PUNITIVE DAMAGES, SOCIAL NORMS, AND ECONOMIC ANALYSIS

ROBERT D. COOTER*

INTRODUCTION

From its beginnings, the economic analysis of law has traced the incentive effects of liability on potential wrongdoers who pursue their material advantage.1 Recently the economic analysis of law has turned to the study of social norms.2 This paper relates these two literatures to punitive damages. People...
who internalize social norms feel righteous anger against those who violate them. The institution of punitive damages allows judges and juries to express righteous anger through speech and punishment. Expression of emotions by the court demonstrates the strength of its commitment to the law in question. Perception of this commitment shapes the expectations of citizens and changes their behavior. In technical terms, expressions of emotion signal commitment, and commitment provides a focal point in a game with multiple equilibria.

Conventional morality provides a basis for ordering wrongdoing by its seriousness, but conventional morality does not attach precise punishments to wrongs. Consequently, people order wrongdoing consistently by its seriousness, but they attach inconsistent money damages to wrongdoing. Social norms provide a better guide to the need for punitive damages than to their extent. Inconsistent awards of punitive damages muddle the message conveyed by the courts concerning the seriousness of the wrong. Courts should not ask juries to determine the extent of punishment without providing instructions for its computation from the facts of the case.

Consistency requires mapping wrongs ordered by seriousness into punishments ordered by severity. The goal of deterrence provides a basis for consistent punishment. The court in a typical case should impose the minimum punishment required to deter wrongdoing. In jury trials, the court should instruct the jury on how to compute the minimum punishment required for deterrence. To compute the minimum, I distinguish between intentional and unintentional wrongdoing by a theory of rational self-monitoring. Imperfect self-monitoring results in lapses that cause unintentional violations of legal obligations. I show that liability for compensatory damages provides efficient incentives for self-monitoring. Deterring intentional wrongdoing, however, requires additional liability. In circumstances explained in Part V. B., optimal deterrence requires increasing compensatory damages by the reciprocal of the enforcement error.


3. The most thorough and far-reaching application of deterrence theory to punitive damages is found in Polinsky & Shavell, Punitive Damages, supra note 1.
4. See id. at 957 (providing sample jury instructions).
II

PUNITIVE DAMAGES IN PRACTICE

“Punitive damages” usually refers to damages awarded to the plaintiff in a civil trial in order to punish the defendant. Private law in most countries prohibits or narrowly circumscribes punitive damages, perhaps because civil procedures afford less protection to the defendant than criminal procedures. United States statutes occasionally prescribe punitive damages, most notably in antitrust cases where the law allows treble damages. However, statutory authorization for punitive damages is unusual, and courts seem reluctant to take the initiative in fashioning such remedies. The common law traditionally allowed punitive damages only in a narrow class of torts where the injurer typically harmed the victim’s dignity, such as slander. American state courts and legislatures, however, have expanded this practice beyond dignitary torts to allow punitive damages in many different kinds of civil cases, provided the defendant’s behavior is bad enough.

The standard for “bad enough” is set so high that few cases qualify. Most states have a statute concerning the incidence and extent of punitive damages. According to the usual formulation, punitive damages can be awarded when the defendant’s behavior is malicious, oppressive, gross, willful and wanton, or fraudulent. The statutes typically identify factors for the court to consider when setting the amount of punitive damages, including the seriousness of the wrong and the ability of the defendant to pay damages.

These phrases provide guidelines for courts, but the guidelines have never been formulated into rules. Punitive damages are supposed to bear a “reasonable relationship” to compensatory damages, but the law does not prescribe anything definite. For example, the law does not say whether punitive damages should exceed compensatory damages by two times or 100 times. In the typical case involving punitive damages, the jury is asked to decide whether punitive damages should be awarded, and, if so, how much to award. The jury, however, does not receive definite instructions for answering either question. Instructions are so thin that the jury’s intuitions determine damages.

Since juries receive minimum guidance from judges, awards are unpredictable. To illustrate, the ratio of punitive damages to compensatory damages awarded recently by juries falls as low as 1/10 and rises to a number approaching infinity as compensatory damages approach zero. In one recent case, the

6. See, e.g., CAL. CIV. CODE § 3294 (a) (West 1998). The California Civil Code provides the following: “In an action for the breach of an obligation not arising from contract, where the defendant has been guilty of oppression, fraud, or malice, the plaintiff, in addition to the actual damages, may recover damages for the sake of the example and by way of punishing the defendant.” That is not much detail to decide disputes involving millions of dollars.
7. The ratio of 1/10 is found in WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 305 tbl.10.7 (1987).
jury awarded $3 billion in punitive damages. Many private insurance contracts do not cover liability for punitive damages, and some states outlaw insurance against liability for punitive damages. Although rarely awarded, punitive damages represent a significant risk to some American defendants. Little wonder that defendants continue to argue (without success so far) that punitive damages fall under the prohibition of excessive fines by the Eighth Amendment to the United States Constitution. Jury intuition about the extent of punishment provides an uncertain and unstable foundation for law.

III

Why Punish?

In economic models of liability, “perfect compensation” leaves the victim indifferent between no harm and harm with compensation. In other words, perfect compensation restores the victim to the same indifference curve as if no injury had occurred. A liability system with perfect compensation would make victims indifferent about the behavior of injurers. If victims are indifferent, then injurers should be free to decide how to act. These facts pose the theoretical question, “Why not permit everything and hold injurers liable for the consequences?” To develop the foundation for a theory of punitive damages, I will sharpen this question and answer it.

A. Internalize or Deter?

Permission to do something, provided that one pays a price set by the court, has been called a liability right. To illustrate, companies may be permitted to engage in activities that create a nuisance provided that they compensate the victims. Alternatively, law may forbid people to harm others and punish those who intentionally violate the prohibition. To illustrate, law and morality forbid people to commit fraud or drive recklessly.

Liability rights and prohibitions locate decisions differently. For a liability right, the injurer is free to decide whether or not to act, provided he pays damages. In contrast, punishment subordinates the individual’s freedom to the group’s judgment about right and wrong. With a liability right, perfect compensation internalizes the harm caused by the injurer. With a prohibition, punishment ideally deters wrongdoing.

To understand deterrence, consider the parallel between victim’s compensation and injurer’s disgorgement. Disgorgement makes the injurer give up his

10. The concept of perfect compensation is explained and developed in several passages in ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 297-98, 363-65, 444-48 (1996).
gains from the injury. “Perfect disgorgement” is a sum of money that leaves the injurer indifferent between the injury with liability for damages or no injury.\textsuperscript{12} In other words, perfect disgorgement restores the injurer to the same indifference curve as if no injury had occurred. Damages that exceed the amount required for perfect disgorgement are extra-disgorging. Extra-disgorging damages make the injurer worse off than if he had not committed the injury.

By definition, an injurer is indifferent between no injury and an injury with liability for perfectly disgorging damages. Consequently, an injurer who faces certain liability for extra-disgorging damages prefers not to cause the injury. When the liability system works perfectly, liability for extra-disgorging damages deters intentional wrongdoing.

\begin{table}
\centering
\caption{Liability Right Versus Prohibition}
\begin{tabular}{|l|l|l|l|}
\hline
\text{Injurer’s Act} & \text{Law’s Goal} & \text{Damages to Achieve Goal} \\
\hline
Liability Right & Permitted & Internalize & Perfect Compensation \\
Prohibition & Forbidden & Deter & Extra-disgorging \\
\hline
\end{tabular}
\end{table}

Table 1 summarizes the contrast between a liability right and a prohibition. A liability right gives freedom to injurers and perfect compensation fully protects victims. Furthermore, when the liability system works perfectly, liability for perfectly compensatory damages causes a potential injurer to balance his costs of precaution and the victim’s costs of harm, as required for economic efficiency. Liability for perfectly compensatory damages provides incentives for efficient precaution by injurers in many economic models.\textsuperscript{13} In contrast, a prohibition forbids wrongdoing and deters it with extra-disgorging damages.

Notice that Table 1 implicitly suggests two different baselines for deciding when damages become punitive. Damages may be described as punitive when they exceed the amount required to compensate the plaintiff for the injury. According to this definition, punitive damages are extra-compensatory. Alternatively, damages may be described as punitive when they exceed the amount required for the injurer to disgorge the gain from wrongdoing. According to this definition, punitive damages are extra-disgorging. The two definitions of punishment reverse the role of victim and injurer in the baseline.

Although similar in form, damages computed under one definition do not generally equal damages computed under the other definition. The two con-


\textsuperscript{13} The first formalization of this idea for torts is in John Brown, Toward an Economic Theory of Liability, 2 J. Legal Stud. 325 (1973). Perfect compensation also destroys the victim’s incentives for precaution. Any modern discussion of bilateral precaution explains the problem. For example, see Cooter & Ulen, supra note 10, at 365, 409, 435.
cepts of punishment are not materially equivalent. When the goal of law is to
deter injurer’s from committing forbidden acts, the injurer’s gain, not the vic-
tim’s loss, provides the correct baseline for computing damages. Under ideal
conditions—including certain liability—a small increment added to perfectly
disgorging damages will deter. When liability is uncertain, the increment must
increase enough to offset the uncertainty, as I discuss later in detail. The ideal
conditions, which are unrealistic but useful for analysis, include no court errors
and actors who are rational, informed, and solvent.

B. Channeling Transactions

Now I return to the question, “Why not permit everything and hold injurers
liable for the consequences?” A variety of answers can be found both within
and outside of the framework of economic analysis. Some scholars who work
within the economic framework concentrate on the desirability of channeling
transactions through markets rather than courts. To illustrate, the law prohib-
its you from stealing my golf clubs. If you want them, you have to buy them
from me. Alternatively, the law might permit you to steal my golf clubs and
compensate me. For most transactions, the market determines values at less
cost than a court. For example, the market determines how much I value my
golf clubs at less cost than a court can determine liability. So protecting my en-
titlement by a property right rather than a liability right saves transaction costs
by channeling transactions into markets and away from courts. When markets
can price goods at lower transaction costs than courts, prohibitions provide a
more efficient structure for transactions than liability.

C. Uncompensable Losses

Instead of the transaction structure, I focus on a different reason why effi-
ciency sometimes requires prohibitions rather than liability. Implementing per-
fected compensation is easy when the harmed object has a perfect substitute in
the market. To illustrate, if one motorist destroys another’s automobile, the
market price enables the court to estimate perfectly compensatory damages.
Implementing perfectly compensating damages is difficult when the injury
harms a unique object with an imperfect substitute, such as an oil painting. Im-
plementing perfect compensation may be impossible when the injury seriously
harms a subject rather than an object. To illustrate, a parent may feel that no
amount of money could make him indifferent over the death of his child. If a
grieved parent feels that way, then perfectly compensatory damages are impos-
bable in principle. Impossibility comes from the fact that the parent cannot
make any sense out of the idea of being indifferent between his child and
money. (Some scholars conclude that money and children’s lives are incom-
mensurable.14)

14. See Margaret Jane Radin, Compensation and Commensurability, 1993 Duke L.J. 56, 67; Cass
R. Sunstein, Is the Rule of Law Good? (1994) (paper read at first draft of Tanner lectures). See gener-
D. Liability for Uncompensable Losses

In spite of this fact, courts routinely award “compensatory damages” to the victims of uncompensable harms. In cases where compensation is impossible, the phrase “compensatory damages” badly misleads people concerning the purpose of damages. Legal language obscures another purpose for damages in such cases. The purpose is to deter people from exposing others to unreasonable risks. I will explain how to implement this goal through damages.

Social norms often restrict and regulate behavior that causes uncompensable losses. To illustrate, conventional morality forbids assault. Similarly, current social norms in California apparently require the driver to make sure that all children in cars use seat belts. Social norms, however, often do not stipulate the damages owed to victims of wrongdoing. To illustrate, social norms do not specify how much to compensate someone who was punched in the nose or who lost a child in an accident. Given these facts, juries find obligations more reliably than they do damages. For example, juries feel more confident about an obligation to reduce a particular risk to a child’s life than about the damages that would compensate the parent if the risk materializes.

When compensation is impossible, an appropriate goal of liability law is to deter wrongdoers from violating community standards. Setting damages at the appropriate level makes rationally self-interested people conform to social norms. To illustrate, successful deterrence prevents adults from exposing children to more risk than allowed by social norms. I can describe the method of computing damages for deterrence by using Judge Hand’s notation from United States v. Caroll Towing, Co.15

Let B denote the burden of care, let p denote the reduction in the probability of an accident caused by taking B, and let L denote liability. Judge Hand had in mind a situation where no community standard exists. If there is no community standard, Judge Hand required the actor to take precaution until the burden is at least as great as the reduction in expected liability: B > pl.

I have in mind a different situation where a community standard exists but the harm is uncompensable. The community standard establishes the burden B whose omission constitutes negligence. Objective facts establish the reduction in the probability p of an accident caused by B. Given B and p, the court should compute L by solving L = B/p. I call this approach the “risk-equivalent” method, because it aims for equivalence between money and the risk, not equivalence between money and the loss.16

15. 159 F.2d 169 (2d Cir. 1947).
To illustrate, consider the computation of damages for the accidental death of a child. Courts find standards of negligence in social norms and refine them. If social norms as found by courts require an actor to spend $100 on precautions to reduce the probability of a child’s accidental death by .0001, then the equivalent value of the risk equals $100/.0001, or $1 million. In this example, the burden $B$ equals $100, the probability $p$ equals .0001. These facts imply that the liability $L$ should equal $1 million.

Rationally self-interested actors decide whether or not to conform to the legal standard of care by balancing the costs of precaution and the expected costs of liability. The risk-equivalent method sets liability at the level that causes such actors to conform to the legal standard. When $L = B/p$, a rational actor will pay up to $B$ to avoid facing liability with probability $p$. For example, when faced with liability of $1 million, the rationally self-interested actor will spend up to $100 to reduce the probability of being liable by .0001. To illustrate concretely, a profit-maximizing bottler of Coca-Cola will spend up to $100 on quality control to reduce by .0001 the probability of being liable in tort for $1 million.\(^{17}\)

Current regulations create incentives for irrational and inconsistent levels of precaution. To illustrate, Viscusi found that the cost per life saved by various regulations varies from $100,000 for unvented space heaters to $72 billion for the 1987 formaldehyde standard.\(^{18}\) The risk-equivalent method for computing liability can help to rationalize precautions by providing uniform incentives. To illustrate, $1 million in liability for the wrongful death of a child creates incentives for firms to spend $100 on precaution to reduce the probability of death by .0001, regardless of whether the firm manufacturers space heaters or formaldehyde.

E. Bad Names

Sometimes the legal methodology for computing compensatory damages omits categories of harm. To illustrate, the law may fail to compensate fully for consequential damages, remote damages, economic losses, exposure to risk, emotional distress, or pain and suffering. To achieve the goal of internalization, the law needs to fill the gap between allowed compensation and perfect compensation. One way to fill the gap is by allowing additional damages and calling them “punitive.”


17. The same proposition can be stated in terms of “perfect disgorgement,” which I defined as the damages that leave the injurer indifferent between doing right and doing wrong. A rationally self-interested actor must face damages of $1 million in order to be indifferent between committing a wrong that results in liability with probability .0001 or doing right at a cost of $100. A distinction can be drawn between ex ante and ex post disgorgement. From the ex ante viewpoint, perfect disgorgement of the expected gain from wrongdoing requires damages equal to $1 million. From the ex post viewpoint, perfect disgorgement of the actual gain from wrongdoing requires damages equal to $100.

Gap filling was apparently one of the original inspirations for awarding punitive damages in common law cases involving dignitary torts. To illustrate, consider the difficulty of estimating the damages required to compensate perfectly for the indignity of getting insulted and punched in the nose while peacefully enjoying a beer in a public house. Courts lack a methodology for computing compensatory damages in such cases. Punishment might fill the methodological gap and provide damages approximating the amount required for compensation.

Applying “punitive” to gap-filling damages spreads confusion. If the law’s goal is perfect compensation of the victim, then the court should not use language suggesting that its goal is punishing the injurer. Instead, the law should adopt the best available methodology for implementing perfectly compensatory damages and say so. The best method to pursue a goal will advance it farther than pursuing a different goal by a better method.

IV
WHEN TO PUNISH

Channeling activities into voluntary transactions and deterring uncompensable harms sometimes requires backing prohibitions with penalties. Now I turn to the problem of when to impose punitive damages on wrongdoers. One writer has described a theory that judges and juries award punitive damages when, upon hearing the facts, they grip the arms of their chairs and think, “Oh my God!”

This claim explains little and suggests much. I will develop a theory of punitive damages by trying to explain such behavior.

A. Commitment and Indignation

Moral commitment increases trust, which makes cooperation easier and more productive. In a cooperative venture, people often get paid according to their contribution. Consequently a person who can signal moral commitment often gains an advantage in cooperative activities. Signaling commitment is difficult, however, because people can fake moral commitment. Gaining an advantage by commitment requires signaling that it is genuine.

Feelings can help to distinguish between fake and genuine commitment. Genuine moral commitment has an emotional aspect. People feel committed to internalized values. This connection between commitment and emotion has a useful function. Telling a cool lie is easier for many people than faking emotion. For example, a child who can tell a cool lie is often incompetent at faking emotion. Aspiring actors devote much time and effort to perfecting the art of faking emotion. Since faking emotion is difficult, the emotion attached to the

expression of moral commitment helps authenticate it. According to one theory, emotions evolved among people partly to provide the means to signal commitment.21

Whereas economic rationality seems relatively cool, discussions of law can become relatively hot. The heat comes from the connection between emotion and commitment. People feel committed to values at issue in contested law, such as abortion, discrimination, or even closing the range to cattle. The presence of emotion suggests a prominent role in law for expressing internalized values. People who internalize social norms feel righteous anger against those who violate them. The institution of punitive damages allows judges and juries to express righteous anger through speech and punishment.

B. Systems of Social Norms

Punitive damages express indignation through the language of the law and the magnitude of the punishment. Expression of emotions by the court demonstrates its commitment to the law in question. Perception of commitment can shape the expectations of citizens and change their behavior. Usually, perception of the court's commitment changes behavior marginally by deterring wrongdoing. Sometimes, however, perception of the court's commitment changes behavior dramatically.

The system of social norms determines whether signaling commitment causes marginal deterrence or dramatic change. With a uniquely stable equilibrium in the system of social norms, the law's sanction changes behavior marginally. In contrast, with multiple equilibria, some of which are unstable, the law's expression can sometimes change expectations dramatically and permanently. Jumps occur because the court's commitment focuses expectations on a new equilibrium. If coordination problems prevent people from reaching a higher standard of behavior to which they aspire, law can sometimes solve the coordination problem by urging people to achieve the higher standard.

C. Graphs22

Before turning to deterrence, however, I will elucidate the preceding discussion of legal expression with the help of some graphs. The reader who dislikes graphs can skip to the next section.

Many productive activities require cooperation among people. Norms provide a guide to organization and a basis for trust that facilitates cooperation. Organized activity typically increases production above the level obtainable without organization. Some organizations, however, are superior to others. Social norms do not automatically prescribe the most productive organization. In addition, organizations distribute power and status. Social norms do not

automatically distribute power and status fairly. I will briefly describe some formal characteristics of norms relevant to their efficiency and fairness.

A social norm imposes an obligation that partitions the set of possible actions into permitted and forbidden zones. People conform to a norm by staying in the permitted zone, and people violate a norm by entering the forbidden zone. When an actor adopts the pure strategy of doing right or the pure strategy of doing wrong, the resulting payoff depends upon the strategy pursued by others. I assume that doing right contributes to production by increasing trust and cooperation, whereas doing wrong reduces production by undermining trust and cooperation. For example, truth-telling and promise-keeping contribute to production by solving the principal-agent problem.

Given a socially productive norm, an increase in the proportion of wrongdoers decreases the productivity of cooperative relationships. A decrease in productivity reduces the payoffs to wrongdoers and rightdoers. Although everyone's payoffs fall, the reduction may be different for wrongdoers and rightdoers. These differences have important consequences for equilibria in the system of norms. In evolutionary equilibrium, all behavior that persists yields the same objective payoff. In one common pattern, as the proportion of wrongdoers increases, the payoffs to wrongdoers fall faster than the payoffs to rightdoers. This possibility results in a stable interior equilibrium.

Figure 1 depicts this possibility. The vertical axis represents the payoff, the horizontal axis represents the proportion of wrongdoers, one curve represents payoffs for conforming to the norm, and the other curve represents payoffs for violating the norm. The intersection in the curves depicts an interior equilibrium where rightdoers and wrongdoers receive the same expected payoff.

**Figure 1**

**STABLE EQUILIBRIUM**
Now consider why the equilibrium is stable. If the proportion of wrongdoers increases beyond the equilibrium level, the payoff to rightdoers rises above the payoff to wrongdoers. With rightdoers receiving higher payoffs, some wrongdoers change their behavior. The number of wrongdoers declines until equilibrium is restored. Conversely, if the proportion of wrongdoers decreases below the equilibrium level, the payoff to wrongdoers rises above the payoff to rightdoers, so wrongdoers will increase in number until equilibrium is restored.

As explained, stability results from assuming that an increase in the proportion of wrongdoers harms wrongdoers more than rightdoers. Now change the assumptions and assume that an increase in the proportion of wrongdoers harms wrongdoers less than rightdoers. Figure 2 depicts this situation. The intersection of the curves represents an interior equilibrium that is unstable. Beginning from the interior equilibrium, an increase in wrongdoers causes the payoffs of wrongdoers to rise above the payoffs to rightdoers, so the number of wrongdoers continues to increase. The process ends at the stable equilibrium at the lower right corner where everyone does wrong.

Conversely, beginning from the unstable interior equilibrium, a decrease in wrongdoers causes the payoffs of wrongdoers to fall below the payoffs to rightdoers, so the number of rightdoers increases. The process ends at the stable equilibrium at the upper left corner where everyone does right.

For norms of cooperation, the curves expressing payoffs to rightdoers and wrongdoers slope down to express the loss in productivity as the proportion of wrongdoers increases. Theory, however, does not prescribe the gradient of the curves. The curves might intersect more than once. Figure 3 depicts this possi-
bility. Figure 3 has a stable interior equilibrium with few wrongdoers, an un-
stable interior equilibrium with many wrongdoers, and a stable corner equilib-
rium with all wrongdoers. In Figure 3, the system stabilizes when few people
(25%) do wrong or when everyone (100%) does wrong.

**Figure 3**
**Multiple Equilibria**

Now I discuss how law can influence the systems of social norms depicted in
Figures 1, 2, and 3. Assume that lawmakers raise the social obligation to the
level of a legal obligation and attach a sanction to its violation. The legal sanc-
tion causes the wrongdoer’s payoff curve to shift down in the figures. If the
wrongdoer’s curve shifts down in Figure 1, then the equilibrium moves to the
left. Because wrongdoing decreases, the new equilibrium has more cooperation
and higher production. Given a stable interior equilibrium in the system of so-
cial norms, imposing legal sanctions deters wrongdoers marginally.

Deterrence, however, has different consequences in Figure 2. Recall that
the interior equilibrium is unstable. Since the system does not settle at the in-
terior equilibrium, small changes in the interior equilibrium have little or no
consequences for the social system. Attaching a legal sanction to the social
obligation causes the wrongdoer’s payoff curve to shift down in Figure 2. A
downward shift in the wrongdoer’s payoff curve shifts the interior equilibrium
and does not change the corner equilibria. Either everyone continues doing
wrong, or else everyone continues doing right. In either case, the imposition of
a legal sanction has no deterrence effect.

---

23. Changes in the interior equilibrium affect the size of the shock needed to cause the system to
jump from one corner equilibrium to another corner equilibrium.
Besides deterring wrongdoing, law coordinates expectations. Given a system of social norms with multiple equilibria, the coordination effects of law can be dramatic. In Figure 2, everyone does right or everyone does wrong. Assume that everyone does wrong. Now the state raises the social obligation to the level of law. If people respect law, enacting the obligation into law may cause people to change their expectations. If people think everyone else will switch from doing wrong to doing right, then the prophecy will fulfill itself. Enacting the law will cause the equilibrium to jump from the right corner, where everyone does wrong, to the left corner, where everyone does right.

The expressive theory of law represents an alternative to the deterrence theory of law. According to the expressive theory of law, the expression of social values is an important function of the courts, or possibly the most important function of the courts. The classical formulations, however, did not analyze the behavioral consequences of expressive law. The main finding of such an analysis is that, given multiple equilibria in a system of social norms, expression of legal commitments can cause dramatic jumps in behavior.

I have explained that legal sanctions change behavior marginally in a stable system of social norms, and legal expression can change behavior non-marginally in an unstable system of social norms. In reality, normative systems are likely to be mixed, as depicted in Figure 3. In mixed systems of social norms, law can have deterrence effects and expression effects. To illustrate by using Figure 3, assume that the system of social norms is stuck at the right corner where 100% of the actors do wrong. A stable interior equilibrium exists where only 25% of the actors do wrong. Everyone’s payoff would increase if the system could move from the corner equilibrium to the stable interior equilibrium.

Notice that the unstable, interior equilibrium in Figure 3 occurs with 70% wrongdoers and 30% rightdoers. In order to move to the stable interior equilibrium, at least 31% of the actors must change strategies and do right. Once 31% of the actors do right, rightdoers will continue to increase until 75% of the actors do right. Assume that enacting a law induces at least 31% of the citizens to change, so the system in Figure 3 moves to the stable interior equilibrium. Although improvement is dramatic, 25% of the actors continue doing wrong. Further reductions in wrongdoing would require deterrence. Supplementing social sanctions with legal sanctions shifts down the curve representing expected payoffs to wrongdoers in Figure 3, thereby reducing the equilibrium number of wrongdoers. Thus, the combination of expression and deterrence brings wrongdoing down to a low level.

24. For an argument that expressing social judgments is one of the uses of criminal law, see H.L.A. Hart, Punishment and Responsibility (1968).
25. This was apparently Durkheim’s view, as analyzed in David Garland, Punishment and Modern Society: A Study in Social Theory 23-46 (1990). Note that I draw no connection between the emotive theory of law, which belongs to jurisprudence, and the emotive theory of the meaning of value, which belongs to epistemology.
I have explained how law can change behavior dramatically in a social system with multiple equilibria. In such a social system, law can also fail dramatically. To illustrate concretely by Figure 3, recall that dramatic improvement occurs when, beginning from a situation in which everyone does wrong, enacting the law causes at least 31% of the actors to change and do right. Dramatic failures occur if less than 31% of the actors change and do right. Assume that enacting the law causes 25% of the actors to change and do right. The actors change from wrong to right because they expect other people to make the change. But not enough people make the change to sustain it. In that event, the system reverts to the original equilibrium where everyone does wrong.

In a system of social norms with a uniquely stable equilibrium, the law’s sanction changes behavior marginally, but the law’s expression cannot permanently change expectations or behavior. In a system of social norms with multiple equilibria, the law’s sanction changes behavior marginally, and the law’s expression dramatically succeeds or dramatically fails to change expectations and behavior. Using punitive damages for marginal deterrence risks little, whereas using law expressively risks a lot. The high risk of expression is that law can change social behavior dramatically or discredit itself.

V

HOW MUCH TO PUNISH?

Given the complexity of coordination problems and the fragility of symbols, I doubt that expressive theories of law can provide the basis for calibrating punitive damages. If punitive damages are based upon expressive law alone, the current problems of inconsistency have no remedy. Inconsistent awards of punitive damages muddle the message conveyed by the severity of punishment. For purposes of expression, the courts need a more consistent calibration of punitive damages than an expressive theory can provide. The goal of deterrence can provide a more consistent calibration of punishments. I will analyze the use of punitive damages for marginal deterrence, which leads to more precise conclusions about punishment.

I explained earlier that punitive damages arise only when the defendant behaved very badly. In the usual case, only intentional wrongs, or nearly intentional wrongs, are bad enough for punitive damages. The extent of damages thus depends upon distinguishing between intentional and unintentional wrongs. I will develop a distinction appropriate to an economic theory of damages.

A. Lapses

Table 1 describes incentive effects that influence rational actors. Tort liability, however, often arises from diminished rationality. Rationality diminishes when people mistakenly assess facts. In cases of unintended wrongs, the actor does not know that he is doing something wrong. To illustrate, I may drive too fast because I do not notice that I am near a school. When a wrong is
unintended, the actor does not foresee possible punishment, so it cannot deter him. Similarly, in cases of under-appreciated risks, the actor under-estimates the risk that he imposes on others. To illustrate, I mistakenly think that asbestos dust poses only a small risk of lung cancer to my employees. When a risk is under-appreciated, liability will not cause the actor to internalize the full benefit of precaution.

Lapses in attention often cause unintended wrongs and under-appreciated risks. More vigilance by the actor reduces the frequency of lapses in attention. By “vigilance” I mean self-monitoring activities by which an actor controls his behavior. An account of rational self-monitoring provides the foundation for a theory of deterrence applied to unintended wrongs and under-appreciated risks.

A rationally self-interested actor balances the cost of self-monitoring and the reduction in expected liability from lapses. Liability for perfectly compensatory damages makes the injurer internalize the expected cost of harm. So, assuming certain ideal conditions including perfectly compensatory damages, a liability rule causes efficient self-monitoring by a rational actor.

As explained, perfect compensation is impossible in principle for wrongs resulting in uncompensable harms. I propose to compute damages in such cases by the risk-equivalence method. Risk-equivalent damages create incentives to rationalize precaution. Specifically, risk-equivalent damages provide incentives to equalize the increase in safety obtained by investing more resources in all types of precaution, including self-monitoring. So, assuming certain ideal conditions including risk-equivalent damages, a liability rule causes efficient self-monitoring by a rational actor.

The law often does not apply a legal standard to self-monitoring. To illustrate, consider failing to stop for a stop sign. Failing to stop at a stop sign is negligent, and the driver is liable for the consequences. Everyone, however, occasionally lapses and fails to stop at stop signs. The law, however, does not

---

26. The assumptions include costless, error-free application of liability rules by courts. Also included are certain assumptions about the connection between lapses and liability. Note that actors can have deficient incentives for self-monitoring under a negligence rule when some lapses result in non-negligent injuries. For example, a lapse in my attention causes me to apply the breaks slower than otherwise, thus resulting in an accident. The law, however, may decide that I applied the breaks soon enough to satisfy the legal standard of care. Consequently, the law concludes that I was not negligent and not liable. In this example, I externalize part of the benefit of more intensive self-monitoring. Specifically, I externalize the reduction in non-negligent harms caused by more intensive self-monitoring. In this case, liability law does not create incentives for efficient self-monitoring.

This problem cannot arise when all injuries caused by lapses result in liability. To illustrate, consider self-monitoring of stop signs by drivers. Every time I unintentionally run a stop sign, I am negligent and liable for the resulting harm. Consequently, I internalize the reduction in harm caused by more intensive self-monitoring of stop signs. I assume that self-monitoring to stop at stop signs does not spill over to the reduction of liability from accidents with causes other than failing to stop at stop signs.

In this discussion, I have not distinguished between the two major forms of the liability rule: strict liability and negligence. The former is simpler than the latter, but my generalizations hold with some qualifications under both rules. For an especially careful explanation of the difference in the context of punitive damages, see Polinsky & Shavell, Punitive Damages, supra note 1, at 878-85.
allow the driver to defend himself by proving that he inadvertently runs stop signs less often than most other people do. A high level of self-monitoring provides no defense for a lapse.

In this respect, self-monitoring by individuals resembles quality control by manufacturers. To illustrate, a Coca-Cola manufacturer has to monitor the bottles to make sure they contain no defects that could cause them to explode. More intensive and costly quality control results in fewer defective bottles. Defective bottles sometimes slip through even the best system of quality control. When a defective bottle injures a consumer, the law holds the manufacturer liable, without allowing a defense that its quality control was best.\(^{27}\) The manufacturer’s quality control resembles the driver’s self-monitoring, and the defective bottle resembles the driver running the stop sign. The court does not have the information to create and apply a standard of self-monitoring to individuals or a standard of quality control in manufacturing.

Now I return to the connection between intention and damages. Liability for unintended harms causes injurers to internalize the benefits of self-monitoring. Thus liability for unintended harms causes actors to balance the cost of self-monitoring against the benefit of fewer lapses, as required for efficient incentives. Conversely, imposing extra-compensatory damages measured relative to the baseline of perfect compensation or risk-equivalent damages, provides incentives for excessive self-monitoring. In other words, extra-compensatory damages create incentives for unreasonably costly vigilance. Thus the liability of injurers should typically equal, but not exceed, compensatory damages.

I used the concept of rational self-monitoring to build a theory of lapses. In this paper, I will not attempt to achieve greater realism by adjusting the model of rational self-monitoring in light of psychology.\(^{28}\)

B. Punishment

Instead of lapsing, however, some actors impose excessive risk on others by intentionally violating standards of law and morality. Deterring their wrongdoing requires punishment. I will distinguish three motives for intentional wrongdoing, each with its own logic of deterrence.

First consider intentional wrongdoing motivated by errors in enforcing compensatory damages. To illustrate, assume that perfect compensation of the victim requires damages of $10,000, and assume that problems of detection and proof cause the injurer to be held liable with probability 1/2. An injurer’s expected liability for compensatory damages thus equals $5,000. Assume the cost of compliance with the legal standard of care costs $6,000. Under these as-

\(^{27}\) See, e.g., Escola v. Coca-Cola Bottling Co., 150 P.2d 436 (Cal. 1944).

sumptions, the rationally self-interested injurer intentionally does wrong in order to save $1,000 in expected costs.

Punitive damages can offset enforcement errors and restore efficient deterrence to the liability system. In this example, a punitive multiple of 2/1 restores the injurer’s expected liability to $10,000, as required for deterrence. Since the punitive multiple of 2/1 equals the reciprocal of the enforcement error 1/2, I call this method for computing punitive damages the rule of the reciprocal. The rule of the reciprocal applies to most cases of fraud by individuals and self-seeking by firms.

When applying the rule of the reciprocal, courts should not award punitive damages in cases like the Exxon Valdez, where an oil tanker ran aground. Since the whole world was sure to know about the accident, a rational actor like this large corporation would not expect to escape liability.

Some kinds of enforcement errors do not distort incentives and do not require correction by adjusting damages. Expected costs usually jump at the boundary between permitted and forbidden acts. Because of the jump in costs, most rational actors will conform to the legal standard in spite of enforcement errors. If costs jump by a large amount at the legal standard, then a damage multiple is unnecessary to offset enforcement errors and to deter wrongdoing by rational actors. Conversely, if costs jump by a small amount at the legal standard, then reciprocal damages are necessary to offset enforcement errors and to deter wrongdoing by rational actors. In general, the necessity of reciprocal damages to deter wrongdoing depends upon the relative size of the enforcement error and the jump in liability costs at the legal standard.

Malice is a second motive for intentional wrongdoing. The assailant who insults me and breaks my nose presumably derives something valuable from his act, such as satisfaction. Subjective satisfaction is difficult to observe and often impossible to measure. Consequently, no rule exists for computing the punishment required to deter malicious acts. Although tailoring punishment to individuals is difficult, statistical evidence can show the average relationship between punishment and deterrence. As with criminal fines, the calibration of punitive damages for malice must rely upon such aggregate statistics.

Self-seeking firms may pursue profits unscrupulously but not maliciously. Malicious acts by employees divert the firm’s resources from the pursuit of profits. The directors and managers of the firm have an incentive to monitor employees to prevent the diversion of its resources. Consequently, malice by employees typically results from imperfect monitoring by the firm. Imperfect monitoring of employees by the firm resembles imperfect self-monitoring of the individual, which brings me to the next kind of wrongdoing.

A third motive for wrongdoing occurs in cases involving reckless disregard for the safety of others. Actors who disregard risks are often in special emo-

tional states, such as the young man suffering from Saturday night fever. Heightened emotions can cause excessive discounting of risk and futurity. In such circumstances, the actor needs the threat of punishment to acquire better discipline and control over his emotions. Efficient self-monitoring requires punitive damages in such cases, not merely compensatory damages. Social scientists have not developed or tested precise theories about how punishing reckless behavior causes people to gain self-control. (An elaboration of the model of lapses can provide the foundation for developing such a model.30) Consequently, I know of no rule for computing optimal punishment in such cases.

A similar analysis applies to a firm. Idiosyncratic directors and managers may apply very high discount rates to risk and futurity, which results in reckless behavior. Deterring such behavior requires the punishment of the directors and managers. I know of no rule for computing optimal punishment in such cases. Similarly, a lapse in monitoring by directors and managers may allow a deviant employee recklessly to endanger customers or the public. The imposition of punishment on the firm, as opposed to the individual wrongdoers, punishes the stockholders or consumers for not monitoring the directors, managers, or employees who committed the wrong.

VI

CONCLUSION

Law and morality forbid acts that cause uncompensable harms or create unreasonable risks. When people intentionally commit such acts, courts express anger and indignation at the wrongdoer through speech and punishment. Civil courts correctly draw upon the righteous anger of jurors over intentional wrongs to chastise the actor and harm his reputation. In a system of social norms with multiple equilibria, expressions of emotion signal commitment and coordinate expectations. In a system of social norms with a unique, stable equilibrium, sanctions deter marginally. Social norms, however, provide a better guide to the need for punitive damages than to their extent. Courts make a serious error in asking jurors to assign punitive damages without providing detailed instructions on their computation. The court in a typical case should impose the minimum punishment required to deter wrongdoing.