Economic Analysis of Punitive Damages: Theory, Empirics, and Doctrine

Catherine M. Sharkey
NYU School of Law, catherine.sharkey@nyu.edu

Follow this and additional works at: http://lsr.nellco.org/nyu_lewp

Part of the Law and Economics Commons, Remedies Commons, and the Torts Commons

Recommended Citation
http://lsr.nellco.org/nyu_lewp/289

This Article is brought to you for free and open access by the New York University School of Law at NELLCO Legal Scholarship Repository. It has been accepted for inclusion in New York University Law and Economics Working Papers by an authorized administrator of NELLCO Legal Scholarship Repository. For more information, please contact tracythompson@nellco.org.
1. INTRODUCTION

Punitive damages have been a part of the civil law landscape in the United States since the nineteenth century, but the past two decades have witnessed a firestorm of renewed interest and debate over this supra-compensatory remedy, whose goals are to punish and to deter wrongful behavior.

At first glance, this intense interest may seem puzzling given how rarely punitive damages are awarded. Punitive damages are at the tip of the iceberg in the civil justice system. The number of cases going to trial is very small, on the order of 3–5 percent of civil cases filed in state court (Galanter, 2004: 509 tbl.5). Of those tried cases, punitive damages are awarded in a small minority—less than 5 percent of cases in state courts (Del Rossi & Viscusi, 2010: 138 n.36; Eisenberg & Heise, 2011: 6; Hersch & Viscusi, 2004: 24 tbl.2).

So, what explains the sustained interest in what is, as an empirical matter, a very extraordinary, rarely imposed remedy? First, punitive damages are awarded much more frequently in certain types of cases. Thus, the aggregate 5 percent figure might underestimate the significance of punitive damages, at least for certain categories of cases. In cases of fraud, slander/libel, intentional torts, employment discrimination, and products liability, punitive damages are more common, and are awarded in 12–30 percent of the cases that go to trial (Sharkey, 2003a: 351 & n.12). In addition, a recent empirical study has shown that, in tried cases in which the plaintiff prevailed and punitive damages were sought (a variable usually not present in datasets), the estimated overall rate of issue of punitive damages was 29 percent (Eisenberg et al., 2010a: 9–10 tbl.3).

Second, even if punitive damages are rarely awarded, the magnitude of the awards (known to reach into the millions, and even billions, of dollars) may exert an oversized influence on our litigation system. Defendants (especially if risk averse) will take precautionary measures—perhaps even excessive ones—to avoid facing the risk of a small probability of a very large damages award. Moreover, the infrequency of punitive damages awards may obscure a much larger “shadow” effect on settlements: defendants may be driven to settle rather than face potentially crippling punitive damages.

The statistics cited above are taken from comprehensive surveys of state courts around the country, giving rise to a fairly detailed portrait of punitive damages in average, mine-
run cases. The datasets are comprehensive and representative, but, as a general rule, have not picked up many of the rare “blockbuster” punitive damages awards that, some argue, are the real problem. So, a raging empirical debate persists, with one camp in the legal academy asserting, based upon state court data, that punitive damages are predictably related to the size of compensatory awards and relatively stable over time (Ted Eisenberg and various collaborators), and a competing camp arguing that when the blockbuster punitive awards are analyzed separately, punitive damages awards appear indefensible, or at least not rationally related to compensatory damages (Kip Viscusi and collaborators).

Third, punitive damages, regardless of their frequency or magnitude, maintain scholars’ and practitioners’ active attention because their underlying theoretical justifications raise interesting questions about the very role of the tort system. Punitive damages connote punishment, which, at least initially, would seem to be the domain of criminal law, not the civil justice system. Indeed, punitive damages were historically awarded only in cases of malice or willful and wanton conduct, a subset of intentional tort cases. The paradigmatic case was that of intentional battery or assault, including acts of physical violence and dignitary affronts such as spitting upon one’s adversary (Sharkey, 2003a: 359 n.23). The standard verbal formulations of the doctrine require mental states ranging from “intent to harm without lawful justification or excuse,” to “reckless disregard of the interests of others.” Gross negligence typically does not suffice. With its emphasis on malice and willful and wanton conduct, the conventional definition focused on egregious conduct, or—in economic terms—conduct that substantially deviated from the optimal level of care.

A newer generation of punitive damages cases, however, falls outside this narrow band of malicious, intentionally wrongful conduct and deals almost exclusively with the recklessness side of the equation (Sharkey, 2003a: 358 n.19, 364). Indeed, the gradual acceptance of insurance for punitive damages over the last 50 years stems, in part, from the evolution of punitive damages themselves: whereas punitive damages were once awarded predominantly for acts that satisfied malice aforethought or intentional wrongdoing, now many punitive damages awards arise from what was essentially accidental conduct, albeit committed recklessly (Sharkey, 2005b: 438–50).

This contemporary landscape of punitive damages cases presents new theoretical challenges. Should punitive damages be restricted to malicious or willful and wanton conduct? If not, when is it appropriate to award such supra-compensatory damages in the event of a defendant’s reckless conduct? Should theories treat knowing breaches that create a high risk of injury differently from accidents caused by recklessness?

This chapter addresses these issues from one vantage point: the economic perspective. The primary economic rationale for supra-compensatory damages—itself traceable back more than a century to Jeremy Bentham,1 but not formalized in the specific context of punitive damages until recent decades—is optimal deterrence (or loss internalization): when compensatory damages alone will not induce an actor to take cost-justified safety precautions, then supra-compensatory damages are necessary to force the actor to internalize the full scope of the harms caused by his actions. Alternative economic

---

rationales—disgorgement of ill-gotten gains and enforcement of property rights—have been proposed to align the theory with the historical and conventional focus of punitive damages on intentionally wrongful behavior. The Calabresi-Melamed (1972) property rule/liability rule dichotomy provides one framework for choosing between the loss internalization (liability-rule) and gain elimination/voluntary market transfer (property-rule) models (Sharkey, 2003a: 368 n.56).

Notwithstanding its academic prominence, the economic deterrence rationale has not dominated doctrine. In fact, the U.S. Supreme Court has all but rejected economic deterrence, by instead placing increasing emphasis on a competing retributive punishment rationale. But, since punitive damages lie squarely within the purview of state law, state legislatures and courts possess a degree of freedom to articulate state-based goals of punitive damages—such as economic deterrence—even in the face of heavy-handed federal constitutional review imposed by the U.S. Supreme Court.

2. ECONOMIC GOALS OF PUNITIVE DAMAGES

A. Theory

The predominant law and economics theory of punitive damages is based upon optimal deterrence or loss internalization and focuses on the under-enforcement problem: supra-compensatory damages are needed when under-detection of harms or other factors leads to inefficiently low expected liability, which is insufficient to induce optimal care. Other contenders include gain elimination or disgorgement and inducement of voluntary transfers, also known as the property rights perspective.

The domain of optimal deterrence theory is a socially productive, yet externality-producing, activity. According to the “Hand formula” (put forth by Judge Learned Hand in United States v. Carroll Towing, 159 F.2d 169, 173–74 (2d Cir. 1947), and made the centerpiece of economic analysis of tort law by Posner (1972: 32–3)), defendants should be held liable for negligence if the burden ($B$) of taking a precaution is less than the probability ($P$) of harm resulting if the precaution is not taken, multiplied by the magnitude of the loss ($L$): $B < P*L$. In order to induce defendants to act efficiently, compensatory damages should be equal to the loss, resulting in expected liability equal to $P*L$.

Purely compensatory damages will not induce optimal care if negligent injurers expect to avoid liability for some of the harms they cause. This can occur for multiple reasons: under-detection; under-compensation; and other imperfections in the litigation system such as false negatives in adjudication and the cost of adjudication. The primary goal of punitive damages is to address the under-enforcement problem by increasing damages for

---

2 Cogent theoretical accounts of punitive damages from a non-economic perspective include Colby (2008); Hampton (1992); Sebok (2003); Zipursky (2005).

3 Alternative mechanisms (beyond the scope of this chapter) likewise respond to the under-enforcement problem. See Richard Epstein, Chapter 23 (this volume) (arguing for broad regulatory and criminal sanctions in cases that do not result in harm to offset the under-compensation in the cases that do).
Economic analysis of punitive damages

the harms that are detected and sanctioned by a sufficient amount to ensure that injurers’
expected liability equals the social cost of their negligence. Punitive damages are war-
ranted in all cases where under-enforcement is an issue; there is no reason to distinguish
between intentional and purely negligent harms.

By contrast, the domain for theories of gain elimination or prevention of wrongful
takings (sometimes referred to as “complete deterrence” theories) is intentional, conscious
wrongdoing. There is a distinct shift in focus away from losses suffered by the plaintiff
or society and toward defendant’s conduct. In this realm, one worries about under-
deterrence, but less so about over-deterrence. The primary goal of gain elimination is the
complete deterrence of socially unproductive activities; the primary goal of the property
rights model is facilitation of voluntary market transfers.

Optimal deterrence may be internally coherent (i.e., may explain the normative attrac-
tiveness of punitive damages properly conceived), but its fundamental premise fits poorly
with the prevailing historical and popular understanding of punitive damages as punish-
ment. The historical and conventional requirement for punitive damages that there be
malice or willful or wanton misconduct aligns more with the disgorgement theory or
property rule protection.

This could suggest that, although optimal deterrence theory may have considerable
force as a prescriptive law and economics theory, as a matter of positive (or descriptive)
law and economics analysis, the alternative gain elimination and property rights models
have the upper hand. But, the contemporary expansion of punitive damages into reck-
less indifference (most prominently in products liability cases) suggests additional room
for expansion of the loss internalization rationale, with both descriptive and prescriptive
payoff.

1. Optimal deterrence: loss internalization

The classic law and economics account of tort liability is one of cost internalization:
actors will have incentives to take reasonable care (i.e., cost-effective safety precautions)
as long as they are forced to pay for the harms that are caused by their taking unreas-
sonable risks (Shavell, 1980).

Note how the success of this optimal deterrence model hinges upon the validity of two
assumptions: (1) actors will in fact pay compensatory damages in each instance in which
they take unreasonable risks and cause harm to others, and (2) compensatory damages
can be set accurately to reflect the social cost of the harms inflicted.

Yet, in reality, negligent actors are not forced to pay for the harms that they cause
in each instance. Under-enforcement results from a variety of factors including under-
detection of wrongful conduct, failure to sue, plaintiffs’ inability to prove negligence and
causation, and error. Punitive damages can be used to mitigate enforcement error (Cooter,

Building on Becker (1968), Cooter (1989: 1148) proposes the “rule of the reciprocal”:
to determine total damages, one should multiply harm to the plaintiff by the enforcement
error.4 In cases of substantial enforcement error, utilizing the Hand formula, the potential

4 Where enforcement errors are difficult to determine, Cooter suggests a “rule of thumb”: the
cost of precaution divided by the expected harm stemming from its absence (Cooter, 1989: 1189).
tortfeasor might determine that it would be profitable for him to intentionally expose others to risk because the precaution costs that he is saving exceed expected liability costs (Cooter, 1989: 1149–53). In Cooter’s model, which utilizes a negligence standard of liability with due care set equal to the efficient level of care, punitive damages should be awarded only in the case of intentional acts, where there is substantial enforcement error, and where the defendant’s tortious conduct would allow him to derive “material benefit” that “more than offsets the risk of liability for compensatory damages alone” (Cooter, 1989: 1190).

In 1998, Polinsky and Shavell offered an elaboration of this optimal deterrence theory and proposed “a simple formula to assure that injurers will pay for the harms they cause”: “the total damages [compensatory plus punitive damages] imposed on an injurer should equal the harm multiplied by the reciprocal of the probability that the injurer will be found liable when he ought to be” (Polinsky & Shavell, 1998: 889). Polinsky and Shavell’s approach, linking under-detection to under-deterrence, harkens back to Bentham (1838–43), and Becker (1968) but offers a new gloss by providing a simple “punitive damages multiplier” for determination of punitive damages based upon the probability of detection. Polinsky and Shavell’s punitive damages multiplier is not tethered to intentional wrongdoing; it applies equally to breaches that are unintentional. In fact, Polinsky and Shavell exclude “malicious conduct”—which they define as benefit derived solely from causing harm to another—from their cost internalization model. Such conduct, unlike even reprehensible conduct that is nonetheless socially productive, should be deterred completely. In other words, benefits derived from socially illicit conduct do not count towards social welfare. Though they do not employ the terminology, in effect they insert a lone property rule amidst the loss-internalization liability rule domain.

Scholars have noted limitations of the multiplier model. Sharkey (2003a: 366–70) expanded upon the model, incorporating factors in addition to under-detection that could lead to under-liability, such as diffuse harms or negative value cases or cases in which the victim does not know the identity of the wrongdoer. Seen in this light, “the [Polinsky-Shavell] punitive multiplier, which focuses on underenforcement error due to nondetection of harms, is . . . a subset of the broader economic deterrence goal of internalization of total costs” (ibid. at 369).

Calfee and Craswell (1984: 1002) point to additional difficulties with achieving optimal deterrence via a negligence liability standard, given uncertainty. Whereas under-detection will give rise to under-compliance, the probability of detection is likely affected by the egregiousness of a defendant’s conduct (ibid. at 979–80). They highlight the risk that a punitive damages multiplier will induce over-compliance rather than optimal cost internalization where the probability of being found liable increases as a defendant’s conduct

In a later paper, Cooter widens the berth for punitive damages to include cases of “reckless disregard” (Cooter, 1997: 90–1).

Per Cooter’s model, when the enforcement error is small, it is unprofitable for the defendant to intentionally deviate from the appropriate standard of care and expose the victim to a degree of risk. This is so because the defendant’s costs jump substantially at the standard of care set by the court, not only because of the resulting compensatory damages, but also because of litigation and potential reputation costs (Cooter, 1989: 1156–9).
Economic analysis of punitive damages

becomes more egregious. They conclude that a multiplier is warranted only in the limited circumstances where (1) there exists no optimal level of the offender’s conduct (an “either/or” decision), (2) the probability of being found liable does not increase based on the conduct of the defendant, and (3) the under-compliance incentives are greater than over-compliance incentives (ibid. at 994–7). Craswell (1999: 2200–5) offers an additional critique of the Cooter/Polinsky-Shavell multiplier. He finds that it can exceed the optimal multiplier depending upon the rate at which the probability of punishment increases when one acts more egregiously and the absolute level of expected damages facing a defendant who in fact takes the efficient level of care (e.g., due to the application of causation or other error).

While law and economics theorists generally emphasize the importance of relaxing the assumption of full enforcement via compensatory damages, more recently scholars have directed attention to the second assumption, examining whether compensatory damages are set too low. The case of wrongful death is the classic example, given strict limitations on the recovery of compensatory damages. At early common law, there were no compensatory damages for wrongful death; the victim’s cause of action ceased to exist at death. States enacted wrongful death statutes to provide remedies, but these were measured by the losses of the deceased’s survivors. Today, it is rare for states to provide for hedonic damages or damages for loss of enjoyment of life by the deceased. For this reason, wrongful death damages will systematically under-deter, as they do not encompass the entirety of the harm inflicted (Arlen, 1985: 1121–34). The conventional response (primarily from theorists who have focused on under-detection as the primary justification for punitive damages) is to address any under-estimation directly, by adjusting the amount of compensatory damages (Arlen, 2000: 710–14). Polinsky and Shavell, for example, advocate inclusion of hedonic damages in compensatory damages for wrongful death (Polinsky & Shavell, 1998: 939–41).

By contrast, Geistfeld (2008) and Hersch and Viscusi (2010b) would employ punitive damages in wrongful death cases to counteract systematically insufficient compensatory damages. These scholars highlight the inherent differences between death (and serious bodily injury) and property or economic damage cases. First and foremost, death and serious bodily injury cases involve irreplaceable losses, with the likelihood in the latter cases that compensatory damages will not suffice to return the victim to his pre-injury state. Hersch and Viscusi point out that, given the current method for setting compensatory damages, the Polinsky-Shavell multiplier will not establish efficient levels of deterrence for wrongful death and bodily injury cases. Nor are they satisfied with the Polinsky-Shavell proposed solution to increase compensatory damages by providing for hedonic damages, because this will lead to over-insurance, given that victims would not choose ex ante to insure fully for noneconomic losses. Instead, Hersch and Viscusi

---

7 But see Kahan (1989: 437) who finds that the effect of uncertainty will always lead to under-compliance rather than over-compliance, if defendants may avoid liability even when they fail to exercise due care. This is the case under his model, which explicitly takes into account causation (i.e., only those defendants who exercise suboptimal levels of care and cause harm are found liable under a negligence standard).

8 Most states, however, prohibit punitive damages in wrongful death cases (Hersch & Viscusi, 2010b: 246).
propose that punitive damages be linked not to the value of compensatory damages, but
to the value of a statistical life. The Hersch-Viscusi formula is: total damages (compensa-
tory plus punitive damages) = value of statistical life (Hersch & Viscusi, 2010b: 231).9

Scholars have long recognized the dilemma posed by wrongful death (and other non-
economic losses): namely that the amount necessary for optimal deterrence of wrong-
ful conduct will inevitably lead to over-insurance (Arlen, 2000: 696–709; Arlen, 1985:
1135–6). To elaborate:

A conundrum thus emerges: Even though noneconomic damages are properly part of what
defendants should pay (on deterrence grounds), they may exceed the amounts victims should
receive (on insurance grounds). Put differently, even when deterrence requires compensation
for nonpecuniary losses, optimal damages do not provide full recovery for nonpecuniary losses
because of the balance that must be struck between the deterrence and insurance goals of tort
law (Sharkey 2005a: 402–3).

Another way to reconcile this dilemma would be to force defendants to pay punitive
damages in an amount calculated for optimal deterrence but give the money to the state or
some entity other than the victim (Polinsky & Che, 1991: 562; Hersch & Viscusi, 2010b).
Sharkey (2003a: 416–22: 375–80) proposed such a scheme, patterned off the existing
split-recovery schemes in several states, which transfer some portion of punitive damages
awards to either the state or a state-operated fund.

2. Disgorgement of illicit gains: gain elimination
Hylton (1998) has been the most vociferous critic of the loss internalization justification
for punitive damages. He proposes a competing economic model based upon disgorgement
of ill-gotten gains, or gain elimination. Deterrence-focused scholars have adverted
to this alternative economic goal, with the caveat that such gains should be removed only
if they are deemed socially “illicit” (Cooter, 1982: 86–9; Polinsky & Shavell, 1998: 908–10).
But, according to Hylton, gain elimination is the primary goal of punitive damages,
because complete deterrence (i.e., stopping the wrongful conduct altogether), rather than
some form of optimal deterrence, is the goal. In other words, where punitive damages are
used for disgorgement purposes (to teach that “torts do not pay”) rather than for incentive
purposes (to teach that “precautions pay”), society has implicitly chosen a goal of abso-
lute deterrence rather than relative deterrence in tort. Here, the realm of “efficient torts”
is correspondingly limited; the focus instead is on eliminating the conduct altogether.

Hylton’s theory harkens back to Becker (1968), who embraced the loss-internalization model in the criminal law context, which, he argued, could deter unwanted conduct as well as the gain elimination model, at least when losses exceed gains, and was also more efficient to administer on account of the costs associated with calculating a defendant’s gain. Hylton contends that, in some circumstances, gains are not difficult to measure (Hylton, 1998: 432–3). Moreover, he worries that the costs the defendant internalizes may not reflect the true societal costs associated with his activities, which likely inflict “deriva-
tive and consequential harms” upon those impacted by the victim’s loss and “secondary

9 This formula is generalized to situations in which the probability of detection is less than 1
harms” upon “unassignable third parties,” who forego beneficial opportunities and take unjustified and excessive precautionary measures in response (ibid. at 434–9). Thus, according to Hylton, the additional costs associated with the gain elimination model—even if difficult to measure—are in many instances outweighed by the societal costs associated with the insufficient deterrence due to incomplete loss internalization, especially where an activity should be completely deterred (ibid. at 436–9).

3. Prevent wrongful takings/ induce voluntary transfers

The gain elimination theory resonates with the “property rule model” set forth by Haddock, McChesney & Spiegel (1990). Like Hylton, Haddock et al. (1990) stress that even if the probability of being held liable were 100 percent, compensatory damages alone would not suffice to deter certain acts. Haddock et al.’s model targets conduct that disregards property rights. Under the Calabresi-Melamed (1972) conception of property rules versus liability rules, whereas liability rules enable an actor to “pay as he goes”—breach a duty of reasonable care and simply pay compensatory damages for the harm caused—property rules force actors to bargain with entitlement holders. Economic theory suggests that, where transaction costs are low, property rule protection is favored, as this ensures voluntary transfers that better protect subjective valuations.10

Haddock et al.’s property model suggests that the award of punitive damages bars the tortfeasor from converting property rule protection into liability rule protection. In other words, for certain kinds of acts, an actor will not be inclined to “take” from a putative victim if the actor will face not only compensatory damages (i.e., liability rule protection) but also punitive damages (i.e., property rule protection). Allowing individuals to “take” under a liability regime leads to the following negative outcomes: (1) it “diminishes [the possessor’s] incentive to invest scarce or unique resources in tradeable assets”; (2) “substitute investments less threatened by opportunistic takings will be enlarged, even where such investments are socially less valuable at the margin than the foregone alternatives”; and (3) encouragement of defensive spending by the present possessor and excessive (and inefficient) offensive spending by the individual who seeks to take from the possessor (Haddock et al., 1990: 16–17). Alternatively, faced with the prospect of punitive damages, the actor will be induced to seek voluntary transfer of the entitlement. The primary goal of punitive damages is to induce the defendant to utilize the market in procuring goods. Punitive damages should leave the defendant no better off than if he had used the market, thereby removing the incentives for him to circumvent the market and engage in extra-market activity.

Consistent with this model, punitive damages are favored in situations where there is no real market: bodily injuries, injuries to character and personality (e.g., defamation, libel, slander), trespass to land, and malicious interference with business relations (Haddock et al., 1990: 26).

This model mirrors a conception of deterrence that often animates penal and civil

---

10 In an amicus brief submitted for litigation, Hylton (2003: 18–22) explicitly adopts this rationale, arguing that if the transaction cost of using the market is lower than the cost of enforcing the law against an offender, the defendant’s gain should be stripped. Hylton invokes the Calabresi-Melamed framework, arguing that when transactions costs are low, punitive damages impose essentially a “property rule” upon the victim’s property interest (ibid. at 12 n. 16).
penalties. That is, the purpose of a penalty (e.g., a speeding ticket, a fine from a safety inspector) is to communicate a norm of conduct to a population of actors. The scale of the penalty may depend on many factors, including some empirical sense of what penalty suffices to induce compliance, but there is no explicit admission by the penalty-setting agency that the penalties have a direct economic relationship with the social cost of the activity they are designed to deter. In other words, even at the theoretical level, there is a disconnect between the size of punitive damages and the economic cost of the activity to be deterred. This approach provides an alternative view of optimal deterrence—a rival theory to loss internalization—that is based on a public-law model of punitive damages.

Distinguishing between alternative economic rationales for punitive damages is not simply of theoretical interest. The underlying rationales can yield very different results in individual cases. Most prominently, law and economics theorists offered diametrically opposing views on whether the Exxon Valdez oil spill warranted the large ($5 billion) punitive damages award. Polinsky and Shavell argued that punitive damages were not justified on under-detection grounds, given that the oil spill was open and notorious (Polinsky & Shavell, 1998: 903–4). Hylton, by contrast, contended that punitive damages were justified on disgorgement grounds (Hylton, 1998: 452–4). The difference in approach rests primarily on the “characterization of the offender’s conduct” (Hylton, 2003: 8). On one view, the Exxon Valdez incident arose out of socially beneficial activity—piloting supertankers—and thus the optimal deterrence/ internalization of losses model pertains. But, on the opposing view, Exxon’s conduct—namely, countenancing the operation of a large tanker by a lapsed alcoholic captain—is the kind of activity that society seeks to deter completely (Hylton, 1998: 452–4).

B. Doctrine

To what extent have the aforementioned theoretical insights and analyses made inroads into punitive damages doctrine? The short answer is: not much. The U.S. Supreme Court has taken a keen interest in punitive damages, with a string of cases beginning in 1996 with BMW of North America, Inc. v. Gore, 517 U.S. 559 (1996), that have created an edifice of federal constitutional review of punitive damages awards, setting forth a template for restrictions on punitive damages. In each of its cases, the Court has reiterated the twin purposes of punitive damages: to punish and to deter. But, the Court has never really specified what it means by deterrence and has intimated that punishment is the predominant purpose, with deterrence perhaps an incidental effect.

In a trilogy of cases—BMW v. Gore, State Farm v. Campbell, 538 U.S. 408 (2003), and Philip Morris USA v. Williams, 549 U.S. 346 (2007)—the Court set forth and reinforced its framework for appellate review of punitive damages awards, identifying three principal guideposts: (1) reprehensibility of conduct; (2) the ratio of punitive to compensatory damages; and (3) alternative comparable sanctions (such as statutory penalties). The

---

11 Originally—and more precisely—the Court’s ratio was between punitive damages and the harm caused, with compensatory damages merely serving as a proxy for the said harm. This formulation leaves room for arguments that compensatory damages do not accurately reflect the harm in a particular case and so should be adjusted prior to computing the relevant ratio.

12 These cases are constitutional due process punitive damages cases. The Court’s guideposts
Economic analysis of punitive damages

Court has emphasized that reprehensibility is the most important factor in reviewing punitive awards, but it has repeatedly stressed the punitive-compensatory ratio, which lower courts have taken to be a “rule of thumb” in favor of single digit ratios (Sharkey, 2009: 28).

From an economic perspective, however, neither reprehensibility nor the punitive-compensatory ratio is related to economic deterrence. Reprehensibility connotes noneconomic justifications like retributive punishment. And economic deterrence is concerned with total damages, not the relationship between punitive and compensatory damages.

The U.S. Supreme Court, amidst its forays—or incursions—into the state law realm of punitive damages, has always insisted that its analysis begins with the legitimate state interests served by punitive damages. To date, the Court has given a quick nod to state-defined goals of punishment and deterrence, before diving headlong into an analysis that, by and large, has emphasized retributive punishment goals. But the states have, in essence, enabled the Court to go down this path. They have, as yet, not exploited their power to redirect the Court—and thus the evolution of punitive damages doctrine.13

1. U.S. Supreme Court

The U.S. Supreme Court has not warmly embraced economic rationales for punitive damages. In Cooper Industries, Inc. v. Leatherman Tool Group, Inc.—a case in which the Court rejected the argument that punitive damages were “facts” found by juries and courts and thus subject to the Seventh Amendment prohibition upon redetermination of facts during appellate review—the Court distanced itself from an optimal deterrence view of punitive damages:

However attractive such an [optimal deterrence] approach to punitive damages might be as an abstract policy matter, it is clear that juries do not normally engage in such a finely tuned exercise of deterrence calibration when awarding punitive damages. . . . After all, deterrence is not the only purpose served by punitive damages. . . . Moreover, it is not at all obvious that even the deterrent function of punitive damages can be served only by economically “optimal deterrence.” “[C]itizens and legislators may rightly insist that they are willing to tolerate some loss in economic efficiency in order to deter what they consider morally offensive conduct, albeit cost-beneficial morally offensive conduct; efficiency is just one consideration among many” (532 U.S. 424, 439–40 (2001) (citing Galanter & Luban, 1993: 1450).

The Court reasserted this attitude toward economic deterrence in State Farm v. Campbell when it gave short shrift to the lower courts’ findings with respect to the link are meant to set the point at which a punitive award to a plaintiff violates the defendant’s right not to be deprived of property without due process of law, as guaranteed by the Fourteenth Amendment to the U.S. Constitution.

13 State legislatures have even more tools at their disposal than state courts. For example, they could establish statutory damages schemes that track the likelihood of under-detection. And they could clearly articulate non-retributive goals, such as effectuating broader cost internalization for widespread, possibly diffuse harm (Sharkey, 2010: 472–6). State legislatures could likewise articulate economic goals when enacting split-recovery schemes (which direct a portion of punitive damages to the state or state-directed fund). In so doing, these legislatures have the opportunity to recast the underlying purpose of punitive damages as economic deterrence, setting up, for example, designated funds directed toward remediating the type of widespread harm involved (ibid. at 476–7).
between likelihood of under-detection and size of punitive damages. In reinstating a $145 million punitive award (on top of a $1 million compensatory award) in a Utah case alleging bad faith failure to settle against an insurance company, the Utah Supreme Court relied (in part) upon the trial court’s finding that “State Farm’s actions, because of their clandestine nature, will be punished at most in one out of every 50,000 cases as a matter of statistical probability.” One concern might be with respect to the evidentiary basis necessary to support a finding on likelihood of detection in any individual case. Indeed, Polinsky and Shavell argued in an amicus brief to the Supreme Court that the $145 million punitive award constituted an excessive punitive “multiplier” that was not specifically tailored to the conduct that led to the suit, but instead included a wide range of unrelated misconduct, which would result in overdeterrence (Polinsky & Shavell, 2003: 3–4).

But the U.S. Supreme Court’s critique of this under-detection rationale went beyond evidentiary concerns: “[T]he argument that State Farm will be punished in only the rare case . . . had little to do with the actual harm sustained by the Campbells” (State Farm v. Campbell, 538 U.S. 408, 427 (2003)). Moreover, the Court worried that such an argument would justify “a departure from well-established constraints on punitive damages” (ibid.). Nor was the Court persuaded by Hylton’s gain-elimination or market-circumvention rationales. In an amicus brief, Hylton characterized State Farm’s policies as circumvention of the market through “predatory” and “inherently fraudulent” means (Hylton, 2003: 24, 26). As such, Hylton argued, the conduct should be completely deterred, with no risk of over-deterrence of socially beneficial conduct; moreover, given that the conduct was “extremely profitable,” the under-deterrence concern was significant (ibid. at 28).

Despite its criticism of the optimal deterrence rationale and its silence with respect to the alternative gain-elimination or market-circumvention rationales, it would be wrong to conclude that the U.S. Supreme Court has closed the door entirely on economic rationales of punitive damages. A more nuanced view would be that the Court recognizes optimal deterrence as one, but not the sole, underlying justification for punitive damages. Ever since its first articulation in BMW v. Gore, the Court has reiterated that a higher punitive-compensatory ratio might be warranted where “the injury is hard to detect” (BMW v. Gore, 517 U.S. 559, 582 (1996)). And, in Exxon Shipping Co. v. Baker—the Court’s fullest exposition of the policy goals underlying punitive damages—the Court elaborated:

> [H]eavier punitive awards have been thought to be justifiable when wrongdoing is hard to detect (increasing chances of getting away with it) . . . or when the value of injury and the corresponding compensatory award are small (providing low incentives to sue). . . . That concern has no

---

14 A separate line of inquiry is whether the Court’s decisions (regardless of what underlying justifications it embraces) have any effect on outcomes in future cases. Compare Eisenberg & Heise (2011: 23–8) (finding that for both judges and juries punitive-compensatory ratios have increased, the opposite of any expected State Farm effect) and Eisenberg & Wells (1999: 75–80, 83) (concluding that BMW v. Gore had little to no effect on punitive damage levels or the punitive-compensatory ratio in published opinions), with Del Rossi & Viscusi (2010, 138–46 tbls. 6,7,8; 151–3 tbl. 11) (based upon a sample of blockbuster punitive damages awards, reporting a drop in the frequency of punitive damages awards, the level of punitive damages awards, the effect of compensatory damages on punitive awards, and the punitive-compensatory ratio after State Farm).

traction here, in this case of staggering damage inevitably provoking governmental enforcers to indict and any number of private parties to sue (554 U.S. 471, 494, 511 (2008)).

By rejecting the under-detection rationale in the case at hand, the Court both affirmed the relevance of the economic justification and implicitly sided with the loss-internalization model over that of the gain-elimination model, by distinguishing the Exxon Valdez situation from those involving “odds of detection that have opened the door to higher awards” (ibid. at 513).

That said, the Court has chosen to emphasize the retributive punishment rationale over any economic rationale which, in the final analysis, inevitably circumscribes the degree to which the Court will be willing to incorporate economic analysis of punitive damages into doctrine (Sharkey, 2009: 26–30, 50–3). Moreover, the Court’s repeated invocation of the punitive-compensatory ratio as a metric of the appropriateness of the size of the punitive award bespeaks a commitment to a goal other than economic deterrence. In BMW v. Gore, the Court stated that, “perhaps [the] most commonly cited indicium of an unreasonable or excessive punitive damages award is its ratio to the actual harm inflicted on the plaintiff” (517 U.S. 559, 580 (1996)). With State Farm v. Campbell, the Court’s single-digit ratio rule of thumb was ensconced: “Our jurisprudence and the principles it has now established demonstrate . . . that, in practice, few awards exceeding a single-digit ratio between punitive and compensatory damages, to a significant degree, will satisfy due process” (State Farm, 538 U.S. at 425). By Exxon Shipping, the ratio guidepost was firmly entrenched: “[T]he potential relevance of the ratio between compensatory and punitive damages is indisputable, being a central feature in our due process analysis” (Exxon Shipping Co., 554 U.S. at 507).

The relevance of wealth (or net worth) of the defendant—an issue that looms on the U.S. Supreme Court’s horizon—figures in the debate surrounding economic justifications for punitive damages. Economists are divided on the question whether the wealth of the defendant is relevant to economic deterrence (apart from the limited case of the judgment-proof tortfeasor, when all agree it is relevant). Cooter (1989: 1176–7) argues that consideration of the defendant’s wealth is “inappropriate to deterrence of economically self-interested decisionmakers” because the “controlling factor in a purely self-interested calculus . . . is the cost of compliance relative to the cost of liability.” Polinsky and Shavell (1998: 910–14) concur, asserting that wealth should never be taken into account for corporations as long as they can correctly balance the “costs of precaution against the resulting reduction in harm.” By contrast, Arlen (1992: 428) argues that to the extent society wants to maximize social utility in the presence of risk-aversion, then negligence regimes must take into account wealth differences in establishing the duty of care and further argues that both strict liability and negligence regimes should consider differences in wealth in setting the level of compensatory damages. Hylton (2008: 930) also suggests that, “the wealth of the defendant is relevant in the determination of a punitive award when either the victim’s loss or the defendant’s gain from wrongdoing is unobservable and correlated with the defendant’s wealth.”

Many states in fact allow the jury to consider the wealth of the defendant. The U.S. Supreme Court has yet to decide the issue; to date, its pronouncements on the matter have been rather cryptic, cautioning that, “[t]he wealth of a defendant cannot justify an otherwise unconstitutional punitive damages award” (State Farm, 538 U.S. at 427). Speaking
more candidly in *BMW v. Gore*, Justice Breyer expressed concern that consideration of the wealth of the defendant “provides an open-ended basis for inflating awards when the defendant is wealthy” (*Gore*, 517 U.S. at 591 (Breyer, J., concurring)). The majority in *State Farm v. Campbell* indicated similar reservations about the use of evidence of defendant’s wealth: “[R]eference to its assets (which, of course, are what other insured parties in Utah and other States must rely upon for payment of claims) had little to do with the actual harm sustained by the Campbells” (*State Farm*, 538 U.S. at 427).

2. Lower federal and state courts

Of course, the U.S. Supreme Court is only one (albeit influential) actor; lower federal and state courts (as well as legislatures) are key players given that punitive damages fall within the purview of state common law.

The loss-internalization model received an endorsement in a now-infamous case involving two siblings’ stay at a bed bug-infested hotel in Chicago. In *Mathias v. Accor Economy Lodging, Inc.*, federal appellate judge Richard Posner justified a nearly 40:1 punitive to compensatory ratio by invoking the under-detection rationale for punitive damages:  

The award of punitive damages in this case . . . serves the . . . purpose of limiting the defendant’s ability to profit from its fraud by escaping detection and (private) prosecution. If a tortfeasor is “caught” only half the time he commits torts, then when he is caught he should be punished twice as heavily in order to make up for the times he gets away (*Mathias v. Accor Economy Lodging, Inc.*, 347 F.3d 672, 677 (7th Cir. 2003)).

A prominent state court case embraces the property-rule model approach for cases of intentional trespass. In *Jacque v. Steenberg Homes, Inc.*, 563 N.W.2d 154 (Wis. 1997), the Wisconsin Supreme Court upheld a punitive damages award of $100,000 in a trespass to land case, in which the plaintiff was awarded only nominal compensatory damages. Defendant Steenberg Homes intentionally crossed plaintiff Jacque’s property (in order to deliver its product to a purchaser), despite specific exhortations to keep off Jacque’s land. The court recognized that liability rule protection—here, requiring the defendant to pay only a “halfpenny” in damages—was inadequate to stave off intentional trespasses of land. Punitive damages were necessary, in the court’s view, to protect the landowner’s legal interest in private property, prevent self-help, and ensure the long-term viability of the legal system. The court also implicitly endorsed Hylton’s gain-elimination model, by suggesting that, “punitive damages must be in excess of the profit created by the misconduct so that the defendant recognizes a loss” (ibid. at 630).

16 The precise amount of punitive damages awarded by jury has an ad hoc quality to it. Judge Posner remarked that “[i]t is probably not a coincidence that $5,000 [in compensatory damages] + $186,000 [in punitive damages] = $191,000/191 = $1,000, i.e., $1,000 per room in the hotel” (*Mathias*, 347 F.3d at 678). Sharkey (2005b: 445–6) suggests that the jury may have awarded damages sufficient to measure not only harm to the individual plaintiffs in the case, but also to the other harmed individuals (guests in a 191-room bug-infested hotel).

17 Hylton (1998: 445–6) endorses the result in *Jacque*, which he argues is consonant with his gain-elimination model—but contrary to the Polinsky/Shavell loss-internalization model—because, while there was no damage to the plaintiff’s actual property, secondary costs would be substantial and society’s losses would be difficult to determine.
3. EMPIRICAL ANALYSIS OF PUNITIVE DAMAGES

A. Existing Studies

Over the last decade, the empirical debate surrounding punitive damages has overtaken the theoretical debate within the academy. While the key economic question remains the extent to which punitive damages serve a deterrent function, empiricists have yet to provide a compelling answer. Instead, their efforts have focused primarily on subsidiary questions of the frequency, size, and predictability of punitive awards. Two opposing camps have emerged: Kip Viscusi (and collaborators) has taken aim at the unpredictable nature of punitive damages, focusing on the problem posed by “blockbuster” awards (of at least $100 million) of punitive damages given by jurors that seem unrelated to any rational explanatory factor; Ted Eisenberg (and collaborators) instead argues that punitive damages are rationally related to compensatory damages and have been relatively consistent and stable over time.

The empirical debate is intermittently fueled by the release of subsequent iterations of the Civil Justice Survey of State Courts (a project of the National Center for State Courts and the Bureau of Justice Statistics). The Civil Justice Surveys include data on tort, contract, and property cases collected directly from state court clerks’ offices. The first tranche included cases resolved by trial in fiscal year 1991–92; subsequent datasets have covered calendar years 1996, 2001, and 2005. The Civil Justice Surveys comprise the most representative and most comprehensive sample of state court trials in the United States.18

The first dataset (1991–92) includes only jury trials, but the subsequent datasets include both jury and bench trials, enabling researchers to probe (albeit in a limited fashion, given selection effects) differences between judges’ and juries’ award of punitive damages.

While the Civil Justice Survey data are undeniably the most comprehensive state court sample, Hersch and Viscusi (2004; 2010a) have noted significant limitations that affect their usefulness for assessing punitive damages. While these datasets have been richly mined and analyzed, given disagreements about appropriate model specifications, assumptions and statistical techniques, no consensus has emerged on questions of the significance and predictability of punitive damages.

Moreover, due to the difficulties associated with separating out certain variables or the availability of relevant data, empiricists have encountered difficulties in studying the “shadow” effect of punitive damages in tried cases on settlements. Another area that remains under-explored is the “substitution effect” of punitive damages caps on punitive damages awards; namely, questions of whether, and to what degree, a fact-finder will mitigate the effects of such caps by increasing the compensatory damages to be awarded.

Finally, a different genre of empirical work, based not upon real world data, but upon experiments conducted with mock jurors and juries, has tried to shed light on how jurors and juries actually make decisions regarding punitive damages. An advantage of this

---

18 The surveys cover state courts in a random sample of 46 of the 75 most populous counties in the United States. For a summary of the data and methodology, see Bureau of Justice Statistics Bulletin: Civil Justice Survey of State Courts, 2005: Civil Bench and Jury Trials in State Courts, 2005 (Oct. 2008).
kind of experiment is that it can be carefully controlled and manipulated to isolate the effects of a particular factor of interest. But such experiments are open to critiques that they cannot approximate real world decisionmaking, given the complexity and nuance involved with actual trials, as contrasted with the more streamlined vignettes or case summaries given to mock jury participants.

1. Frequency and size
   Notwithstanding the high decibel of disagreement, there are some agreed-upon empirical facts concerning punitive damages. First, punitive damages are awarded infrequently. An oft-cited RAND study finds that punitive damages were awarded during the early 1960s in only 1/10 of 1 percent of all civil jury trials in Cook County, Illinois, and even fewer in San Francisco, California; for the early 1980s the corresponding figures are 2.5 percent and 8.3 percent, respectively (Peterson et al., 1987: 9 tbl.2.1). The findings from such earlier studies are confirmed by analyses of the Civil Justice Survey datasets, which reveal that punitive damages are awarded in less than 5 percent of tried cases (Eisenberg & Heise, 2011: 6). Certain case types, however—such as fraud, slander/libel, intentional torts, employment discrimination, and products liability—exhibit much higher rates of issuance of punitive damages (Sharkey, 2003a: 351 & n.12). Moreover, where it was possible to limit the sample to jury-tried cases with a plaintiff winner in which punitive damages were sought (a variable included only in the 2005 Civil Justice Survey data), the rate of punitive damages awarded rose to 29 percent (Eisenberg et al., 2010a).

   The size of punitive damages awards—or the issue of whether size should be a cause for alarm—is more controversial. Eisenberg and Heise (2011) report, based upon the Civil Justice Survey datasets, that fewer than 14 percent of the cases in which punitive damages were awarded involved punitive awards larger than $1 million. Indeed, more than half of the awards were less than $100,000. But Viscusi and collaborators counter that it does not matter that the size of punitive damages is of little concern in the average, mine-run case. What is key, according to them, are the outlier, blockbuster awards, which they have compiled into a dataset, consisting of all known punitive judgments greater than $100 million awarded between 1985 and 2003, that they argue is the relevant sample (Viscusi, 2004; Hersch & Viscusi, 2004; Del Rossi & Viscusi, 2010).

2. Predictability: Relationship to compensatory damages
   Predictability relates to the deterrent effect of punitive damages and to fairness. The most scrutinized relationship has been the one between the size of punitive awards and corresponding compensatory awards, or the punitive-compensatory ratio. This ratio, however, does not align with the theoretical justification for the Cooter/Polinsky-Shavell multiplier, for it is not related to the probability of detection or any other under-enforcement factor. Nor, under the gain-elimination or property rights theories—where willful and wanton, or at least intentional wrongdoing is the typical target—is there any necessary link between the relative sizes of the punitive and compensatory awards. In sum, the ratio guidepost is “theoretically bankrupt” (Sharkey, 2009: 28).

   Using the Civil Justice Survey datasets, Eisenberg and collaborators provide evidence suggesting that the size of punitive awards is strongly correlated with the size of compensatory damages; moreover, this relationship has been relatively stable over time (from 1991–2005, the lifespan of the datasets) (Eisenberg & Heise, 2011: 20 tbl.4 (2005 data);

Following the contours of their debate regarding the size of punitive damages, Viscusi and Hersch counter that this stable relationship disappears in their dataset of blockbuster awards. Hersch & Viscusi (2004: 9–10) find that the levels of blockbuster punitive damages are not significantly related to compensatory damages, but after taking logs there is a stronger, though still not statistically significant, relationship because taking logs diminishes the influence of outliers. They argue that, from the standpoint of predictability, the real world operates in levels, not logs (Hersch & Viscusi, 2004: 10; Viscusi, 1998b: 386; Del Rossi & Viscusi, 2010).

Hersch & Viscusi (2004) also present a discussion (but not multivariate estimates) of the prominence of cigarette and oil/chemical company blockbuster awards. Del Rossi & Viscusi (2010) show a variety of such industry group effects in a multivariate regression using a larger sample of blockbuster awards, with the cigarette industry, energy-chemical industry, and finance-investment-insurance industry being prime players. Hersch & Viscusi (2010a) also highlight some of the systematic differences in punitive damages awards by case type. Regression analyses of the levels of punitive damages are not pure noise, but they argue there is no evidence that blockbuster punitive damages awards have as sound a basis as Eisenberg et al. claim.

3. Deterrent effect

Do punitive damages in fact deter? To date, the question has eluded empiricists. This is perhaps understandable, given the inherent difficulties of measurement and empirical testing. In terms of specifying and measuring relevant dependent and independent variables, empiricists must rely on imperfect proxies for tortious conduct: lawsuits or statistics regarding accidents (which may or may not be the result of tortious conduct) (Cardi et al., 2011: 4–5). To isolate a discernible effect of punitive damages, one must also be able to control for the independent deterrent effects exerted by market forces and regulation (Viscusi, 1998a: 335). Eisenberg (1998: 353–6) points out that one must also consider the effects of nationwide regulation and the capacity of tort liability to influence neighbor states and the conduct of nationwide corporations. Schwartz (1994: 382–3) suggests that the existence of insurance could obscure the deterrent effect of tort liability. Moreover, in order to detect effects, one must ensure that there are sufficient observations to allow one to effectively draw a conclusion.

Ideally (but as a practical matter, impossible), one would test the deterrent effect of punitive damages by comparing two identical jurisdictions—with the same political climates, populations, industry mixes, etc.—that differ only in whether or not they allow punitive damages (Viscusi, 1998a: 339–40). An alternative (more feasible) test is to analyze a single jurisdiction over time or several jurisdictions with different levels of tort liability.

---

19 Several other studies—using disparate datasets across different types of cases—have likewise revealed that the size of a punitive award is strongly correlated with the size of the compensatory award (Sharkey, 2006: 29 tbl. 2).

20 Eisenberg and Wells, however, challenge this finding. Using a different methodology (and accounting for an extreme outlier), they find a statistically significant relationship between punitive and compensatory awards in the Hersch-Viscusi dataset (Eisenberg & Wells, 2006).
at the same time. Such analyses, however, encounter substantial problems of endogeneity and identification (Mello & Brennan, 2002: 1614), and are typically plagued with additional statistical problems, such as spurious correlation or multicollinearity (Brennan et al., 1999: 206 nn.1,3). These statistical challenges are apt to be particularly formidable in analyses of punitive damages, where greater levels of tortious conduct likely lead to greater imposition of punitive damages.

Viscusi (1998a; 1998b) did offer limited evidence of a negligible deterrent effect of punitive damages, by measuring differences in rates of environmental and safety torts in states with and without punitive damages. Viscusi focused on a dozen risk measures that should have been (but in fact were not) affected by punitive damages, including insurance rates, toxic releases, environmental accidents, and medical malpractice deaths. The sample was “relatively small” in that there are only 50 states, providing one observation per state for each risk outcome, but the lack of an effect of punitive damages was evident across all these groups even though other variables often did matter. Eisenberg (1998: 348–53), however, argues that Viscusi’s relatively small sample size could not pick up any statistically significant effect, questions Viscusi’s methodology, and also challenges Viscusi’s choice of environmental and safety torts as well as his coding of state punitive damages laws.

4. Judge versus jury
Another area of significant disagreement among empirical scholars is whether judges and juries systematically differ in their determination of the award of punitive damages. Many criticisms of punitive damages are, at their core, criticisms of juror-awarded punitive damages, or juror decisionmaking in civil cases more generally. Eisenberg et al. (2002; 2006) find no substantial evidence suggesting that jurors and judges differ in the rate of awarding punitive damages or in the relationship between the amounts of punitive and compensatory damages.

By contrast, analyzing the 1996 Civil Justice Survey data (the same data used by Eisenberg et al., 2002), Hersch and Viscusi (2004) find that “juries are significantly more likely to award punitive damages than are judges and award higher levels of punitive damages” (Hersch & Viscusi, 2004: 1). Moreover, in the dataset of 63 “blockbuster” punitive awards of more than $100 million, 95 percent were awarded by juries.

---

21 Given Viscusi’s use of “count-based data,” Eisenberg argues that a Poisson or negative binomial regression should have been used, rather than weighted least squares regression (ibid. at 350 n.21).
22 Eisenberg et al. (2002: 746, 749 tbl.1) report overall percentage of plaintiff-win cases in which punitive damages are awarded as 4.03 percent (jury) and 4.00 percent (judge). Eisenberg et al. (2006: 269 tbl.1) report 4.85 percent (jury) and 3.93 percent (judge).
23 Hersch and Viscusi (2004: 14) find the mean of the punitive-compensatory ratio for juries is 4.6, compared to 1.25 for judges; moreover, the standard deviation for juries is eight times that of judges. They also find that, controlling for other variables, juries have a .013 greater probability of awarding punitive damages than do judges.

In their latest round of analyses, Eisenberg et al. (2010a: 604) uncovered a modicum of judge-jury difference in the decision to award punitive damages. The difference was based on whether cases involved personal injury: judges awarded punitive damages at a higher rate in personal injury cases, whereas juries awarded them at a higher rate in non-personal injury cases. Eisenberg & Heise (2011) also detected a difference in the punitive-compensatory ratios in judge versus jury trials: in the 2005 iteration of the Civil Justice Survey data (but not in previous years), juries awarded higher punitive-compensatory ratios than judges. In other words, juries awarded higher punitive damages per unit of compensatory damages than did judges.

Selection effects seriously impugn empirical studies of judge-jury differences in awarding punitive damages. It is not as if cases are assigned randomly to judges and to juries; instead, parties direct their cases to one decisionmaker or another. It could be, then, that the same underlying factors or attributes of cases that make it more likely for cases to end up before juries would also account for higher punitive damages awards (and vice versa with respect to judges). In other words, even if similar cases would be treated similarly by judges and juries, they are in fact adjudicating different cases, which might best explain the disparate results.

5. Shadow effect
None of the aforementioned empirical studies can get a definitive grasp on what has been called the “shadow effect” of punitive damages—namely the effect of the threat of punitive damages on settlement of cases, both in terms of rate and amount. Polinsky (1997: 677) argues that “punitive damages may be a significant factor in litigation despite the fact that only a small fraction of cases . . . involve punitive damage judgments,” in large part because of their effect on settlement dynamics. Koenig (1998: 193) reports that, in a sample of insurance adjuster data, 11 percent of total settlements were allocated to punitive damages. Eaton et al. (2005: 365), analyzing data from six Georgia counties, conclude that, “the decision to seek punitive damages has no statistically significant impact” on “whether a case that was disposed was done so by trial or by some other procedure, including settlement,” or “whether a case that was disposed by means other than a trial was more likely to have been settled.”

However, there is no comprehensive dataset available to study these effects. Many settlements are confidential and sealed. Even with respect to those that are publicly available, one would not expect that punitive damages amounts would be explicitly designated in the settlement. Rather, both parties have an incentive to reclassify any putative punitive amount as compensatory: plaintiffs are not taxed on settlement amounts designated as compensatory; and defendants avoid the stigma associated with punitive damages and do not risk losing their insurance coverage (at least in states that prohibit insurance for punitive damages).

“sensitivity to the treatment of Harris County [a jurisdiction with a large number punitive damages awarded by judges, relative to juries]” and ii) an “inclusion of two highly correlated jury variables.”

24 Internal Revenue Code § 104.
25 For a detailed treatment of the insurability of punitive damages, see Sharkey (2005b).
6. **Substitution effect of caps**

A little-studied phenomenon may have an underappreciated effect upon proposed
document a substitution effect: namely that one effect of limiting punitive damages is an
increase in compensatory damages. This substitution effect could be the result of plain-
tiffs’ attorneys expending greater effort to increase the noneconomic damages component
of a case in jurisdictions that cap punitive damages, or it could arise from jurors’ taking a
more holistic view of damages, treating compensatory and punitive damages as essentially
fungible and focusing more on the total amount awarded. This substitution effect has yet
to be replicated in other datasets, although there is anecdotal (and some experimental)
evidence to support it (Sharkey, 2008: 86–94).

7. **Mock juror studies**

Mock juror studies provide a window into the jury decisionmaking process that is other-
wise impenetrable to empirical analysis. Moreover, the experimental framework provides
the only viable way to test issues such as whether like cases are treated alike or whether
judges and juries would handle cases similarly (e.g., by presenting to mock jurors vignettes
of cases decided by real-world judges).

Cass Sunstein et al. (2002) provide a wealth of data collected from hundreds of con-
trolled experiments run on over 600 mock juries involving more than 8,000 jury-eligible
citizens as mock jurors. One of the authors’ central findings is that jurors are not inspired
by economic deterrence principles when awarding punitive damages; instead, they are
motivated primarily (and according to the authors, solely) by a sense of moral outrage
at reprehensible conduct (Sunstein et al., 2002: 236–8). In one experiment, mock jurors
were given scenarios that held constant the type of wrongdoing, but varied the likely-
hood of detection. The jurors’ award of punitive damages was impervious to changes in
the detection rate (Sunstein et al., 2002: 135–9; Sunstein et al., 2000: 241–6). In another
experiment, jurors were explicitly instructed to award punitive damages according to the
Polinsky-Shavell multiplier model; they were even given a cheat-sheet table to help them
calculate the relevant figures. Despite this instruction, jurors steadfastly refused to do so
(Sunstein et al., 2002: 151–64; Viscusi, 2001b: 325–37). The authors take this as a defini-
tive rejection of the economic deterrence model by jurors. Another study, using a judge
sample, showed that judges performed better than jurors in terms of risk assessment and
application of cost-benefit analysis (Sunstein et al., 2002: 189–93, 201–4; Viscusi, 1999:
31–3; Viscusi, 2000: 50–3; Viscusi, 2001a: 114–15, 130–2). Based upon these studies, the
authors conclude that, to the extent optimal deterrence is the (or a) goal of punitive
damages, decisionmaking should be taken out of the hands of jurors and placed into the
hands of judges (Sunstein, 2002: 248–52).

Arlen and Talley (2008), drawing upon the field of experimental economics, question
the validity of experimental results of the kind offered by Sunstein et al.’s mock juror

---

26 See Hyman, et al. (2009: 403). Born et al. (2009) find that there is a negative effect of punitive
damages on insurance company losses—which would suggest that, if there is an offset, it does not
seem to be a complete offset once you include settled cases as well as those that go to trial, which
are subject to selection effects reflecting the role of damages caps and other factors.
studies that fail to satisfy the requirements of control, internal consistency, and external validity. Of particular relevance to the Sunstein et al. mock juror studies, internal consistency requires that the payoffs that the subject receives mirrors the payoffs the researcher is seeking to evaluate. This is a problem if a choice has emotional consequences that are experienced in real life (such as when a juror imposes large punitive damages upon a defendant), but not by someone making a hypothetical choice (such as the decision of the mock juror to impose large punitive damages) (Arlen & Talley, 2008: xxxiv). Moreover, external validity is a concern given that the methodologies employed in the mock juror studies, such as providing the participants with succinct summaries of the cases at hand, differ from the real world juror decisionmaking process, whereby jurors make punitive damages decisions only after sitting through days of direct and cross-examination of witnesses, hearing arguments by attorneys, having been instructed in the law by the judge, and after having determined a compensatory damages award (Vidmar, 2004: 6–7, 22–3).

Even in light of these methodological critiques, and even if the authors’ policy prescription is premature based upon their more limited findings (Sharkey, 2003b), Sunstein et al. nonetheless provide a seminal body of empirical work that cautions against too readily assuming that the economic theory of punitive damages translates readily into operation, at least where jurors are involved.

B. U.S. Supreme Court’s Misuse of Empirical Data

For years, researchers submitted their empirical findings on punitive damages to the U.S. Supreme Court via amicus briefs. By and large, the Court seemed to ignore this body of work—or at least, the Court did not find it relevant to the task at hand, namely fashioning a framework for constitutional excessiveness review. But Exxon Shipping Co. v. Baker, 554 U.S. 471 (2008), presented the Court with an opportunity to delve into these findings. Because Exxon Shipping was taken up under the Court’s federal admiralty jurisdiction (resting upon federal common law, not constitutional principles), the Court had license to weigh the full spectrum of public policy considerations with respect to the punitive damages remedy.

The Court took sides in the raging empirical debate (to many observers’ surprise, given that, in this case—unlike previous cases—the parties and amici had not briefed or argued the empirical findings). The Court rejected the view that punitive damages are wild and out of control; to the contrary, looking across data from the mass of cases, the Court found relative moderation and stability in the size of punitive damages (ibid. at 497–9). But something was amiss. The root of the problem of punitive damages, according to the

27 An experimenter must control differences between the control and experimental groups, including “both the variation in the subject pools and . . . the differences in the factors relevant to the subjects’ choices” (Arlen & Talley, 2008: xxxiii).

28 Internal consistency demands that the factors driving the subjects’ decisionmaking must be that which the experimenter is testing (ibid. at xxxiv). Essentially, this requires that i) the experiment and control groups differ in the way hypothesized by the experimenter, and ii) they do not differ in any other meaningful ways.

29 An experimenter should “test[] a choice that matches up with the real-world settings from which the experiment is meant to extrapolate” (ibid. at xxxvii).
Court, is unpredictability—the fact that like cases are not treated alike and that there does not appear to be any good explanation for some outlier punitive awards. In reaching this conclusion, the Court relied upon data from the Civil Justice Surveys and highlighted the standard deviation—a statistical measure of the spread of data—for punitive damages, which, the Court argued, suggested too much variability (ibid. at 499–500).

Having diagnosed the root of the punitive damages problem, the Court proposed a solution. The Court canvassed alternative reforms, from verbal ones such as improved jury instructions to quantitative ones such as caps on punitive damages, set either in absolute dollar terms or else as a punitive-compensatory ratio (ibid. at 503–4, 506–12). The majority of states that have imposed restraints on punitive damages have gone the ratio route, with the most common limit set at 3:1 for punitive-compensatory awards (ibid. at 509–11). The Court settled upon the ratio solution, but instead of following the states’ lead by fashioning a 3:1 benchmark, the Court turned to statistical data—primarily from the Civil Justice Surveys—as its guide. The Court used the median punitive-compensatory ratio (0.65:1) as its benchmark, rounding up to a 1:1 suggested ratio (ibid. at 513).

The Court’s analysis drew equal criticism from both camps in the empirical debate (Eisenberg et al., 2010b; Hersch & Viscusi, 2010a). Hersch and Viscusi (2010a: 266–77) articulate four main criticisms: (1) there is no law and economics basis for selecting the median ratio as the guidepost; (2) there will always be a disparity between the median and the mean (the Court’s index of unpredictability) for distributions that are truncated at zero, as is the punitive-compensatory ratio; (3) the ratios differ by case type perhaps for quite sensible reasons; and (4) the very high ratios tend to be statistical artifacts for which the compensatory damages are very small, driving up the ratio.

As for the Court’s reliance upon the standard deviation of aggregate Civil Justice Survey data, Eisenberg et al. (2010b) demonstrate that this was a crude, not to mention misleading, measure. Eisenberg et al. disaggregate the data, group it according to the size of the compensatory award and then reanalyze the relevant descriptive statistics. They find that the source of the variation in the size of the punitive award—the entirety of the contribution to the high standard deviation—was the very low (less than $10,000) compensatory awards (Eisenberg et al., 2010b: 16). This finding certainly casts the U.S. Supreme Court’s conclusion (based upon the Civil Justice Survey data) in considerable doubt. To begin, low compensatory damage cases are the very ones that the Court has previously suggested warrant higher punitive-compensatory ratios (Gore, 517 U.S. at 582; State Farm, 538 U.S. at 425). Moreover, to the extent that the variability in punitive awards derives solely from the subset of very low compensatory awards, the Court’s more general pronouncements regarding unpredictability and outlier awards are misdirected (Sharkey, 2009: 41–2).

4. CONCLUSION

To date, the theoretical and empirical insights from economic analysis of punitive damages have had only a modest effect upon doctrine. The partisan political rhetoric of tort reform reaches high decibels in any real-world debate about punitive damages. The economic deterrence view of punitive damages does not align squarely with the conventional plaintiff’s or defendant’s preferred position on punitive damages. Consider, for example, the
opposing economic positions in the Exxon Valdez case, or the position advanced by Judge Richard Posner in Mathias v. Accor Economy Lodging, which, to the dismay of the defense bar, used economic reasoning to argue for an increased punitive-compensatory ratio. Perhaps for this reason, there is no natural champion of the economic deterrence view—at least in a tort reform environment where warring sides are in search of stalwart allies. Economic deterrence, in other words, might be too much of a fair-weather friend, supporting high punitive damages awards in some contexts, but quite limited ones in others.

Politics, then, might turn out to be a formidable barrier to the economic perspective on punitive damages. It would be wrong to think, however, that the U.S. Supreme Court has already definitively closed that door. State legislatures and courts certainly have at their disposal—as a theoretical matter, if not a politically realistic one—tools they could wield in terms of crafting supra-compensatory remedies (hitherto known as punitive damages) based expressly on economic deterrence principles.

The academy faces a different set of challenges. The contemporary landscape of punitive damages, with an emphasis on reckless endangerment cases, poses new questions, yet to be taken up in full in the theoretical literature. Should one restrict punitive damages to malicious or willful and wanton conduct? If not, when is it appropriate to award damages in excess of the harm caused in the burgeoning category of recklessness cases? Should theories treat knowing breaches with high-risk injury differently from accidents caused by recklessness?

Moreover, scholars have devoted insufficient attention to the interplay with other mechanisms—both regulatory and market-driven—likewise designed to deter wrongful conduct. Should punitive damages be available, for example, in situations where the tortfeasor has complied with the relevant regulatory standards for conduct? Should negative reputational effects be factored into the analysis? Karpoff and Lott (1999) provide a rough measure of tarnished reputations and adverse publicity by analyzing stock market declines at the time of announcement of suits involving punitive damages. Polinsky and Shavell (2010) highlight tort doctrine’s lack of attention to the deterrent effect of market-driven reputation; an extension of their theory would try to incorporate this directly into the Polinsky-Shavell multiplier model.

These as-yet underexplored areas, combined with the far-from-complete task of better integrating doctrine and jurisprudence, make punitive damages an area of law that will likely continue to attract great scholarly interest for many years to come.

BIBLIOGRAPHY

Articles and Books

Arlen, Jennifer (2000), Tort Damages, in Encyclopedia of Law & Economics, 2, 682 (Boudewijn Bouckaert & Gerrit De Geest eds.).


Eisenberg, Theodore & Martin T. Wells (1999), *The Predictability of Punitive Damages Awards in Published Opinions, the Impact of BMW v. Gore on Punitive Damages Awards, and Forecasting Which Punitive Awards Will Be Reduced*, *Supreme Court Economic Review*, 7, 59–77.


Research handbook on the economics of torts


**Cases and Statutes:**

*Internal Revenue Code § 104 (2006).*
*Jacque v. Steenberg Homes, Inc.*, 563 N.W.2d 154 (Wis. 1997).
*Mathias v. Accor Economy Lodging, Inc.*, 347 F.3d 672 (7th Cir. 2003).
*United States v. Carroll Towing*, 159 F.2d 169 (2d Cir. 1947).

**Briefs and Data:**