus the rational plaintiff decides to assert a claim because the cost of doing so is less than the expected payoff from settlement or trial. Similarly, the plaintiff gauges his demand in settlement bargaining based in part upon the judgment he expects in the event of a trial.

A thorough analysis of the process of resolving disputes is beyond the scope of this Article, but some comments based upon the economic analysis are appropriate.38 Given the dependency of prior stages upon subsequent stages, the discussion should proceed from trial and work back to the assertion of claims. The institution of punitive damages creates additional rights in plaintiffs which they can assert or not assert. In any trial, consequently, the institution cannot weaken, and may strengthen, the plaintiff's hand. As a consequence, the expected judgment from trial, viewed ex ante before the trial actually begins, cannot decrease and may increase as a consequence of punitive damages. In addition, the institution raises the stakes and creates additional claims to be decided, which makes trials longer and more complicated. The expected judgment

37. See infra Technical Appendix.
increases for the plaintiff, and the cost of trial increases for both parties.

The increase in the expected judgment strengthens the plaintiff's hand in settlement bargaining. The terms of settlements should, consequently, shift in favor of accident victims and other plaintiffs. It is uncertain, however, whether the institution increases or decreases the frequency with which disputes can be settled out of court. Raising the costs of trial and the stakes at risk make disputants more eager to settle. The greater uncertainty created by the institution of punitive damages, however, makes settlement bargaining more difficult. It is impossible to determine which effect is stronger without gathering empirical data that is currently unavailable.

Since the institution of punitive damages increases the expected net value of trial to the plaintiff and also strengthens the plaintiff's hand in settlement bargaining, accident victims are almost certainly more likely to assert claims as a result of their power to ask for punitive damages. Punitive damages thus help to overcome the tendency toward underenforcement due to the high cost of filing a complaint and proceeding with a legal dispute. These facts, however, give potential injurers extra incentives to take precaution, which results in fewer accidents, in turn resulting in fewer claims. With the assertion of claims, as with the frequency of settlements, it is impossible to determine without empirical data which of the two effects working in opposite directions is stronger.

The tendency toward underenforcement, due to the high cost of the legal process, may be more than overcome by forces tending towards over enforcement. The possibility of overenforcement is illustrated by nuisance disputes. Nuisance disputes are disputes without merit, which can be defined for purposes of discussion as disputes whose expected judgment in the event of trial is zero or

42. It is possible in principle for punitive damages to increase trial costs by so much that the expected value of pressing any particular claim falls rather than rises, but this possibility seems remote. The underlying logic is developed in Katz, supra note 39.
close to it. The only reason for the plaintiff to pursue such a dispute is asymmetries in the cost of proceeding. If the cost of defense at trial, for example, is far larger than the plaintiff’s cost, or if delay in resolving the dispute is far more costly to defendant than plaintiff, then it may be rational for the defendant to settle on positive terms even though plaintiff’s expected judgement at trial is nil. Punitive damages may create or aggravate asymmetries in the cost of the dispute to the parties. For example, the possibility of punitive damages may require disproportionate increases in the cost of expert witnesses to the defendant. Or punitive damages may increase the cost of defendant’s response to discovery requests. Such asymmetries give defendants reason to settle claims of little merit.

So far the discussion of the costs of dispute resolution has concerned predictions. There is, however, an important normative issue that relates to the model developed in this Article. Injurers cause costly accidents and accidents cause costly disputes. The internalization of social costs by injurers, as required for efficient deterrence, must encompass the harm done by accidents. Must the cost of resolving disputes also be born by injurers for the sake of efficient deterrence? In particular, are the plaintiff’s costs of proceeding with the dispute part of the “enforcement error” that the punitive multiple should capture?

Allowing injurers to escape paying the costs of resolving disputes externalizes part of the costs that they impose upon others, which typically, although not inevitably, gives them insufficient incentives for precaution. However, requiring defendants to pay the litigation costs of successful plaintiffs, without requiring unsuccessful plaintiffs to reimburse the litigation costs of defendants, creates an asymmetry that would open the doors far wider for nuisance

43. The definition of nuisance suits requires refinement in a fuller theory. To illustrate, Doug Leslie prefers defining them as suits in which the expected judgment is less than the plaintiff’s litigation costs.

44. See R. Cooter & T. Ulen, supra note 1, at 486; Bebchuk, Suing Solely to Extract a Settlement Offer, 17 J. LEGAL STUD. 437 (1988); Cooter & Rubinfeld, supra note 38; Rosenberg & Shavell, A Model in Which Suits Are Brought for Their Nuisance Value, 5 INT’L REV. L. & Econ. 3 (1986).

45. It can be shown that this element of externalized cost can result in too much or too little precaution by injurers, at least in principle, but the conditions for the former effect are unlikely to be met in practice. See Polinsky & Rubinfeld, The Welfare Implications of Costly Litigation for the Level of Liability, 17 J. LEGAL STUD. 151 (1988).
disputes. These comments are the preface to a large topic that is beyond the scope of this Article.

F. Limitations of the Model

The gross shortfall standard and the rule of the reciprocal have been explained using words and graphs. The explanation was based upon a simple model that is sufficient for dealing with a broad range of cases. There are, however, circumstances in which this model is inadequate. Limitations in the simple model of deterrence and the two prescriptions derived from it are spelled out in this section.

1. Large or biased error in standards.—The preceding account of uncertain legal standards assumes that errors are random in the sense that the court is just as likely to overestimate the injurer’s precaution as to underestimate it. To illustrate the possibility of errors that are biased instead of random, suppose that the burden of proof upon plaintiffs is increased. As a result, courts often make the mistake of finding that injurers whose precautions fall below the legal standard are not liable, whereas courts seldom make the mistake of finding that injurers whose precautions exceed the legal standard are liable. The expected cost of violating the legal standard falls as a consequence, without any change in the expected cost of conforming to it, which in turn causes precaution to decrease.

Another example of biased error is provided by the use of radar machines to enforce speed limits. The machines presumably estimate the speed of a car with random error normally distributed about the true mean. Suppose police adopt the practice of not issuing tickets unless the machine estimates that a car was exceeding the speed limit by, say, at least 10 miles per hour. This practice creates incentives for motorists to exceed the speed limit by a modest amount.

Furthermore, if uncertainty about legal standards becomes extreme, rather than moderate, injurers’ incentives for precaution will be eroded. This is true even though error is unbiased. To illu-

46. In technical terms used in the appendix, see infra Technical Appendix, uncertainty is represented by a mean-preserving spread in the probability distribution of errors.

47. In technical terms, there are few type II errors (finding a violation where conformity occurred) and many type I errors (finding conformity where a violation occurred).
trate, suppose that enforcement is so riddled with random errors that people who violate the legal standard are no more likely to be found liable than those who conform to it. This state of affairs provides no incentive for injurers to bear the cost of precautions. Instead, potential injurers may withdraw from the activity that creates the possibility of liability. To illustrate, a 1986 survey by the Conference Board shows that forty-seven percent of United States manufacturers withdrew products from markets for liability reasons, thirty-nine percent decided against introducing some new products, and twenty-five percent discontinued some new product research. 48

Errors in applying the legal standard that are biased or extreme can cause the injurer’s cost-minimizing precaution to fall below the standard by a little or a lot. Such errors, consequently, undermine the usefulness of gross underprecaution as a criterion of intentionality. This criterion would have to be reconsidered for cases involving very large or biased errors in applying the legal standard.

2. Strict liability.—The simple model of continuous choice was based upon a negligence rule that conditions liability on conforming to a legal standard. Strict liability is an alternative rule, which conditions liability on causing harm, whether or not a legal standard was violated. To illustrate, the manufacturer of a defective product is liable for the harm it causes to consumers, regardless of the level of care taken in manufacturing it. Similarly, the person who engages in a dangerous activity, like dynamite blasting, is liable for all the harm it causes to others, regardless of his level of care. The model of punitive damages developed in this Article turns upon the jump in injurers’ costs at the legal standard. If liability is not conditional upon violating a legal standard, there is no such jump in costs. This fact raises a question about the applicability of the model developed in this Article to strict liability rules.

If there is no legal standard for determining liability for compensatory damages and no jump in costs, a purely self-interested decisionmaker will equate the cost of precaution and expected lia-

48. Priest, How to Control Liability Costs, FORTUNE, April 24, 1989, at 323. Priest also documents an increase in claims for product liability injuries that is occurring in spite of a decrease in the actual number of injuries. See also Huber, No-Fault Punishment, 40 ALA. L. REV. 1037, 1040-42 (1989).
bility at the margin, assuming that liability is the only cost that injurers bear from accidents.49 Purely self-interested decisionmakers consequently respond to enforcement errors in a rule of strict liability by gradually lowering their precaution, which may reduce precaution a small amount below the level of reasonable care.50 Consequently, the criterion of a gross shortfall from reasonable care is less useful for determining whether the violation is intentional. Sorting out intended from unintended violations of a standard of reasonable care seems an impossible task when small shortfalls may be intentional. Thus punitive damages should not be applied when the shortfall from the requirement of reasonable care is small, as is typically the case with manufacturing defects. In any event, replacing a negligence rule with a rule of strict liability makes recovery much easier for accident victims. The enforcement error should, consequently, be small.51 If the enforcement error is small, there is not much scope for punitive damages.

Consumer product injuries are the most important class of modern torts covered by strict liability rules. These cases can be distinguished according to whether the injury was caused by defective manufacture or defective design. Defective manufacture is a stochastic occurrence that quality control limits but cannot altogether eliminate. Manufacturers are held liable for the harm to consumers from manufacturing defects regardless of their level of quality control.52 In the case of defective designs, however, proving a defect often involves showing that the manufacturer violated accepted standards of safety in designing the product. Liability conditional on a shortfall from the legal standard is the defining characteristic of a negligence rule. Perhaps liability for design defects is best conceptualized as a negligence standard, even though the law persists for historical reasons in applying the label “strict

50. In economic jargon, the response to errors in a rule of strict liability is on the intensive margin, whereas with a negligence rule it is on the extensive margin.
51. This would not be true if less than full liability were combined with strict liability, as was allegedly the political bargain in the enactment of workers compensation. Friedman & Ladasky, Social Change and The Law of Industrial Accidents, 67 Colum. L. Rev. 50 (1967).
52. An attempt to recover punitive damages in such cases would presumably have to show not just that the product was defective, but that there was gross negligence in quality control or something similar. Thus the recovery of punitive damages requires a showing of a shortfall relative to a legal standard, even though the recovery of compensatory damages does not.
liability." If so, the limitation of this Article's model discussed in this section does not apply to design defects. Alternatively, the design standard applied by some courts may be so high as to be unattainable in practice, in which case liability is effectively strict, and the limitations discussed in this section apply.

3. Nonpecuniary harms.—Full compensation of accident victims for pecuniary harms makes them as well off as if they had not been injured. Potential victims should, consequently, be indifferent regarding escaping pecuniary harm or being fully compensated for suffering it. In the case of nonpecuniary harms, however, the concept of indifference becomes attenuated or inapplicable. To illustrate, full compensation for the wrongful death of a child obviously does not leave parents indifferent concerning whether or not the accident occurred. Nonpecuniary harms characteristically involve losses of things that are not bought and sold—limbs, lives, affection, pain—so, pricing them is difficult or impossible.

The problem that nonpecuniary harms raise for the theory developed in this Article is computing enforcement error. To compute the extent of enforcement error, if there is any, dollar values must be assigned to nonpecuniary losses. One solution to this problem is simply to consider the compensatory damages that courts actually give and to use these numbers as indicators of nonpecuniary value. This approach, while essentially sound, is better if some guide can be used when evaluating the range of numbers supplied by courts. There is, fortunately, an approach that solves this problem adequately from a deterrence perspective.

The approach involves a form of comparative risk analysis. Although lives are not bought and sold, people make investment decisions every day affecting the probability that people will die. Examples are investment in maintaining railroad track, servicing airplanes, quality control in manufacturing, and the testing of drugs. These investment decisions place implicit value upon lives. To illustrate, suppose that additional expenditures costing $200 reduce the probability of fatal accidents by .0001. Dividing the first number by the second yields $2,000,000. Suppose this level of precaution is deemed reasonable by the court. Assuming no enforcement error, an incentive for purely self-interested decisionmakers to conform to the rule will be created by awarding compensatory damages of $2,000,000 for each death. If enforce-
ment error were ½, the total damages for purposes of deterrence should be $4,000,000.

4. Malice and idiosyncratic compliance costs.—The model of punitive damages developed in this Article applies to injurers motivated by their own economic advantage. Its extension to people who act from malice or similar motives raises special problems. Actual malice refers to enmity which makes a person eager to harm someone else, even though there is no gain or advantage enjoyed by the injurer. Malicious injurers, like everyone else, will experience a jump in costs at the legal standard, so the gross shortfall criterion is still useful for determining whether or not to punish. Malice gives the injurer an additional reason for violating a legal standard apart from enforcement error. The additional element of malice is not captured by the rule of the reciprocal for awarding punitive damages, which is based upon enforcement error alone. Malicious motives, consequently, make the rule of the reciprocal insufficient as a guide to punitive damages.

To illustrate by extending the earlier example, suppose the injurer must spend 110 to avoid exposing someone else to the risk of losing 600 with probability ½. A reasonable decisionmaker will conform to the legal standard. Assume there is enforcement error equal to ½, so the injurer’s expected liability equals ½ x 600 x ½ = 100. A self-interested decisionmaker would respond by not taking precaution. To correct for this fact, compensatory damages of 600 are supplemented by punitive damages of 600, so the injurer’s expected liability now equals ½ x (600 + 600) x ½ = 200. A malicious person, however, might be willing to bear expected costs of 200 in order to expose her enemy to the risk of losing 600. Thus the rule of the reciprocal does not set punitive damages at a level sufficient to deter a malicious injurer.

Given these facts, punitive damages determined by the rule of the reciprocal may not be enough to deter behavior motivated by actual malice. An additional development of theory beyond this Article would be required to handle deterrence in such cases.

Malice can be described as increasing the psychological cost of complying with the legal standard. A variety of idiosyncratic motives can have a similar effect—envy, jealousy, revenge, etc. These motives are not ordinarily attributed to business organizations. Indeed, the attribution of human emotions to an institution such as a corporation is a metaphorical form of speech. The assumption that
businesses aim to maximize their profits, which fits with the model of punitive damages proposed in this article, is more sound.

5. Lapses and akrasia.—Tortious behavior warranting punishment is sometimes explained by a motive other than material self-interest or actual malice. Most people are subject to occasional lapses in which they do something wrong that they would not ordinarily do. The lapse may result from a temporary weakness of will. Weakness of will is a significant topic for moral philosophers, who call it "akrasia." Economists have recently begun to study lapses. One way to conceptualize a lapse is to divide the person in two, consisting of better and worse selves. The better self tries to find ways of constraining the worse self to prevent lapses in times of weakness. This approach in effect treats the better self as the principal who tries to control the worse self who is his agent. Strategies of controlling oneself are thus viewed as similar to strategies for controlling others.

Organizational lapses figure prominently in disputes over punitive damages involving business defendants. An organizational lapse occurs when a person makes decisions on behalf of the organization that are against its interests. To illustrate, an agent of the corporation runs risks without the knowledge of his supervisors that undermine its profitability. Organizational lapses will be considered later under the heading of vicarious liability.

54. This claim is advanced convincingly in Schwartz, Contributory and Comparative Negligence: A Reappraisal," 87 YALE L.J. 697, 713-19 (1978).
55. Tom Schelling’s pioneering research on addictive behavior, which is in progress, deals directly with weakness of the will. See T. SCHELLING, ETHICAL AND LEGAL IMPLICATIONS OF SELF-MANAGEMENT, in CHOICE AND CONSEQUENCE (1984).
6. ** Victim's precaution.**—The victim, as well as the injurer, is often able to take precautions to reduce the probability of an accident. It is easy to see why victim's incentives for precaution might be problematic. To illustrate by an extreme case, suppose the harm suffered by the plaintiff was purely pecuniary. In that case, supplementing compensatory damages that are fully compensatory with punitive damages converts a reluctant victim into an eager one. The victim who stands to gain from an accident obviously does not have incentives to prevent it.

This Article notes the problem of victim's incentives without considering its solution, which would require a significant extension of the model.

II. **Consequences for Court Practices and Examples**

Adopting the gross shortfall standard and the rule of the reciprocal would clarify usual court practices and modify occasional practices in several respects that are discussed below.

A. **Court Practices**

1. **Plaintiff's burden of proof.**—Adopting these two rules would change the burden of proof upon the plaintiff and the instructions to the jury concerning it. The plaintiff attempting to recover punitive damages from a business defendant would have to prove a gross shortfall from the legal standard and develop a theory about the motives for such behavior. The theory might be developed that the injurer acted from actual malice, although this is implausible for the typical business defendant. Alternatively, the theory might be developed that the injurer acted from material self-interest, which is very plausible for business defendants.

If the second theory is adopted, plaintiff must show that violating the legal standard would be advantageous for defendant in the absence of punitive damages. This step of the proof shows that an incentive problem exists with compensatory damages alone.

56. Economists refer to such a possibility as "bilateral precaution." For a discussion of how tort law has failed to provide victims with incentives for precaution, see Priest, *Punitive Damages and Enterprise Liability*, 56 S. CAL. L. REV. 123 (1982).

57. There are mechanisms by which this incentive problem can be overcome. The obvious mechanism is to allow a defense of contributory negligence.
From the deterrence perspective, it is unnecessary to prove that this defendant in this case believed he would escape punitive damages and this belief caused him to violate the legal standard grossly. Rather, it is necessary to prove that a similarly situated defendant would have the incentive to act in this way but for punitive damages.\textsuperscript{58}

The core of this demonstration is an account of the nature and extent of enforcement error. For example, the plaintiff might argue that most victims in similar cases never recover the compensatory damages to which they are entitled by law. Alternatively, the plaintiff might argue that compensatory damages recoverable by victorious plaintiffs in such cases are incomplete relative to the accident’s social cost.

Next, the plaintiff would have to show that enforcement error is so large that the cost of conforming to the legal standard exceeds the injurer’s expected liability for compensatory damages. This proof, if successful, would complete the demonstration that violating the legal standard is profitable for defendant in the absence of punitive damages.

Finally, the plaintiff would have to provide an estimate of the enforcement error to be used in setting the punitive multiple. The parties might employ experts on costs and deterrence to provide the proof or refutation of this estimate. The burden of proof and the method of calculating the punitive multiple would be matters of law to be explained to the jury.

Notice that plaintiff’s burden of proof as sketched above does not require an investigation into the injurer’s state of mind or intentions. It is the behavior of business defendants and the incentives faced by them, specifically the fact and profitability of a gross shortfall in precaution relative to the legal standard, not their mentality, which matters to deterrence theory.

2. Defendant’s wealth.—Courts routinely consider the total assets or wealth of the defendant when determining the level of punitive damages.\textsuperscript{59} This may be appropriate for retributive pun-

\textsuperscript{58} I am grateful to Krish Ladha whose persistent questions made me clarify this point.

\textsuperscript{59} Thirty-four recent cases are discussed briefly in L. Schluter & K. Redden, 1987 Cumulative Supplement to Punitive Damages 62-67.
ishment of malicious acts, but it is typically inappropriate to deterrence of economically self-interested decisionmakers. The controlling factor in a purely self-interested calculus of whether to conform to a legal standard is the cost of compliance relative to the cost of liability, assuming liability is the only cost of noncompliance. Whether or not a firm expects to enjoy an advantage from violating a legal standard typically has little or nothing to do with its total wealth or the value of its assets. To think otherwise is to commit the cardinal sin in economics—confusing marginal and total values.

If the gross shortfall standard were applied, considering total assets would be the exception rather than the rule. There are two special circumstances in which the defendant’s assets are relevant to setting punitive damages. If the judgment threatens the firm with bankruptcy, its assets are immediately relevant to the ability of present and future plaintiffs to collect judgments. This is not, however, a special problem of punitive damages. The problem of maximizing the damages actually collected from a firm approaching bankruptcy is independent of whether the damages are purely compensatory or involve an element of punishment.

The second circumstance is one in which the stakes in the case amount to a significant proportion of the defendant’s wealth. It is well known that most decisionmakers are averse to risks of this magnitude. Given risk aversion, full deterrence can be obtained by setting the punitive multiple at a value somewhat less than that given by the rule of the reciprocal.

3. Ratio rule.—Court practice acknowledges the “ratio rule,” according to which punitive damages must bear a reasonable ratio

60. European criminal justice systems often adjust fines according to wealth or income. For example, criminals are often assessed fines denominated in work days (e.g., a five-day fine equals the convicted person’s ordinary earnings over five days of employment). See Comment, The West German Day-Fine System: A Possibility for the United States?, 60 U. Cin. L. Rev. 281 (1993).


62. A discussion of the need to avoid driving firms to bankruptcy is in Sales & Cole, supra note 5, at 1141-45.

63. The reciprocal rule is based upon expected values, whereas a risk-averse decisionmaker will discount the expected values by a factor indicating his aversion to risk. The effect of risk aversion upon the common law has been stressed by Polinsky. See, e.g., Polinsky, Risk Sharing Through Breach of Contract Remedies, 12 J. LEGAL Stud. 427 (1983).
to compensatory damages. Absent any instructions for determining a reasonable ratio, this "rule" is little more than a rationale for judges to set aside punitive damages that strike them as excessive. 44 Recent cases exhibit wide variation in ratio; as mentioned above, a 60 to 1 ratio was deemed reasonable by the judge 46 as was a 300,000 to 1 ratio. 44 At the lower end, punitive damages can be 10 percent or less of compensatory damages. The range of these examples, while not typical, 47 suggests the extent to which juries are unconstrained by rules when setting punitive damages. It is almost inconceivable that enforcement error could be of the magnitude of 60 to 1 because such large errors would provoke a corrective response. Rather, enforcement error is more likely to be of the magnitude of 2/1 or smaller, possibly rising to as high as 10/1 in cases of fraud, but not 60/1; this conclusion is supported by evidence from criminal law in the next section of this Article.

The major forms of enforcement error are failure to assert claims of victims legally entitled to recover and undercompensation of successful plaintiffs. There is disagreement about whether the rapid expansion of compensatory damages in recent years has eliminated undercompensation of successful plaintiffs. It is difficult to imagine circumstances in which undercompensation for economic torts would be of the order of 10 to 1, or even 2 to 1. 48 Corrective measures would normally be taken before such a level of

64. The vagueness of the ratio rule is discussed in Sales & Cole, supra note 5, at 1145-47. The view that punitive damages might be used to overcome enforcement error is stated by Ellis in an important article. He describes two grounds where punitive damages might improve efficiency. The first is "where expected liability for compensatory damages is less than expected harm to society." Ellis, Fairness and Efficiency, supra note 1, at 77. This satisfies my definition of enforcement error. The second ground is "where harm is deliberately caused and the satisfaction obtained by the actor is 'illicit.'" Id. This corresponds to malice in fact, which does not presuppose enforcement error. He mentions a third ground—where the ex ante probability of being held liable for the loss is less than the actual probability of the harm, id. at 25-26—which is another form of enforcement error. Ellis does not, however, develop the principle that the punitive multiple should equal the reciprocal of the enforcement error. Also see discussion in Note, Formulating Standards, supra note 5, at 2053.

65. Auburn Harpwell Ass'n v. Day, 438 A.2d 234 (Me. 1981); see supra note 3 and accompanying text.


68. Compensatory damages may be far too small relative to the actual harm in cases of libel or insult, although the harm is so intangible that there is scope for wide disagreement.
undercompensation developed, especially in a system where juries determine awards. In any case, if the rule of the reciprocal were used by courts, juries would have to base the ratio upon factual evidence, rather than moral intuitions.

The failure of some victims to assert legal claims, as opposed to the undercompensation of successful plaintiffs, may result in a larger enforcement error, especially where fraud and deceit are involved. There are, however, criminal sanctions to correct this problem, which introduces the next topic.

4. Criminal fines and statutory penalties as guidelines.—There is a close relationship between the rule of the reciprocal proposed in this Article and the magnitude of criminal fines. Most white collar crimes are economic in the sense that the criminal’s motivation is material self-interest. To deter such crimes by fines, they must be large enough to overcome the expected gain from committing the crime. If enforcement were perfect in the sense that criminals were always apprehended when the stolen money could be recovered, requiring them to return the money would be a sufficient deterrent. A fine in addition to return of the stolen money is needed for deterrence because some crimes go unsolved and others are solved after the thief has spent the funds. The size of the fine needed for deterrence thus depends upon the size of the enforcement error.

Deterrence is presumably an important consideration in setting statutory civil or criminal penalties, such as fines for white collar crimes or treble damages for antitrust violations. What matters for deterring injurers is the extent of their expected liability, not whether the money collected from them goes to the state as a fine or to the victim as damages. The two rules derived in this Article—gross shortfall criterion and the rule of the reciprocal—are required for efficient deterrence, whether they are applied by courts in civil disputes or criminal trials. As a consequence, the magnitude of statutory fines provides evidence concerning the

---

In any event, these are not forms of economic tort which are the central concern of this Article.

judgment of authorities about the requirements for deterrence and the extent of enforcement error. Fines for economic crimes that are also torts thus provide a guide to setting punitive damages consistent with the rule of the reciprocal. It might be appropriate, for example, for a court to impose special burdens on a plaintiff seeking to recover punitive damages that exceed the statutory level for comparable acts. Thus a plaintiff seeking punitive damages in excess of treble damages might have a special burden of proof.

5. Vicarious liability.—The master is liable at common law for the torts of his servant committed in the course of his employment. If the employee’s act warrants punishment, should the employer be held liable for punitive damages as well as compensatory damages? Jurisdictions disagree on this point, some holding that liability for compensatory damages implies liability for punitive damages, and others holding that the employee’s act must bear the stamp of his employer’s approval.

The theory developed in this Article provides the best answer for purposes of deterrence. There is an analogy between defective products and the tortious acts of employees. The employer is strictly liable for the torts of employees acting in the course of their employment (although the employee may have to be negligent in order to commit a tort). This legal principle can be rationalized on the same grounds of deterrence as strict product liability—it internalizes the risk of loss by placing liability on the party best situated to avoid it. The burden of proof upon the injured plaintiff is light, so enforcement errors are few.

70. I am grateful to John Jeffries for this suggestion.
72. What is probably a majority of the courts which have considered the question hold an employer, especially a corporate employer, who is liable for actual damages also liable for punitive damages for malicious acts of his employees. Other courts hold that the employer is not liable unless he or his managerial officers participate in, order or ratify the employee’s actions (the “complicity rule”).

K. RADDOCK, supra note 2, § 4.5, at 116 (footnotes omitted).

As with consumer product injuries, however, obtaining punitive damages requires demonstrating something more than injury and causation. In addition, the plaintiff must show a shortfall from reasonable care that was gross or intentional. A business organization, rather than being a single mind, binds together different people with their own interests. Attributing states of mind to organizations, such as the intention to violate a legal standard, is unproblematic when everyone’s interests align, but attribution is problematic when their interests diverge. Thus when the tortious act of a employee gives rise to a suit against the company, the case is relatively straightforward if the employee’s act was contemplated and approved as corporate policy. More problematic, however, are suits resulting from a wayward employee whose acts are out of line with company policy. The wayward employee may intentionally violate a legal standard, but insofar as she acts against company policy, the company did not intend for her to do it.

The directors and managers of the company are the principals who define policies that must be carried out by their employees who are their agents. The incentive structure for channelling employee activity aligns their interests and the company’s imperfectly, so employees sometimes depart from policy. For example, an employee may find that she enjoys the “upside” risk of a dangerous activity, which includes a promotion and a raise, whereas much of the “downside” risk falls upon the company, including several million dollars in liability. The departure from policy, once it becomes known by the principals, will provoke corrective action by them, possibly including an attempt to modify the incentives of agents. This situation is analogous to lapses and akrasia, which were discussed earlier in this Article.

The same model of punitive damages applies to monitoring the tortious acts of employees as to monitoring output for defective products. The employer should be liable for punitive damages when there is a gross shortfall from the legal standard, and the punitive multiple should be set by the rule of the reciprocal. The point to notice that is easily missed, however, concerns the act that must fall grossly short of the legal standard. It is the act of the

73. See K. Renosh, supra note 2, § 4.2, at 85.
74. See supra text preceding and accompanying note 53.
employer in monitoring the employee, not the act of the employee. To be specific, the employer's monitoring of the employee should fall far short of the standard of reasonable supervision to warrant the employer's liability for punitive damages.

The rationale for this requirement follows immediately from the central arguments of this Article. Most employers find that a reasonable level of monitoring of employees is in their own self-interest. The employer whose negligent monitoring allows an employee to commit a tort, even an intentional tort, usually does so by mistake. Liability for compensatory damages is enough to call the mistake to his attention and eliminate the problem in the future. The exceptional employer, however, finds that monitoring employees at a level below the legal standard would be advantageous if liability is limited to compensatory damages. For this employer, punitive damages are needed to restore the incentive to monitor employees at the level required by law.

6. Insurance.---Should potential defendants be allowed to insure against the possible award of punitive damages? There is currently a split among jurisdictions over this question in the United States, with a majority apparently allowing insurance coverage. The question has both a backward-looking and a forward-looking aspect. The backward-looking question, not discussed here, is whether to extend coverage to punitive damages for contracts that do not explicitly mention punitive damages and apparently did not contemplate them. The forward-looking question, which

75. An argument along much these same lines that is based upon monitoring can be developed to explain why courts apportion punitive damages among multiple tortfeasors, rather than following a rule of joint and several liability. A discussion of court practice and its rationale is in Note, The Apportionment of Punitive Damages Among Joint Tortfeasors, 25 Ark. L. Rev. 579 (1963).

76. At present, the majority of the states in and territories of the United States permit the insurability of punitive damages directly or vicariously assessed against an insured. Less than half of the states in the United States prohibit the insurability of punitive damages that are directly assessed against an insured, and only three states prohibit the insurability of punitive damages that are assessed vicariously as a result of acts of another for whom an insured is responsible.

PUNITIVE DAMAGES: A GUIDE, supra note 2, at 35.

This contradicts Sales & Cole: "A majority of jurisdictions conclude that public policy forecloses the application of insurance coverage for punitive awards." Sales & Cole, supra note 5, at 1153. The former conclusion, however, is apparently based upon more systematic data.

77. According to George Priest, the relevant clause in traditional insurance contracts disclaims coverage for "acts neither expected nor intended by insurers." Interpreting this
is discussed in this section, is whether to allow such coverage when the contract indicates the coverage was contemplated by the parties.

If the victims of accidents for which the losses are purely pecuniary receive punitive damages as well as full compensation, there will be many eager victims. It is hard to see any public policy reason to forbid insurance in such cases. Insofar as the losses involve death, dismemberment, disfigurement, or permanent injury, there will be few or no eager victims, regardless of the level of compensation. In these circumstances, potential injurers should take precaution until the relationship between injuries and the cost of avoiding injuries is the same in this industry as in other industries where precaution is generally regarded as reasonable. The objection to insurance of punitive damages in such cases is that it allegedly reduces incentives for policy holders to take precaution, thus resulting in unreasonable behavior and excessive accidents.

Insurance involves both a moral hazard, which is the temptation of injurers to risk liability because they know the insurance company will bear all or part of it for them, and adverse selection, which is the tendency of high-risk policy holders to drive low risk policy holders out of the insurance market. The claim that tort insurance of all kinds, including coverage of compensatory damages as well as punitive damages, reduces injurer's incentives for precaution by shifting part of the cost of accidents is correct in a purely short run analysis. A long-run analysis, however, identifies forces that promote precaution which may be stronger than the forces undermining it.

phrase to exclude coverage for punitive damages becomes more difficult as they become more common. Certainly the possibility of punitive damages is now contemplated by businesses and their insurers. See Priest, Insurability and Punitive Damages, 40 ALA. L. REV. 1009 (1989).

78. A cost-minimizing injurer weighs the cost of precaution against the burden of liability. If these values are balanced at the margin, a reduction in the burden of liability will cause the injurer to take less precaution. A risk-averse injurer adds a premium to the expected costs of liability when computing its burden. Insurance companies reduce risk to their customers by spreading it. This insurance against liability reduces the risk premium attached to it by a risk-averse injurer, which in turn reduces the burden of liability. The lower burden tilts the balance in favor of less precaution. Furthermore, the lower burden of liability makes the industry more profitable, which causes existing firms to expand or new firms to enter the industry. A higher activity level usually results in more injuries, assuming the level of precaution does not increase.
Courts and insurance companies can be thought of as monitors to prevent potential injurers from harming others. As monitors, they differ with respect to extent and type of information they obtain, the cost of obtaining it, and the time they obtain it. Insurance companies in a competitive market try to hold down their costs, which involves avoiding claims. To avoid claims, insurance companies may require applicants to prove a history of careful conduct before writing a policy, especially if the policy extends to punitive damages. Insurance companies may require policy holders to supply information proving conformity to practices specified in the policies before allowing claims. If a company passes these tests and its precaution nevertheless falls short of the legal standard, its liability claims will be out of line with more careful companies, so the insurer will respond in the future by refusing to renew the policy or by raising rates. Insurance companies are similar to government safety regulators in that they obtain some information in advance and have discretion in its interpretation. In contrast, courts require detailed evidence to decide disputes, but the information is not supplied until after an accident occurs, and formal standards of law constrain the court's use of the information. Insurance thus substitutes ex ante regulation of policy holders by insurance companies for some of the ex post liability imposed by courts and borne by defendants.

If insurance companies can obtain timely information about policy holders more cheaply than plaintiffs can obtain information about defendants, then a mixed system that allows ex ante regulation by insurance companies may reduce injuries below the level obtained through a pure system of ex post adjudication by courts. To illustrate, if insurance is a practical necessity for a company to participate in a particular market, the fear that insurance may be terminated could be an effective device to guarantee a high level of care. More generally, assume that insurance companies only write policies for firms with good safety records, and assume that insurance lowers the cost of doing business. Under these assumptions, the public policy of allowing insurance conveys a competitive advantage upon firms with good safety records. The combination of
insurance and competitive pressure favors firms with good safety records, which could be called "beneficial selection."\textsuperscript{79}

There is reason to believe that the tendency towards beneficial selection will be especially strong when insurance extends to punitive damages. The sheer magnitude of the potential liability should result in relatively high levels of risk aversion by firms, which implies a large competitive advantage to having insurance.\textsuperscript{80} The same fact motivates a high level of monitoring by insurance companies. The combination of risk aversion by firms and careful monitoring by insurance companies will produce beneficial selection.

Whether the tendency towards beneficial selection is stronger or weaker than the tendency of insurance to undermine short run incentives for precaution is, however, an empirical question that is yet to be answered.\textsuperscript{81} To illustrate the possibilities, it could turn out that insurance does not significantly affect the level of precaution, and its main effect is to lower the cost of doing business for some firms. Until the needed research is done, the case in favor of allowing insurance of punitive liability must be advanced tentatively.

7. **Selling Unmature Tort Claims.**—Insurance transfers responsibility for the cost of accidents from one person to another before accidents occur. Disclaimers and waivers have the same effect. A general problem with the tort system is that the rights of recovery that it extends to tort victims are often worth less to

\textsuperscript{79} Beneficial selection assumes that insurers can distinguish between safe and unsafe firms and that they only write policies for the former. Adverse selection, in contrast, assumes that insurers cannot distinguish between safe and unsafe firms, so by writing policies indiscriminately, they convey a competitive advantage to the unsafe firms who will have more claims.

\textsuperscript{80} Firms are apparently risk averse as evidenced by widespread tort insurance. However, perfect diversification of portfolios by stockholders makes them risk neutral, which suggests that firms should be risk neutral as well. This apparent paradox has never been entirely resolved, so the causes of risk aversion by firms are imperfectly understood.

them than the corresponding costs borne by potential injurers.\textsuperscript{32} There is, consequently, room for a profitable exchange in which potential victims give up some of their tort rights in exchange for money from the potential injurer.

Transferring tort rights through mechanisms such as disclaimers and waivers constitutes a market in unmatured tort claims.\textsuperscript{33} These markets raise the question, “To what extent should the law permit potential victims to waive or sell their rights to recover punitive damages?” To answer this question, consider in turn two effects of such a market transaction. First, injurers are able to reduce their exposure to liability for punitive damages. This fact raises many of the same public policy concerns as insurance. The preceding section argued tentatively in favor of insurability of punitive damages on the ground that its main effect is to substitute a superior system of monitoring potential injurers for an inferior one, with the result that transaction costs go down and the rate of accidents does not go up. A similar argument can be made concerning a competitive market for unmatured tort claims. It can be argued that the problem with waivers and disclaimers in the nineteenth century was legal obstacles to competition in this market. The best solution to the nineteenth century problem was to facilitate competition; unfortunately, the misguided solution in the twentieth century was to close the market by judicial fiat.\textsuperscript{34} There


\textsuperscript{33} Cooter & Sugarman, \textit{A Regulated Market For Unmatured Tort Claims: Tort Reform by Contract}, \textit{in New Direction in Liability Law, supra note 29, at 174.}

\textsuperscript{34} Waivers and disclaimers create bilateral monopoly between injurer and victim. To have a competitive market in unmatured tort claims, the right to recover must be fully transferable by potential victims to anyone. A proposal for such a market is developed in Cooter, \textit{Towards A Market in Unmatured Tort Claims}, 75 VA L. Rev. 383 (1989), with comments by Schwartz, Commentary on \textit{“Towards a Market in Unmatured Tort Claims”: A Long-Way Yet to Go}, 75 VA L. Rev. 423 (1988), and Goetz, Commentary on \textit{“Towards a Market in Unmatured Tort Claims”: Collateral Implications}, 75 VA L. Rev. 413 (1988).

To illustrate, suppose that a union offers its members the option of waiving their rights to recover damages for medical malpractice from a health maintenance organization on an annual basis in exchange for a reduction in the annual cost of health care. For those union members who accept this option, the health maintenance organization in effect buys up their liability claims in advance, so no claim can be made in the event of an accident. Thus the health maintenance organization faces reduced liability, and the union has greater incentive to monitor the health maintenance organization's level of care. Union members would be free under this proposal to sell their right to recover damages for medical malpractice to someone other than the health maintenance organization (e.g., to a law firm).
is reason to believe that a competitive market in unmatured tort claims would effectively monitor injurers, but this argument is beyond the scope of this Article.

Second, victims lose some or all of their rights to recover damages. This may or may not cause concern with respect to compensatory damages, but in any case it should not cause concern with respect to punitive damages. There is apparently a consensus that the purpose of punitive damages is to deter and punish injurers, not to compensate victims. If this consensus is correct, it would be a mistake to think of a reduction in punitive recoveries as victims' losing something to which they are entitled.

8. Multiple plaintiff litigation.—Another issue concerns multiple plaintiff litigation. The general concern is whether punishment might become excessive by virtue of awards being collected by different plaintiffs in different cases. This problem, however, cannot arise if the enforcement error is computed correctly. To illustrate, if enforcement error arises because each victim's compensatory damages equals only $\frac{1}{2}$ the social cost of the accident, then the punitive multiple relevant for every plaintiff is $2/1$. Similarly, if enforcement error arises because only $\frac{3}{5}$ of accident victims successfully recover, then the punitive multiple relevant for everyone who sues and recovers is $3/2$. Notice that computing the punitive multiple of $3/2$ requires taking account of the proportion of future victims who will recover. So long as the bankruptcy constraint does not bind,\textsuperscript{66} multiple plaintiffs whose cases are the same with respect to enforcement error should enjoy the same punitive multiple.

9. Exposure To Risk.—A related issue concerns recovery for exposure to risk. The expected social cost of an accident equals its probability, denoted $p$, multiplied by the actual social harm, denoted $L$, which is denoted by $pL$. The usual court practice is to award damages of $L$ whenever an accident occurs. An alternative advocated by some scholars is to award damages of $pL$ whenever exposure to risk occurs for certain kinds of cases.\textsuperscript{66}

\textsuperscript{66} See supra note 62 and accompanying text.

\textsuperscript{66} Robinson, Probabilistic Causation and Compensation for Tortious Risk, 14 J. Legal Stud. 779 (1985):
These two approaches may result in enforcement error of different sizes, which provides a basis for preferring one approach to another. The size of the enforcement error determines the punitive multiple, according to the rule of the reciprocal. Although these two approaches to compensation may affect the magnitude of punitive damages obtained by using the rule of the reciprocal, they do not affect its applicability. Regardless of which practice is adopted, the punitive multiple should equal the reciprocal of the enforcement error for purposes of deterrence.

B. Examples

In this section of the Article the theory already developed will be applied to hypothetical facts obtained by simplifying actual cases. The purpose of this section is to illustrate in detail the application of the gross shortfall criterion and the rule of the reciprocal.

1. Unauthorized publication of photograph.—Facts: Defendant used plaintiff’s photograph on a commercial product without the plaintiff’s written permission.

Decision: The jury awarded plaintiff $1.00 in compensatory damages at trial and $35,000 in punitive damages, reduced to $15,000 by appellate court.

Analysis: The likelihood that the publication of the photograph would go unnoticed by its subject or that the subject would not assert a claim is apparently small. A court, however, would have trouble fixing compensatory damages because the harm is in-

Assuming that the risk is one that would give rise to liability when the actual loss is suffered, why not adjudicate the entire case by awarding the victim the present value of the risk at the point at which the risk can be identified and given some measurable value? The value is equal to the present value of the future losses multiplied by the estimated probability of their occurrence.

Professor Farber has recommended a different approach incorporating the concept of damages for risk, awarding full compensation to plaintiffs with the highest probabilities of becoming actual victims. Farber, Toxic causation, 71 Minn. L. Rev. 1219 (1987). Under this scheme, if the funds available for payment of claims are limited, then necessarily some plaintiffs with relatively low probabilities of injury will not be compensated.

87. The cases suggesting the stylized facts used in these examples are largely found in S. SCHLUTER & K. REDDEN, supra note 59, § 28.

tangible. Thus limiting recovery to compensatory damages would create an enforcement error to the extent that the social harm exceeds compensatory damages.89

The direct application of the rule of the reciprocal to determine punitive damages requires estimating the social harm, which is difficult in this case. A more practical approach is to use the rule of thumb according to which the enforcement error for the marginal decision maker equals the ratio of costs of precaution to the expected harm resulting from its absence.90 The cost of precaution in this case refers to the cost of buying the right to publish the photograph from its owner.91 The punitive multiple should thus equal the cost of buying the right to publish the photograph from its owner divided by compensatory damages (one dollar). Thus the question of whether the actual award is excessive depends upon whether the amount ordinarily paid for rights of publication of such photographs is more or less than $35,000.

2. Consumer product liability.92—Facts: A defect in a product manufactured by defendant caused plaintiff who used the product to lose a thumb and two fingers.

Decision: A punitive damages award of $440,000 was given by the court in addition to compensatory damages.

Analysis: To obtain punitive damages, plaintiff should first prove that defendant's behavior fell grossly short of a standard of care, either in manufacture or design, depending upon the nature of the alleged defect. Further, the plaintiff must show that the gross shortfall would be profitable to the defendant in the absence of punitive damages. For example, plaintiff may try to show that the defect was caused by quality control that fell grossly short of industry standards. The plaintiff may also attempt to show that lax quality control would be profitable to defendant corporation so long as its liability was limited to compensatory damages. The

89. The enforcement error arising from compensatory damages that are too low would be reduced to the extent that adverse publicity arising from the case harmed the defendant's business.

90. Retaining the standard notation in this Article, the marginal injurer satisfied \( b = pLq \), which implies \( q = b/pL \). In this example, the burden \( b \) is the cost of purchasing the publication right; the probability of apprehension for publishing it without permission is approximately 1. Thus the equation can be written \( q = b/L \). For a more detailed explanation, see supra note 10 and accompanying and following text.

91. I am assuming away reputation effects as indicated supra note 89.

profitability of laxity presupposes the existence of a large enforcement error. Enforcement error in a case like this is unusual. Accident victims are almost certain to bring successful suit and, furthermore, there are apparently no legal obstacles to awarding damages that are fully compensatory. So the award of punitive damages in this case must be viewed with skepticism.


Decision: The court awarded punitive damages of $12,000, even though defendant’s net worth was $3,116.

Analysis: This case is not amenable to the analysis developed in this Article because the injurer’s motivation was not material self-interest. Punitive damages for deterrence cannot be based upon the rule of the reciprocal in this case because enforcement error is at most only part of the motivation for the wrongdoing.

4. Erroneous Accusation.**—Facts: A customer of a retail store was erroneously charged with shoplifting, his vehicle was searched by the security force, and goods that the plaintiff owned were taken from his car.

Decision: Plaintiff was awarded $750,000 in punitive damages, and the award was found not to be excessive when awarded against a multimillion dollar corporation.

Analysis: To receive punitive damages, plaintiff must show that defendant’s behavior fell grossly short of legal standards. Furthermore, plaintiff must show that doing so was in defendant’s material interest due to enforcement error. Presumably, the defendant gained a material advantage from deterring shoplifters with excessive vigor. Plaintiff needs to show that this advantage more than offsets the risk of liability for compensatory damages alone. Several different theories could be developed by plaintiff concerning enforcement errors that create this alleged advantage. The obvious approach is to argue that the social harm is too intangible to be fully captured in compensatory damages.

A direct application of this theory to the rule of the reciprocal would involve estimating the social value of the harm. The indirect approach applied above to the first hypothetical case, which uses a

rule of thumb, might be more practical. In any case, it is inconceivable that the harm could be of the magnitude of $750,000. The total wealth of the corporation has no direct bearing upon the magnitude of the social harm or the level of punishment required for deterrence.

5. Seller misrepresentation.** Facts: Car dealership misrepresented car as new to buyer when, in fact, it had repaired body damage. Decision: Jury found actual damages of $1,250 and awarded exemplary damages of $15,000, reduced by the appellate court to $5,000. Analysis: This act of deception apparently had a high probability of success since the victim was unlikely to discover the harm, or discovery might come too late to provide proof. If the jury reasoned that the probability of discovery was of the magnitude of 1/15, a punitive multiple of 15/1 would be appropriate. If the judge reasoned that the probability of discovery was more like 1/5, the reduced punitive multiple of 5/1 was appropriate. The issue in setting damages is, in any case, the probability that the deception would succeed.

6. Defective product.** Facts: Television sets were manufactured with a design defect that could cause overheating and fire. The manufacturer knew that the design fell below the industry standard of safety, yet did not make changes and suppressed information about the lack of safety. The defect caused a fire that injured the owner of the television set when she tried to extinguish it. Decision: The jury awarded plaintiff compensatory damages of $125,000, punitive damages of $100,000, and attorneys’ fees of $50,000. Analysis: To recover punitive damages, plaintiff must prove that the design defect was a gross shortfall from the legal standard. The manufacturer apparently expected to profit by saving the cost of correcting the defect and escaping liability in those cases where victims did not know or could not prove that the defect caused the fire. The punitive multiple should be determined by the size of the

enforcement error, which apparently equals the ratio of successful consumer suits concerning the defect to fires caused by it. The punitive multiple might be inferred indirectly from the ratio of the manufacturer's costs saved by not correcting the error to the social harm caused by the resulting accidents.

7. Attorney's fees.—Facts: Plaintiffs who bought house built by defendants brought an action for breach of the implied covenant of habitability.

Decision: Plaintiffs who were awarded punitive damages asked for and received expenses in prosecuting the action, including attorneys' fees and costs, as part of exemplary damages.

Analysis: The magnitude of punitive damages should be determined by the rule of the reciprocal. Attorneys' fees and similar costs will be reimbursed insofar as they are an aspect of enforcement error, as for example when the defendant grossly violated a legal standard because high attorneys' fees would prevent most victims from proceeding with an action. Attorneys' fees do not form any separate basis for a claim apart from their contribution to enforcement error.

CONCLUSION

The questions, "When should punitive damages be awarded?" and "How much should be awarded?" have been answered by constructing a behavioral model that can be summarized in a paragraph. Potential injurers usually find that liability for compensatory damages makes conforming to the legal standard cheaper than violating it. However, in those circumstances where the usual cost relationship is reversed and the decision is made to violate the standard, a large change in behavior is needed to offset the jump in liability costs that goes with falling below the legal standard. That is why intentional violation of a legal standard to save costs tends to be gross. The behavioral criterion for awarding punitive damages in suits against businesses should usually be a gross shortfall from the legal standard. Violation of the legal standard is profitable in the absence of punitive damages because enforcement errors reduce the injurer's expected liability below the social cost of the

harm caused by violating the legal standard. To align expected liability and social costs, and to make conforming to the legal standard the least-cost alternative, the punitive multiple should be set equal to the reciprocal of the enforcement error.

This model is consistent with many aspects of current court practice. For example, the deterrence rationale figures prominently in the law of punitive damages. The "gross shortfall" standard is similar to the language used by courts in enunciating the conditions of punitive liability. The gross shortfall standard suggests that punitive damages should rarely be awarded, and empirical evidence indicates that punitive damages are not being awarded in many torts cases. The punitive multiples in typical court cases are apparently within the range suggested by enforcement error and the rule of the reciprocal. Punitive multiples that are very small (say under .10) or very large (say over 4/1) are unusual.

The lack of a generally accepted theory of punitive damages, however, results in inconsistent and unpredictable court decisions. The proposals in this Article would focus thinking about punitive damages upon more definite standards and produce greater predictability. The plaintiff's burden of proof would be clarified and raised, and the jury would be given a more definite guide, presumably with the help of expert witnesses. Defendants' wealth would not affect the level of punitive damages except insofar as risk aversion or bankruptcy were at issue. The ratio rule would be more definite and constrained at the upper end. Criminal fines would be regarded as good guidelines for setting punitive damages for purposes of deterrence. Vicarious liability would require showing a gross shortfall by employers from the legal standard for supervising employees, not just a gross shortfall from the legal standard of care by employees. Insurance would be allowed typically.

Aside from these specific clarifications and modifications of court practice, analyzing punitive damages with economic models of deterrence reorients the inquiry in favor of behavioral theories.

98. Landes and Posner report that of 172 products liability cases in their sample decided in federal appeals courts between 1982 and 1984, only four punitive damages awards survived appeal. W. LANDES & R. POSNER, supra note 7, at 302-04. They also examined 119 cases decided in state courts from 1984 to 1985 and found punitive damages awarded upheld in less than 2 percent of the cases. Id. at 305-06. These results, as noted by Landes and Posner, are consistent with findings by the Rand Corporation reported in PUNITIVE DAMAGES: PRELIMINARY EMPirical FINDINGS, 16-17 (N-2342-ICJ Aug. 1985), and Daniels, PUNITIVE DAMAGES: THE Real STORY, A.B.A. J., Aug. 1, 1986, at 60.
and against subjective moral feelings. Deterrence, being a strictly forward-looking perspective, is more concerned with incentives than motives. Punitive damages are justified from a deterrence point of view because the incentives to conform to the law are inadequate without them. Little wonder that where materially self-interested decisionmakers, like businesses, are concerned, replacing moral intuitions with a behavioral theory increases the clarity of legal standards and provides a firmer foundation for public policy.
Deterrence

Technical Appendix

The technical appendix proves some propositions invoked in the Article about the effects of uncertain legal standards upon the injurer’s precaution. Injurer’s precaution \( x \), whose unit price is \( b \), results in an accident with probability \( p(x) \), where \( p^< 0, p^> 0 \). In the event of an accident, the victim suffers harm costing \( L \). Social efficiency requires injurer to minimize the sum of the costs of precaution and expected harm:

\[
\min \ bx + p(x)L, \quad x
\]

with first order condition

\[ 0 \leq b + p^L. \]

(1)

Injurer’s precaution is observed by the court with error \( e \) whose probability density function is \( r(e) \). Under a negligence rule with legal standard \( x^* \), injurer is found not liable if \( x+e \geq x^* \), and injurer is found liable if \( x+e < x^* \). Thus the expected total costs \( ETC \) of the injurer are given by

\[
ETC(x) = bx + p(x)LR(x^*-x)
\]

where \( R \) is the cumulative density function. The injurer’s expected marginal costs are

\[
EMC(x) = b + p^LR(x^*-x) - pLr(x^*-x).
\]

(2)

Assume that the distribution \( r \) is centered around the true value, which is \( e = 0 \). An increase in court’s uncertainty is represented by a mean-preserving spread of \( r \). One way to spread the distribution while preserving the mean is to decrease the value of \( r(0) \), which represents a fall in the probability that the court’s judgment will be correct. This is the class of mean-preserving spreads that I will consider.

I seek the conditions under which uncertainty causes injurer’s precaution \( x \) to exceed or fall short of the legal standard \( x^* \), which corresponds to a situation in which \( EMC(x^*) \) is negative or positive, respectively. Consider the value of \( EMC \) when injurer’s precaution equals the legal standard \( x^* \). Under this assumption, (3) becomes

\[
EMC(x^*) = b + p^LR(0) - pLr(0).
\]

(4)
When the court has a lot of information, the distribution \( r \) is concentrated around its true mean, rather than being spread. Under these conditions, \( r(0) \) is large. If \( r(0) \) is large enough, then (4) is negative. Specifically, (4) is negative if

\[
    r(0) > b + p\'LR(0)/pL. \tag{4'}
\]

When the court has enough information to satisfy \( 4' \), the injurer’s expected marginal cost is decreasing at \( x = x^* \), so injurer’s optimal value of \( x \).

Consistent with the preceding paragraph, assume that \( EMC \) is negative at \( x = x^* \) when the court has good information. Now consider the consequences of reducing the information available to the court. An increase in uncertainty, as defined above using the concept of a mean-preserving spread, will leave the value of \( R(0) \) unchanged, whereas \( r(0) \) will decrease. The result is an increase in (4). If this process continues, the value of (4) eventually turns positive. Specifically, \( ETC(x^*) \) is positive from (4) if

\[
    b + p\'LR > pLr. \tag{4''}
\]

When the court’s uncertainty is large enough to satisfy \( 4'' \), the injurer’s expected marginal cost is increasing at \( x = x^* \), so injurer’s optimal value of \( x \) must lie below the legal standard \( x^* \).