Punitive Damages as a Social Harm Measure: Economic Analysis Continues

Sergey Budylin

Available at: https://works.bepress.com/sergey_budylin/1/
PUNITIVE DAMAGES AS A SOCIAL HARM MEASURE: ECONOMIC ANALYSIS CONTINUES

Sergey Budylin*

[R]ewards, in their nature, can only persuade and allure; nothing is compulsory but punishment.\(^1\)

- William Blackstone

Although bedbug bites are not as serious as the bites of some other insects, they are painful and unsightly.\(^2\)

- Richard Posner

I. INTRODUCTION

In a tort case several types of remedies may be available to the plaintiff. Most often, the remedy is compensatory damages designed to compensate the victim for the harm inflicted by the tortfeasor.\(^3\) Another possibility is restitutionary damages, also known as disgorgement, designed to withdraw from the offender the illicitly earned benefit.\(^4\) Although the Restatement of Torts includes restitutionary awards into the notion of “compensatory damages,”\(^5\) thus understanding the latter in a broader sense, in this article I will treat the two concepts as separate. Still another remedy is punitive damages, normally imposed in addition to compensatory damages for especially reprehensible behavior of the

\* Sergey Budylin is a lawyer with Roche & Duffay, Moscow. The author would like to thank Marian Dent for useful discussions; the author is also grateful to A. Mitchell Polinsky, Steven Shavell, and Keith N. Hylton for valuable comments to an earlier version of the article.

1. WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 57 (1765).
4. Id. at cmt. b.
5. Id. at § 903.
offender.\textsuperscript{6} Other types of remedies, like injunction, are not considered here.

Presently, a close relation between law and economics is commonplace.\textsuperscript{7} Although there is apparently no problem in explaining, on economic grounds, either compensatory damages or disgorgement, punitive damages seem rather enigmatic from the economic point of view. Obviously, they are designed to deter the defendant and other potential offenders from repetition of the offensive conduct. However, there seems to be no clear-cut economic way to either define the notion of reprehensibility, or, most importantly, to calculate the optimal amount of punitive damages.

A number of approaches to this problem have been proposed. Perhaps the most notable is the theory offered by A. Mitchell Polinsky and Steven Shavell.\textsuperscript{8} According to their approach, economically optimal damages are equal to what can be called augmented compensation, that is, compensation divided by the probability of the offender's liability.\textsuperscript{9} Another notable theory was offered by Keith N. Hylton.\textsuperscript{10} According to his approach, economically optimal damages in most cases are equal to likewise augmented disgorgement.\textsuperscript{11} Although both are mathematically very convincing, albeit mutually exclusive, the theories are nevertheless far from becoming generally accepted methods of treating the punitive damages problem: the former because of arguably being at odds with the general notion of punishment and the latter because of apparent underestimation of the importance of the victim's losses. Both approaches will be discussed in more detail in a later section.

This article attempts to make a further step in the economic analysis of the punitive damages problem. I argue that the economically optimal charge on a tortfeasor is a sum of two components. The first is augmented compensation, in line with Polinsky and Shavell. The second, where applicable, is augmented disgorgement, partly in line with Hylton, where augmented means that it is inflated by the factor

\textsuperscript{6} Id. at § 908.
\textsuperscript{7} See generally Richard A. Posner, Economic Analysis of Law (7th ed. 2007).
\textsuperscript{9} Id. at 889.
\textsuperscript{11} Id. at 444.
reciprocal to the probability of liability.\textsuperscript{12} As we will see later, respective probabilities may be different for the two components.

I further argue that while the first component is an estimation of primary social losses (inflicted to legally recognizable victims), the second component is the best available estimation of secondary social harms (inflicted to unassignable parties). Both components contribute into the overcompensatory part of the total optimal damages. For the reasons discussed below, I will call the overcompensatory part of the first component "supercompensatory" or "quasi-punitive" damages, and the second component, "truly punitive" damages. Taken together, both types of overcompensatory damages, traditionally indiscriminately referred to as punitive, represent the total amount of the inflicted social harm not attributable to the specific plaintiff.

This interpretation provides a sound economic basis for the proposed damages calculation recipe. Facing a prospect of being charged with the so defined amount, a potential tortfeasor is motivated to maximize the total social utility, including that of himself, of potential victims, and of unassignable parties, society as a whole. Importantly, the rule does not conflict with established moral or legal principles; on the contrary, it relies on them in determining the threshold for punishment imposition.

I further address the issue of economically optimal awards to the plaintiff. Traditionally, the whole amount withdrawn from the offender is passed to the victim. However, there is no obvious economic reason why the two amounts should be equal; perhaps a part of the award should go to the state. This idea is reflected in "split-award" statutes, allocating a portion of punitive damages to a state agency, adopted by a number of states.\textsuperscript{13} In this article I argue that only augmented compensation, that is, compensatory and supercompensatory damages should be awarded to the plaintiff; truly punitive damages should be kept by the state.

Finally, I consider the problem of litigation costs, including both direct expenses and personal effort of the plaintiff. I argue that direct litigation expenses should be shared by the plaintiff and the state in proportion to their share of the award. In addition, the plaintiff should be rewarded by the state for her personal effort; however, this part of the

\textsuperscript{12} Id.

\textsuperscript{13} See BMW of N. Am. Inc. v. Gore, 517 U.S. 559, 616 (1996) (Ginsberg, J., dissenting) (listing split-award statutes; note, however, that the list has somewhat changed since); See also Daughety & Reinganum, infra note 117, at 35 (offering a table of split-award statutes).
optimal award is, arguably, relatively small.

Although the whole approach is essentially mathematical, complex calculations are avoided throughout the article.

II. CURRENT STATE OF THE LAW

In this section the current law of punitive damages is reviewed. Punitive damages are commonly understood as any damages not being nominal or compensatory. Punitive damages are viewed as a "tool to punish [a wrongdoer] for his outrageous conduct and to deter him and others like him from similar conduct in the future."\(^{14}\) For proper imposition of punitive damages the defendant’s conduct must be "outrageous," which includes "[r]ecless indifference to the rights of others and conscious action in deliberate disregard of them."\(^{15}\) Whether punitive damages should be awarded and what amount is appropriate "are within the sound discretion for the trier of fact, whether judge or jury."\(^{16}\) The amount of punitive damages is determined basing upon various factors, including the motives of the wrongdoer, the relations between the parties, the extent of the harm inflicted, and the wealth of the defendant.\(^{17}\) Many states find it worthwhile to restrict applicability of the punitive damages: some states decline to award them in the absence of statutory provisions, some disallow them "unless the defendant’s conduct was aimed at the public generally," and some limit them to the amount of counsel fees.\(^{18}\)

While tort law has traditionally been a state matter, the U.S. Supreme Court has limited states’ ability to articulate their law through the application of constitutional limitations on a plaintiff’s right to recover in certain tort cases. A comprehensive review of the issue may be found in Thomas C. Galligan Jr.’s article.\(^{19}\) Here I will briefly review the most important Supreme Court punitive damages cases, generally in line with Galligan.

---

15. *Id.* at cmt. b.
16. *Id.* at cmt. d.
17. *Id.* at cmt. e.
18. *Id.* at cmt. f.
PUNITIVE DAMAGES

In the famous defamation case of *New York Times Co. v. Sullivan*, the Court struck down an award, apparently including a punitive component, on the grounds of the First Amendment Free Speech Clause, applicable to states through the Fourteenth Amendment Due Process Clause. According to the Court, "actual malice" must be shown for awarding remedy if the plaintiff is a public figure or the subject matter is of public concern. The doctrine was further fine-tuned as applied to private parties in *Gertz v. Robert Welch, Inc.* and *Dun & Bradstreet v. Greenmoss Builders*.

In *Silkwood v. Kerr-McGee Corp.*, a nuclear contamination case, a lower court held that punitive damages claim was preempted by federal safety regulations under the Supremacy Clause. The Court disagreed, finding that Congress did not intend to preempt state law tort remedies except where specifically indicated. Notably, four justices dissented. They argued that while compensatory damages are not preempted, punitive damages are, because of being regulatory in nature.

In *Browning-Ferris Industries of Vermont, Inc. v. Kelco Disposal, Inc.* the Court declined to apply the Eighth Amendment Excessive Fines Clause to limit the amount of punitive damages in cases between private parties. However, in *Pacific Mutual Life Insurance Co. v. Haslip* and in *TXO Production Corp. v. Alliance Resources Corp.*, the Court indicated its willingness to engage in a review of punitive damages awards under the Fourteenth Amendment Due Process Clause. Nevertheless, plaintiffs prevailed in those particular cases. In *Haslip* the Court endorsed the following list of factors Alabama courts take into consideration in determining whether the award was excessive or

21. *Id.* at 279-280.
22. *Id.*.
26. *Id.* at 246 (citing *Silkwood v. Kerr-McGee Corp.*, 667 F.2d 908, 923 (10th Cir. 1981)).
27. *Id.* at 255.
28. *Id.* at 263-64, 274-75, 283.
30. *Id.* at 268.
33. *Id.* at n.6.
inadequate:

(a) whether there is a reasonable relationship between the punitive damages award and the harm likely to result from the defendant’s conduct as well as the harm that actually has occurred; (b) the degree of reprehensibility of the defendant’s conduct, the duration of that conduct, the defendant’s awareness, any concealment, and the existence and frequency of similar past conduct; (c) the profitability to the defendant of the wrongful conduct and the desirability of removing that profit and of having the defendant also sustain a loss; (d) the “financial position” of the defendant; (e) all the costs of litigation; (f) the imposition of criminal sanctions on the defendant for its conduct, these to be taken in mitigation; and (g) the existence of other civil awards against the defendant for the same conduct, these also to be taken in mitigation.34

In Honda Motor Co. v. Oberg35 the Court overturned a punitive damages award in a state law personal injury case on the grounds of a procedural due process right violation, finding no meaningful review was available under state law. In BMW of North America, Inc. v. Gore36 the Court, for the first time, struck down a punitive damages award as offensive to substantive due process. In that case the punitive damages award was five-hundred times greater than the amount of compensatory damages. To arrive to the conclusion, the Court used the following factors, derived from common law: (1) the degree of reprehensibility of the defendant’s conduct; (2) “the disparity between the harm or potential harm suffered . . . and [the] punitive damages award;” and (3) “the difference between this remedy and the civil penalties authorized or imposed in comparable cases.”37 In Cooper Industries, Inc. v. Leatherman Tool Group, Inc.,38 the Court held that a U.S. court of appeals reviewing a punitive damages award for constitutionality should not apply a “clearly erroneous” standard, but rather must review the

34. Haslip, 499 U.S. at 21-22.
37. Id. at 575.
proportionality determination de novo. This apparently indicates that a punitive damages award, although being determined by the trier of fact, is not a factual finding.

Recently, in State Farm Mutual Automobile Insurance Co. v. Campbell, the Court overturned a punitive damages award on substantive due process grounds as being neither reasonable nor proportionate to the wrong committed. Here, in addition to the Gore factors, the Court articulated very specific numerical guidelines for determining the maximum ratio between punitive and compensatory damages. While refusing to impose a bright-line ratio that a punitive damages award cannot exceed, the Court indicated that few awards exceeding a single-digit ratio (that is, nine-to-one as a maximum) will satisfy due process; that the four-to-one ratio, first mentioned in Haslip, may be already close to the constitutional limit; and that in cases where compensatory damages are substantial, a lesser ratio, perhaps one-to-one, “can reach the outermost limit of the due process guarantee.” Here, the ratio was one hundred and forty-five-to-one, and the award was struck down.

The most recent U.S. Supreme Court punitive damages case is Philip Morris USA v. Williams. While the bulk of the article has been written before Williams, last edits are made after it. The decision seems rather controversial. I find it useful to consider the case in more detail separately, at the end of the article. It is also noteworthy that the U.S. Supreme Court granted certiorari in another famous punitive damages case, Exxon Shipping Co. v. Baker, but the case is on an early stage of consideration as of this writing. The Exxon case will also be discussed in some detail later.

The existing jurisprudence may leave an economically-minded analyst with a feeling of frustration. While the common-law factors reiterated in Haslip and Gore look reasonable, they do not provide any absolute-amount estimation for a punitive damages award. The trier of fact determines the award at its discretion, and the award may be further

40. Id. at 437 (quoting Gasperini v. Ctr. For Humanities, Inc., 518 U.S. 415, 459 (1996) (Scalia, J., dissenting)).
42. Id. at 425.
reduced on constitutionality reasons, which are mostly also qualitative. On the other hand, while the *Campbell* ratios are quantitative, they do not seem to be based on any economic, or, in fact, on any rational grounds.

Professor Galligan, in particular, is extremely critical of *Campbell*. He argues that the nine-to-one limit "is not based on pre-existing law, principle, reason, or analogy. Consequently, the guideline seems entirely inconsistent with the traditional model of adjudication." Galligan expresses hope "that the Court will effectively abandon the 9-to-1 ratio in favor of some more principled or reasoned constitutional limitation."

This article does not attempt to explain the esoteric *Campbell* ratios. However, it offers a simple quantitative framework consistent with most of the other jurisprudence, including Haslip and Gore guidelines. It also arguably sheds some new light on the Silkwood preemption dispute, partly supporting the dissenting justices' position. The approach, however, does not seem to be in correspondence with the controversial *Philip Morris* decision.

**III. CURRENT STATE OF THE THEORY**

This section is devoted to the current state of the economic theory of punitive damages. As mentioned above, probably the most notable attempt to explain punitive damages in economic terms has been undertaken by Polinsky and Shavell. According to them, economically optimal damages are compensatory damages divided by the probability of the offender's liability. Since the probability is less than or equal to unity, optimal damages are greater than or equal to compensatory ones. The underlying idea is internalization of social losses (the losses of all victims), that is, charging the offender with those losses total amount, respectively apportioned, rather than with the particular victim's loss amount. The expected charge on the offender is then equal to the expected social harm, so a potential tortfeasor is motivated to maximize

---

45. Galligan, *supra* note 19, at 1256. Professor Galligan goes on to opine that the Court may have somehow derived the questionable ratio from the principles of baseball. *Id.* at 1262.
46. *Id.* at 1262.
48. *Id.* at 889.
the total social utility, including his own utility. Only where the offender’s conduct is socially illicit, his own gain is not taken into consideration and, therefore, should be charged extra. The overcompensatory part of the so calculated optimal damages represents economically optimal punitive damages. Note that this rule means the offense is generally allowed to occur if the benefit gained by the offender is greater than the expected harm inflicted to the victim: such actions are considered socially useful. This approach is traceable to Gary S. Becker and further to Jeremy Bentham.

Although Polinsky and Shavell’s argument is mathematically brilliant, it does not seem absolutely convincing, primarily because it is at odds with the commonsensical notion of punishment. According to this notion, punishment is something imposed for a breach of a socially recognized standard of behavior and roughly proportional to the deviation from the standard. For Polinsky and Shavell, punitive damages have nothing to do with punishment; rather they are a part of morally neutral utility redistribution. Such overcompensatory damages of course influence potential offenders’ behavior; however, they can hardly be interpreted as being imposed for breaching any moral or legal norm. Moreover, in certain cases the algorithm does not seem to provide sufficient deterrence at all (e.g. where a socially important norm is breached, but the inflicted harm happened to be small).

Although courts occasionally cite with approval Polinsky and Shavell’s article, especially for their notion of overdeterrence, that is, charging the offender more than needed to optimally deter, their theory is far from becoming a generally accepted punitive damages calculation.

49. “His” is intended to also incorporate feminine and neutral genders. Apparently, the social harm inflicted by using only one gender is less than the social costs incurred by repeatedly saying “he or she or it.”
50. Polinsky & Shavell, supra note 8, at 879.
51. Id. at 909.
52. Id. at 890.
53. Id. at 877 n.14 (citing Gary S. Becker, Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968)).
54. Id. at 876 n.12 (citing Jeremy Bentham, Principles of Penal Law, in 1 THE WORKS OF JEREMY BENTHAM 365, 401-02 (John Bowring ed., 1962) (1838)).
55. Cf. id. at 890.
56. See, e.g., Parks v. Wells Fargo Home Mortg., Inc., 398 F.3d 937, 943 (7th Cir. 2005); Conder v. Union Planters Bank, N.A., 384 F.3d 397, 401 (7th Cir. 2004); Perez v. Z Frank Oldsmobile, Inc., 223 F.3d 617, 621 (7th Cir. 2000); Norcon, Inc. v. Kotowski, 971 P.2d 158, 180 (Alaska 1999).
rule. The U.S. Supreme Court, in particular, seems skeptical about the theory: "However attractive such an 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."\textsuperscript{57} Moreover, ""[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."\textsuperscript{58}

Hylton offers a substantial modification of Polinsky and Shavell's approach.\textsuperscript{59} According to Hylton, the appropriate guideline for calculating optimal damages is often eliminating offender's gain rather than internalizing victim's loss.\textsuperscript{60} That is, the offender should be charged with the amount of his illicit gain. Similar to Polinsky and Shavell's reasoning, the amount of the actual gain should be divided by the probability of offender's liability.\textsuperscript{61} Since the expected gain is then zero, a potential tortfeasor is sufficiently deterred. However, according to Hylton, this approach is only applicable where the offender's gain is probably less than or equal to the victim's loss: in such a case there is no net social benefit in allowing the offense to occur, and therefore there is no danger of overdeterrence.\textsuperscript{62} In fact, in this case the formula gives only a lower estimation, since any greater amount of punishment provides complete deterrence as well.\textsuperscript{63} If the offender's gain is probably greater than the victim's loss, Polinsky and Shavell's formula may be applicable, according to Hylton.\textsuperscript{64} The reason is that applying the gain-elimination rule here may result in overdeterrence, because potential offenders would invest too much in avoiding the offense.\textsuperscript{65} However, even in such cases the loss-internalizing formula provides only a lower estimation, because it does not take into account secondary harms, that is, indirect costs inflicted to unrelated parties (investments into protecting themselves

\textsuperscript{58} Id. (quoting Marc Galanter & David Luban, Poetic Justice: Punitive Damages and Legal Pluralism, 42 Am. U.L. Rev. 1393, 1450 (1993)).
\textsuperscript{59} Hylton, supra note 10.
\textsuperscript{60} Id. at 467.
\textsuperscript{61} Id. at 444.
\textsuperscript{62} Id. at 423.
\textsuperscript{63} Id. at 429.
\textsuperscript{64} Id. at 439.
\textsuperscript{65} Id. at 424.
from similar offenses). If the secondary harms are substantial, the gain-eliminating formula may be again more appropriate. The overcompensatory part of the so calculated damages represents optimal punitive damages. This approach is directly traceable to Bentham and Cesare Beccaria. The theory has been further developed and modified in Hylton's later works, including application to criminal law, and the property rules versus liability rules dichotomy. The original Hylton article has been cited with approval by the Seventh Circuit, albeit only once.

The idea of gain elimination is in much better correspondence with the general notion of punishment than the idea of loss internalization. The illicit gain apparently means the benefit earned in breach of a socially recognized norm. Accordingly, the gain-eliminating charge on the offender may be viewed as being roughly proportional to the

66. Id. at 435-36.
67. Id. at 444.
68. Id. at 421 n.1(citing CESARE BECCARIA, ON CRIMES AND PUNISHMENTS 43 (Henry Paolucci ed., Bobbs-Merrill 1963) (1764)).
69. Id. at 421 n.1(citing JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 166 (J.H. Burns ed., Clarendon Press 1996) (1781)).
70. Keith N. Hylton, The Theory of Penalties and the Economics of Criminal Law, 1 REV. L. & ECON. 175 (2005). Here, the economic approach is applied to criminal law. It is argued, in particular, that where the cost of transacting is less than the cost of enforcing, penalties should be set to eliminate offender's gains; otherwise, loss-internalization policy may be more efficient. Id. at 176-77. See also Steven Shavell, Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent, 85 COLUM. L. REV. 1232 (1985) (analyzing optimal use of nonmonetary penalties); Richard A. Posner, An Economic Theory of the Criminal Law, 85 COLUM. L. REV. 1193 (1985) (arguing that the major function of criminal law "is to prevent people from bypassing the system of voluntary" exchange).
71. Keith N. Hylton, Property Rules and Liability Rules, Once Again, 2 REV. L. & ECON. 138 (2006). It is argued, in particular, that "when transaction costs are high, the liability rule is preferable to the property rule," with a possible exception of certain subjective valuation-related circumstances. Id. at 153. When transaction costs are low, the property rule is preferable. Id. at 159. See also Louis Kaplow & Steven Shavell, Property Rules Versus Liability Rules: An Economic Analysis, 109 HARV. L. REV. 713 (1996) (explaining why possessory interests in things are generally protected by property rules, whereas interests in not suffering from harmful externalities are often protected by liability rules); Ian Ayres & J.M. Balkin, Legal Entitlements as Auctions: Property Rules, Liability Rules, and Beyond, 106 YALE L.J. 703, 706 (1996) (showing that efficiency may be enhanced "by allowing both sides successive options to take the entitlement back at successively higher prices"). The transaction-cost approach to property versus liability rules is traceable to Guido Calabresi & A. Douglas Melamed, Property Rules, Liability Rules, and Inalienability: One View of the Cathedral, 85 HARV. L. REV. 1089 (1972).
deviation from the norm. If no norm is breached, no charge is applicable, even where some harm is inflicted. However, Hylton’s approach does not seem to be complete either. First, the bright-line distinction between the cases where the offender’s gain is greater or less, respectively, than the victim’s loss, is rather ad hoc, that is, not derived from any general economic principles. Second, both parts of the rule are uncertain, providing only lower estimations. Most importantly, the rule apparently can result in underdeterrence in both parts. If the offender’s expected gain is small as compared to the victim’s loss, the offender charged with the amount of his illicit gains is not motivated to consider potential victims’ losses at all. In many cases whether or not an offense occurs is a matter of probability rather than that of a deliberate decision. Knowing that his expected punishment is small, a potential offender may underinvest into protective measures and, as a result, inflict a substantial harm.73 On the other hand, if an offender’s expected gain is large as compared to the victim’s loss, the offender potentially charged with the victim’s loss is motivated to go ahead with the offense, because his expected punishment is equal only to the victim’s loss. As mentioned above, this does not seem to be optimal where an important social norm is breached.

A number of other, more technical, attempts to explain punitive damages economically have been undertaken, but we will not consider them here.74

To summarize, there are two major competing economic approaches to the punitive damages problem: namely, loss internalization and gain elimination.75 Neither of them has yet won general acceptance. One of the approaches seems to be consistent with a part of case law; the other one, with another part (examples may be found in the original articles).76 This article will attempt to combine positive features of both approaches

73. See also A. Mitchell Polinsky & Steven Shavell, Should Liability Be Based on the Harm to the Victim or the Gain to the Injurer?, 10 J.L. ECON. & ORG. 427, 428 (1994) (arguing that in such a situation harm-based liability is superior to gain-based liability because of a possibility of judicial errors).
74. See, e.g., David Friedman, An Economic Explanation of Punitive Damages, 40 ALA. L. REV. 1125 (1989) (deriving punitive damages form the costs of litigation and the “elasticity of the supply of offenses”).
75. For a recent review of both approaches, see Yoram Keinan, Playing the Audit Lottery: The Role of Penalties in the U.S. Tax Law in the Aftermath of Long Term Capital Holdings v. United States, 3 BERKELEY BUS. L.J. 381, 390 (2006).
76. Polinsky & Shavell, supra note 8, at 901-04; Hylton, supra note 10, at 445-54.
and offer a simple utilitarian interpretation of the resulting formula.

IV. UTILITARIAN APPROACH

This section reviews the utilitarian ideology as applied in this article. For an economist, people are “rational actors,” or, essentially, personal-utility-maximizing machines.\footnote{See Thomas S. Ulen, Rational Choice Theory in Law and Economics, in 1 Encyclopedia of Law and Economics 790, 792 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) (“consumers have transitive preferences and seek to maximize the utility that they derive from those preferences, subject to various constraints”). Despite the deliberate-choice rhetoric, the choice may seem rather mechanical, because the utility-maximizing option is always chosen.} To achieve social policy goals, policy enforcers can program those machines by constantly feeding them with appropriate data, that is, carefully measured amounts of positive or negative utility (awards or charges, respectively; not necessarily in pecuniary form).

The technique of motivating individuals may fall into one of two categories. First, the goal can be maximizing the aggregate utility of society. To achieve this goal, each actor is fed with the amount of the social utility, whether positive or negative, attributable to his actions. In particular, if the actor has inflicted a loss to somebody, the actor is charged with the amount of the loss. For socially useful actions the actor is respectively rewarded. Then all members of society are motivated to cooperatively maximize the total social utility.\footnote{See Polinsky & Shavell, supra note 8, at 882.} Second, the goal may be holding actors to a certain socially-approved standard of behavior.\footnote{For the purposes of this paper I do not distinguish rules and standards. But see Lois Kaplow, General Characteristics of Rules, in 5 Encyclopedia of Law and Economics 502, 508 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) (classifying legal commands into rules and standards depending on whether the law is given precise content \textit{ex ante}, at the time of promulgation, or \textit{ex post}, at the time of adjudication).} To achieve this goal, one withholds from the actor the amount of his personal utility attributable to his deviation from the standard, thus de-stimulating the deviation.\footnote{See Hylton, supra note 10, at 425. There may be a temptation to withhold \textit{more} than that amount with a view to better secure compliance: perhaps to withdraw as much as possible. \textit{Cf. id.} at 440. However, excessively harsh penalties may be suboptimal for a number of reasons. \textit{See id.} at 425-26. One of them is a possibility of a court error: the non-zero probability to be demolished by a mistaken punishment induces actors to keep too far away form the socially-approved standard, which is suboptimal if the standard itself is the optimal equilibrium point.} In either case, each rational actor is tricked
into solving the social optimization problem when maximizing his own utility. However, while the first technique is designed "to persuade and allure," the second is intended to be "compulsory," in the words of Blackstone.81

The explicit aim of the first technique is total social utility maximization. Less obviously, the second, standard-based, method can also be reduced to the same principle. Indeed, what is the reason of the existence of socially-approved standards of behavior, whether legal or ethical? Arguably, this is again maximization of total social utility.

Speaking about legal standards, this proposition is well-established.82 Arguably, the purpose of law is maximizing total social welfare, including its non-pecuniary component, which, however, is theoretically counted into total social utility.83 Even if we are unsure about current lawmakers being an assembly of wise men aspiring to reach social happiness, interaction of confronting political powers in the conditions of free elections tends to produce the same result, that is, socially efficient legal norms.84 Arguably, in non-democratic societies the purpose of establishing legal standards is maximizing the utility of only a selected group of governing individuals.

Concerning court created law, consider the Learned Hand’s rule

81. See Blackstone, supra note 1. This is the same type of distinction as in the liability rules versus property rules dichotomy, stated in somewhat more general terms. For liability versus property rules discussion, see supra note 70.


83. Among economists, an allocation maximizing total social welfare is known as Kaldor-Hicks efficient or Marshall efficient. This is to be distinguished from Pareto efficient allocations, where no person can be made better off without making someone else worse off. Each Kaldor-Hicks efficient allocation is also Pareto efficient, but not vice versa (total social welfare often can be improved by making some people better off but some others worse off). See, e.g., Richard A. Epstein, Rent Control and the Theory of Efficient Regulation, 54 Brook. L. Rev. 741, 760 (1988); David D. Friedman, Price Theory: An Intermediate Text, Ch. 15, available at http://www.daviddfriedman.com/Academic/Price_Theory/PThy_Chapter_15/PThy_Chap _15.html (last visited Nov. 7, 2006).

84. This generally accepted intuition is mathematically qualified by the so called Arrow impossibility theorem demonstrating, essentially, that under certain reasonable assumptions no voting system based on individual preferences can possibly be used to define the social welfare function. See, e.g., Friedman, supra note 83, Ch. 15.
regarding negligence law.\textsuperscript{85} It essentially establishes a standard of behavior basing upon considerations of maximizing the expected total utility.\textsuperscript{86} This is despite the fact that the negligence standard is formally based on the economically ambiguous notion of a reasonable man.\textsuperscript{87}

Speaking about ethical standards, the wealth-maximization proposition is less clear. There are a number of competing theories of ethics.\textsuperscript{88} However, the only currently available quantitative approach is utilitarianism, first expressed by John Stuart Mill.\textsuperscript{89} The modern version of utilitarianism is called "rule-utilitarianism."\textsuperscript{90} According to this approach, ethical rules are also established with a view to maximize total social welfare, including its non-pecuniary component.\textsuperscript{91} That is, ethical standards ultimately originate from efficiency related reasons. I adopt this utilitarian approach for the purposes of this article.

Accordingly, both legal and ethical norms tend to be economically efficient, broadly speaking.\textsuperscript{92} To put it simply, society introduces a standard of conduct because it feels that breaching this standard

\textsuperscript{85} Note that the negligence rule is a hybrid of the two discussed techniques, in the sense of being standard-based but normally charging the offender with the loss inflicted.

\textsuperscript{86} U.S. v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947) ("if the probability [of injury] be called P, the injury, L; and the burden [of adequate precautions], B; liability depends upon whether B is less than L multiplied by P;" in short, B<L:P).

\textsuperscript{87} The lawyers' "reasonable man" is definitely not the same individual as the economists' "rational actor." Their motives drastically differ: the former takes into consideration interests of others, the latter is completely selfish. \textit{Cf. Restatement (Second) of Torts} § 283 (1965) cmt. b, and Ulen, \textit{supra} note 77, at 792 (defining the notions). The whole purpose of law is making rational actors behave like reasonable men.


\textsuperscript{89} \textit{Id.} at 6 (quoting John Stuart Mill, \textit{Utilitarianism}, in \textit{Great Books of the Western World} (2d ed., Encyclopaedia Britannica 1990)).


\textsuperscript{91} It is often argued that such a morally-neutral theory of ethics results in paradoxical notions, like socially efficient homicide. \textit{Cf.} Galanter & Luban, \textit{supra} note 58, at 1449 (rejecting such notions). As a counter-argument, consider the fact that human societies do justify homicide in certain cases on the grounds of social efficiency. Such cases include self-defense, death penalty, and war.

\textsuperscript{92} Of course, efficiency of social standards is only a long run correlation at the best; in the short run moral and legal norms may significantly differ from the economically optimal ones because of various kinds of prejudices, misunderstandings, and abuses. I neglect this qualification for simplicity of argument.
generates excessive social harm not covered by the compensatory techniques currently in place. Holding society members to such standards maximizes total social utility.

If we adopt the utilitarian explanation of social standards of behavior, why is the standard-based method of motivating individuals needed at all? Is it not easier to always withhold from an individual the social loss or credit him with the social benefit attributable to his actions, whether by way of private litigation or otherwise? As discussed, this method also maximizes total social utility, that is, brings about the same result.

There are several good reasons to stick to the standard-based technique. First, often there is no direct way to calculate the amount of the social utility produced or extinguished by an individual. For a simple example, consider littering: too many people are harmed in too small an amount to reliably calculate the total.93 Second, the system of enforcement is imperfect: offenders are not always detected, and courts rarely use fine-tuned economic formulas to calculate the amount of punishment. Third, real-life actors are not ideal risk neutral economically rational subjects. An injurer may be a risk-taker, that is, willing to take the risk of being punished in exchange for a disproportionately small benefit. In addition, an injurer may be insufficiently rational, or unable to see what is really beneficial for him. This is especially true where the social harm and, accordingly, possible future punishment is difficult to estimate.94

Introducing a specific norm of behavior society elects not to rely on the market "invisible hand" to strike the optimal balance of interests. Instead, it determines from some external considerations the standard of behavior it believes is optimal and invites everybody to follow this explicit standard. In case of only conditionally rational actors and/or uncertain amount of potential harm, direct punishment is a more reliable way to provide optimal behavior.

As we see from the above, introducing a standard of behavior and sanctions for its breach does not necessarily mean decreasing economic efficiency. On the contrary, this is an attempt to achieve economic efficiency under uncertain economic conditions. The additional social

93. Note that there may be indirect ways of calculating harm. In the case of littering, apportioned municipal expenses for street cleaning might provide a proper estimation.
94. See supra note 71 for further discussion of why property rules may be preferable to liability rules.
utility resulting from maintaining the standard may be characterized as
"structural utility," since it originates from holding society members to a
certain pattern of behavior.

We come to an important conclusion. On the one hand, for holding a
person to a social standard it is appropriate to withdraw from him the
utility attributable to his deviation from the standard. On the other hand,
to provide optimal behavior in the same situation, it is appropriate to
withhold from the person the amount of the social loss he inflicts.
Therefore, the two amounts are equal, provided the standard is optimal.
The gain of a person attributable to his deviation from a socially efficient
standard represents the amount of the social harm he thereby inflicts.
Mathematically, this is a corollary of the necessary condition of
extremality: in a maximum point of total utility its derivative is zero.
Therefore, at that point the marginal expected personal gain is equal to
the marginal expected social loss.\footnote{This statement oversimplifies things a bit. First, the relevant functions are not
necessarily differentiable. Second, if they are differentiable, a \emph{backward} deviation from
the standard produces social benefit; then for optimality purposes the person should be
rewarded, which is not normally done. Third, since the amounts are only marginally
equal, they may differ by an additive constant. It, however, may be disregarded for
deterrence purposes because the constant amount of social harm or benefit does not
depend upon injurer's behavior. Regardless, the "illicit gain" estimation of the inflicted
social loss seems mathematically reasonable, and is often the only available one.}

The logic may seem perverse: instead of deriving damages from
harm, we derive harm from damages. However, this corresponds to the
very notion of an indirectly measurable variable. If we believe that the
observable social world is essentially utility-driven, we have to attribute
to the actions we believe are socially useful or harmful a certain amount
of social utility, positive or negative, respectively. Just like individual-
associated utility, social utility may be measurable not directly, but
derivable from other directly or indirectly measurable variables, in this
case from the optimal amount of damages.

This provides a simple utilitarian interpretation of both methods of
motivating individuals. If we are able to estimate the amount of the
social harm inflicted by a person, we withdraw this amount from him.
Otherwise, using a presumably efficient social standard of behavior, we
withdraw from the person the amount of gain attributable to his deviation
from the standard. As argued, this amount is an indirect estimation of
the social harm inflicted. In both cases the person is motivated to
maximize total social utility, including his own and other people's utility.

V. COMPENSATION AND DISGORGEMENT

Before proceeding to economic analysis of the problem, I attempt to refine the terminology, specifically, the meanings of compensation and disgorgement. According to the Restatement of Torts, a tortfeasor is "ordinarily" liable either for the damage done to the victim, or "for the value of the benefit received through the commission of the tort," at the victim's option.96 Is the either-or rule optimal? Should the offender not be charged with both amounts? In case of a breach of the fiduciary duty of loyalty, for example, the situation seems to be somewhat different. To quote the Restatement of Agency:

If an agent has received a benefit as a result of violating his duty of loyalty, the principal is entitled to recover from him what he has so received, its value, or its proceeds, and also the amount of damage thereby caused; except that, if the violation consists of the wrongful disposal of the principal's property, the principal cannot recover its value and also what the agent received in exchange therefor.97

That is, the agent is generally charged both with the amount of his improper benefit and, in addition, with the amount of the loss inflicted.98 Similarly, under the Restatement of Trusts, a trustee who commits a breach of trust is not only liable for the resulting loss, but is additionally subject to such liability as necessary to prevent the trustee from benefiting personally from the breach of trust.99 Again, the duty-breaching trustee is charged with both his illicit gains and the trust losses.100

96. RESTATEMENT (SECOND) OF TORTS § 903, cmt. b.
98. See, e.g., Tarnowski v. Resop, 51 N.W.2d 801 (Minn. 1952) (an agent employed to purchase a business was bribed by the sellers and made false representations to the principal; upon discovery of the fraud and a successful action for rescission against sellers, the principal may recover from the agent the amount of the bribe and of the losses suffered).
100. Compare with the remedies for copyright infringement under the Copyright Act:
Punitive Damages

In many cases a tort effectively results in transfer of wealth from the victim to the tortfeasor (e.g., fraudulent appropriation). This means that the offender gains about the same value as the victim loses. Then, the amount of compensation and that of restitution are approximately equal, although because of different calculation techniques they may slightly differ. Accordingly, to compensate the victim properly, the offender may be charged with any of the amounts, whatever happened to be greater or else easier to prove. Perhaps this is the reason why the Restatement of Torts includes restitutionary awards into the notion of compensatory damages.

If, however, the tortfeasor, in addition to such a wealth transfer, makes some extra profit as a result of the tort, restitution effectively results in both a backward wealth transfer and disgorging the extra profit. If, in addition to the wealth transfer, some extra loss is inflicted to the victim, compensation means both a backward transfer and coverage of the extra loss, while restitution means a backward transfer and the extra profit disgorgement.

In such a general case, as well as in a situation where the tort committed is not related to any wealth transfer (e.g., trespassing), the Restatement of Torts rule does not seem to allow both compensation and disgorgement. Note, however, that in such cases, the disgorgement

“The copyright owner is entitled to recover the actual damages suffered by him or her as a result of the infringement, and any profits of the infringer that are attributable to the infringement and are not taken into account in computing the actual damages.” 17 U.S.C. § 504(b) (2004) (emphasis added). Consider also the remedies for a trademark infringement under the Lanham Act: a plaintiff is entitled to recover “(1) defendant's profits, (2) any damages sustained by the plaintiff, and (3) the costs of the action.” 15 U.S.C. § 1117(a) (emphasis added). However (despite the statutory “and”), courts, apparently influenced by the general tort rule, are reluctant to award both profits and damages. See United Phosphorus, Ltd. v. Midland Fumigant, Inc., 205 F.3d 1219, 1228 (10th Cir. 2000) (agreeing with the district court that “a plaintiff in a trademark infringement case generally cannot recover its lost profits in addition to the defendant's profits,” but refusing to expand this result “to a general rule that recovery of both types of damages is forbidden”) id. In a trademark infringement case actual damages may be multiplied by a factor not exceeding three. 15 U.S.C. § 1117(a)(2004). Note Judge Posner’s rationale for awarding treble damages: “Treble damages are a particularly suitable remedy in cases where surreptitious violations are possible, for in such cases simple damages (or profits) will underdeter; the violator will know that he won't be caught every time, and merely confiscating his profits in the cases in which he is caught will leave him with a net profit from infringement.” Louis Vuitton S.A. v. Lee, 875 F.2d 584, 588 (7th Cir. 1989). This rationale is the essence of the multiplication approach discussed in this paper.
amount may be included into punitive damages (added to compensation), since profiting from tortious activity is arguably reprehensible.

In principle, the wealth transfer, the victim’s extra losses, and the extra profit made by the tortfeasor all correspond to different events occurring with different probabilities. However, to simplify reasoning, I will assume that the wealth transfer and the resulting extra losses, if any, constitute a single event, while the extra profit, if any, made by the tortfeasor corresponds to a separate event occurring with its own probability and generally constituting a separate offense, like a breach of the duty of loyalty.

For the purposes of this article, I will understand “compensation” as being equal to the amount of the victim’s loss, including any wealth transfer and extra losses. “Disgorgement” will be understood as being equal to the amount of the offender’s profit above any wealth transfer. It will be argued that, generally, both the compensation and the disgorgement amounts should be included into the economically optimal charge on the offender.

VI. THE MODEL

Although the results of this work are intended to be sufficiently general, for calculations I will use a simple one-dimensional product-liability-oriented model. In this model, there is a population of potential injurers (say, widget manufacturers) and a population of potential victims (widget consumers). Each manufacturer exercises a certain level of care investing a certain amount of money to the product security. The level of care may be measured numerically; one-dimensional means that only one numeric parameter is needed for the estimation. Decreasing the level of care and, therefore, manufacturing costs, the injurer increases his profit from the sale of widgets.

With a certain probability, depending on his level of care, an injurer inflicts harm to a victim (say, a widget explodes). Apart from safety considerations, all widgets are identical. I assume that the personal harm inflicted to the victim is perfectly measurable in the monetary form. This “primary loss” is inflicted to legally recognizable victims. Note that in

101. Cf. Polinsky & Che, infra note 110, at 564; Hylton, infra note 122, at 4 (introducing similar models in a somewhat different context).
102. See Polinsky & Che infra note 110.
this model no direct wealth transfer occurs. In addition, let us suppose that society as a whole bears certain casualty-related costs (say, provides free medical treatment to all victims). This "secondary harm" is inflicted to society as a whole. For the sake of argument I assume that these costs are not coverable by private contractual means in this particular society. With a certain other probability the victim manages to collect sufficient evidence to hold the injurer liable by way of private litigation. Then, a court imposes damages on the injurer.

The problem is to determine the optimal amount of damages to impose. "Optimality" is understood as an economic one, in the sense of maximization of the total expected social utility, including benefits of producers and consumers, primary losses of victims, and secondary costs borne by society.

For a while, we assume that the litigation is free of charge and the outcome does not depend on parties' efforts. Further qualifications will be discussed in a later section.

VII. Optimal Charge

In this section I calculate the amount of damages providing optimal deterrence, that is, resulting in the optimal level of potential injurers' care. To begin, assume that the probability of liability is unity. As discussed above, for motivating manufacturers to maximize the total social utility, the amounts of utility corresponding to others' losses should be withdrawn from the manufacturers.

First of all, the optimal damages should include the primary-loss component. Note, however, that these losses should be withdrawn from the manufacturer only if they are really "caused" by him. It does not make sense to motivate a person to diminish losses that are beyond his control. To better reflect causation, a number of legal tools have been developed, including the proximate cause, contributory negligence, and comparative negligence concepts. However, to simplify reasoning, I assume that the safety of widgets is under full control of the

103. The concept of secondary costs is due to Bentham, supra note 69, and further explored by Hylton and others. See supra note 71 and supra note 70 and references therein.
104. Restatement (Second) of Torts § 281 (1965).
105. Id. at § 463.
106. See 57 B AM. JUR. 2D Negligence § 954 (2004).
manufacturers, which effectively results in a strict-liability rule. That is, a manufacturer should be held liable wherever the victim proves that she was injured by that manufacturer's widget.

Second, the optimal damages should include the secondary-harm component. To determine the optimal level of care from the secondary-harm perspective, the benefits of producing widgets (attributable both to manufacturers and to consumers) should be balanced against the amount of the expected secondary harm (against the social costs of medical care). Although both parts of the equation are not necessarily calculable directly, society may nevertheless estimate the equilibrium level of care from some external considerations, perhaps by the trial-and-error method. The corresponding level is identified with the "reprehensibility" threshold, falling below this level is reprehensible.

Although it is not easy to calculate the secondary harm attributable to the specific injurer's behavior, it is straightforward to hold him to the presumably optimal level of care by withholding from him the amount of his utility attributable to the deviation from that level. This will be the amount underinvested by the manufacturer into the safety of a widget, divided by the probability of his liability in case of underinvesting. Equivalently, in line with the gain-elimination technique, this is the total amount underinvested into safety of all produced widgets divided by the number of successful suits. Note that the amount withheld from the manufacturer is not his total income originating from the sales of widgets, but rather the amount of illicit savings originating from underinvestments into safety. It is proportional to the deviation from the socially-approved standard.

As argued above, this amount is an estimation of the secondary harms inflicted. If the socially-approved standard is not breached, no such charge is imposed. Theoretically, for a backward deviation from the standard a manufacturer should be rewarded, but I neglect this possibility here. Perhaps the resulting broken-line charge function does not bear much resemblance to the smooth expected-secondary-harm curve, but apparently this is the best available estimation.

To take into consideration the non-unity probability of holding the injurer liable, both components should be inflated by relevant multipliers. In line with Polinsky and Shavell, the first component (compensation) should be multiplied by the reciprocal of the probability of holding the manufacturer liable where a victim is injured by that manufacturer's widget. In line with Hylton, the second component
(disgorgement) should be multiplied by the reciprocal of the probability
to hold the manufacturer liable where his care is substandard. The latter
probability is a product of the probability to cause an injury where care is
substandard and the probability of holding the manufacturer liable where
an injury is inflicted.

Any overcompensatory damages are normally indiscriminately
referred to as punitive. However, their part originating from the first
component is not punitive in nature: it is not imposed for breaching any
social standard, but rather represents morally-neutral augmentation of
regular compensatory damages. This part may be called quasi-punitive.
Another possible denotation is supercompensatory damages,
emphasizing their origin. In contrast, the second part, the augmented
disgorgement, is a punishment for a breach of a social standard. This
part may be called truly punitive damages.

Note that the punishment is in fact imposed for an offense different
from the injury itself, namely, for a substandard level of care
(reprehensible behavior). In our model the difference is emphasized by
the strict-liability rule for imposition of compensatory and quasi-punitive
damages, on the one hand and a non-trivial threshold for imposition of
truly punitive damages on the other hand. The resulting level of care,
economically optimal from the manufacturer’s point of view, takes into
consideration both primary and secondary social losses; in any case it is
in compliance with the socially-approved standard.

This, however, leaves open certain issues. Punitive damages,
calculated this way, may be very harsh or even destructive for the
defendant: quite comparable with criminal sanctions. On the other hand,
the standard of proof in a civil case is substantially lower than it is in a
criminal case. 107 This suggests that due process may require some
reduction of the amount of truly punitive damages. However, I do not
further elaborate this issue here. 108

Further, society may choose to diminish secondary harms by ways

damages are not a substitute for the criminal process . . .”).
108. For a discussion of the desirability of a heightened standard of proof in punitive
damages cases, see Developments in the Law -- The Civil Jury, 110 HARV. L. REV. 1513,
1531-32 (1997). For further discussion, and a critical view on the economic approach
to punitive damages generally, see Daniel F. Thomas, Comment: Necessary Protection: An
Examination of the State Farm v. Campbell Standards Test and Why Economically
Efficient Rules Do Not Work at the Intersection between Due Process and Punitive
other than imposing punishment in private suits. In particular, manufacturers may be statutorily obliged to comply with a presumably optimal level of care. In line with the above argument, the proper statutory fine for a breach ought to represent the augmented disgorgement of the profit attributable to manufacturer’s deviation from the standard (according to the gain-eliminating principle). If such a statute is in place, it ought to preempt truly punitive damages because the statutory fine represents the same secondary social harm. However, it should not preempt quasi-punitive (supercompensatory) damages representing primary losses still recoverable by way of private litigation. For further calculations I assume that there is no statute preempting truly punitive damages.

To summarize, for the purposes of proper deterrence, the optimal damages consist of the following components: (1) compensatory damages equal to the victim’s loss, (2) supercompensatory damages being augmentation of the compensatory damages with a view to reflect the possibility of injurer’s escaping liability, and (3) truly punitive damages being similarly augmented disgorgement of the profit attributable to a deviation from a socially-approved standard. The compensatory and supercompensatory components represent primary losses inflicted to the plaintiff and other legally recognizable victims; the truly punitive part represents the secondary harm inflicted to society as a whole.

VIII. OPTIMAL AWARD

By now, only considerations related to influencing manufacturers have been considered. The charge on the injurer optimal in terms of proper deterrence has been calculated. Another important question is what the amount of the optimal award to the victim is.

It is not immediately clear that the whole amount withdrawn from the injurer should be passed to the injured. The total-social-utility maximizing strategy is neutral to any wealth transfers. From the total-social-wealth perspective, the money withdrawn from the injurer may be either passed to the victim or be kept by the state for further

109. If a statute is intended to directly mitigate primary losses by introducing a standard and a differentiated fine for a breach, then apparently the whole idea of deterrence by way of private litigation should be abandoned.
redistribution between society members or else divided between the victim and the state in any proportion. Accordingly, some other technique is needed for calculating the optimal award.

While charges on injurers are designed to motivate manufacturers, awards to victims motivate consumers. Accordingly, the criterion should be related to the optimization of consumers' behavior. Awards may influence consumers in two ways. First, covering losses related to the consumption of widgets stimulates this consumption. Second, awards stimulate private litigation efforts.

A number of authors investigated this issue, assuming that minimizing total social litigation costs should be a major policy objective. Notably, A. Mitchell Polinsky and Yeon-Koo Che have argued for "decoupling," that is, making the plaintiff's recovery different from the defendant's payment, for exactly this purpose. They figure that simultaneously increasing the charge on the defendant and lowering recovery for the plaintiff may result in the same level of deterrence of the undesired behavior of the defendants with fewer suits filed by plaintiffs. Indeed, decreasing recovery results in both decrease of probability of a suit and therefore, of liability and, in line with the above discussion, increase of the optimal damages to be charged. The average expected charge is constant, which provides the same level of deterrence. Assuming that litigation expenses are constant, decrease in the number of suits results in reducing litigation costs for society as a whole. This logic leads the authors to the idea of making the defendant's charge "as high as possible." This result has been further qualified by other studies.

111. Id. at 563.
112. Id.
113. See Albert Choi & Chris William Sanchirico, Should Plaintiffs Win What Defendants Lose? Litigation Stakes, Litigation Effort, and the Benefits of Decoupling, 33 J. LEGAL STUD. 323 (2004). The authors indicate that litigation expenses of parties are not at all constant, but depend heavily on the litigation stake. They argue that in some cases this may lead to the negative effect of decoupling: fewer filed cases may still mean more litigation expenses. Id. at 330. Contrary to possible intuition, the authors' calculations suggest that "optimal recovery is no less than optimal damages in high-stakes, deep-pocket suits" with error-prone juries. Id. at 340. See also Philip Bond, Optimal Plaintiff Incentives When Courts Are Imperfect, 2004 Meeting Papers 723, SOC'Y. FOR ECON. DYNAMICS, available at http://finance.wharton.upenn.edu/~pbond/research/plaintiff-incentives.pdf (last visited Sept. 26, 2006). The author concentrates on the idea that the courts are imperfect. Earlier studies implied that the court always reaches the correct decision (finding the
However, most cited authors seem to presume that litigation expenses are only a dead weight borne by society.

I respectfully disagree. Litigation costs are a source of maintenance of both the judicial system and private lawyers. Both play socially useful roles, making and improving law, in particular. It is economically efficient to maintain the judicial system. Being a kind of useful commodity, access to justice definitely carries a certain amount of social utility. Litigation costs represent the price of this commodity. Even high attorneys' fees are arguably close to economically optimal ones because of being formed in the free-market conditions, provided, of course, that the awards themselves are economically optimal. Therefore, suppressing total litigation costs as close to zero as possible does not seem like a good idea: it would effectively result in suppressing lawyers' useful activity.

This means that one should overcome the temptation to press down litigation by way of artificially decreasing awards. On the contrary, for the judicial system to be economically efficient, it should, in the first place, engage in redistribution of economically optimal amounts of utility. Awards and charges should be determined on the basis of most general economic considerations. The optimal level of litigation is then determined by the free-market forces; mathematically, it is derivable from litigation costs, as discussed in the following section.

If so, the only remaining idea is securing economically optimal consumption of widgets by establishing proper awards to the consumers harmed by exploding widgets. The award-optimization technique is based on the idea that insufficient awards may result in underconsumption of a socially beneficial product (after all, widgets are bought because they are useful), while excessive awards may result in defendant guilty when he is guilty and innocent when he is innocent). The author points out that, in fact, the outcome depends on litigation efforts of the parties. The efforts are measured by the parties' litigation expenses, including lawyers' fees, fact-investigators' fees, and judges' bribes. Id. at 8. Courts are classified into two categories. Courts of the first category are more susceptible to influence activities by a party arguing "against the facts" (e.g., a plaintiff arguing the defendant is guilty when he is in fact innocent) than to a similar activity by the other party. Courts of the second category are relatively resistant to influence activities of this sort, favoring the party arguing "with the facts." Id. at 3. In about thirty pages of math, the author comes to the conclusion that in the former case it is desirable to limit the plaintiff's reward; in the latter, full plaintiff incentives are preferable. Id. at 37. The author goes on to suggest that civil law courts tend to fall into the first category, whereas common law courts belong to the second. Id. at 42-43.
overconsumption of an inherently dangerous product.

In fact, the notion of economically optimal consumption is just another twist on the wealth-maximization strategy, because at the economically optimal point the total benefit gained by both manufacturers and consumers is maximized. Generally, the economically optimal level of consumption in a free-market economy corresponds to the price level at which the manufacturers’ supply equals the consumers’ demand. Graphically, this proposition is represented by a famous diagram featuring supply and demand curves intersecting at the optimality point.114

To start with, imagine such a diagram in a case where all widgets are absolutely safe. Further, if widgets do explode occasionally, each widget bears a certain amount of negative utility equal to the expected harm inflicted. For simplicity I assume that consumers are risk neutral or else that the harm may be fully compensated by insurers in exchange for a premium equal to the expected harm that is, at zero net insurance costs. If the whole amount of loss is borne by the victims or covered by their insurance policies, the consumers decrease their demand, and the demand curve shifts respectively. Of course, this is true only if the consumers are properly informed about the danger of explosion. The consumption is then fixed at a new, lower, socially optimal level and at a lower price. On the other hand, if the losses are fully transferred to the manufacturers, whether by way of private litigation or otherwise, the demand curve does not move, but the manufacturers decrease their supply to reflect payments to victims or the insurance costs. The supply curve shifts respectively, and the consumption is again fixed at the socially optimal level, albeit at a higher price reflecting new costs for the manufacturers.115

If, however, both consumers and victims are effectively charged with the losses caused by exploding widgets, both supply and demand

114. See, e.g., Friedman, supra note 83, Ch. 7, Fig. 7-2, available at http://daviddfriedman.com/Academic/Price_Theory/PThy_Chapter_7/PThy_Chapter_7.html (last visited Nov. 7, 2006).
115. The outcome will be economically efficient in both cases, that is, whether or not the losses are transferred from the consumers to manufacturers. This is a special case of the Coase theorem stating that, if there are no transaction costs, an efficient outcome will occur regardless of the initial allocation of entitlements (because the parties will bargain to restore efficiency). See Steven G. Medema and Richard O. Zerbe Jr., The Coase Theorem, in ENCYCLOPEDIA OF LAW AND ECONOMICS 836, 837 (Boudewijn Bouckaert & Gerrit De Geest eds., 2000).
decreases, and the resulting consumption level is suboptimal. This is the economic reason why the total amount of compensatory damages withdrawn from the injurer should be passed to the injured.

Arguably, the same is true for the augmented compensatory damages where injurers have a possibility to escape liability. Indeed, if the whole amount of augmented compensatory damages, including the compensatory and the supercompensatory parts is awarded to the victim, then the loss inflicted to the victim is equal to the award expectation. This is because the award itself is equal to the loss divided by the probability of the injurer’s liability, and the probability of obtaining this award is the same as the probability of the injurer’s liability. If so, the victims on average are fully compensated for their losses. Moreover, if the award is transferable to insurers, the victims may be fully compensated by the insurers and not only on the average, but in each particular case. Note that transferring augmented compensatory awards to insurers allows pressing down insurance premiums to zero in the ideal case of zero net insurance costs. Accordingly, awarding augmented compensatory damages to victims results in optimal consumption.

The situation is different for the truly punitive part of the charge on the injurer, reflecting secondary social harms. By definition, secondary losses, being losses borne by society as a whole, do not influence consumer behavior. If these losses are charged on the manufacturers, manufacturers decrease their supply to reflect secondary social costs, and the consumption is again fixed at a new, lower, optimal level. If, however, the amount of secondary losses is both withdrawn from the manufacturers and credited to the customers, then in addition to the supply decrease the customers increase their demand and the consumption is fixed at another, greater than optimal, level. Accordingly, truly punitive damages should not be awarded to victims, but rather should be kept by the state and perhaps applied to cover the secondary harm inflicted: in our model to pay medical care expenses.\(^\text{116}\)

In reality, the issue of motivating consumption has only marginal importance, at the best, for judicial practice. Arguably, taking into consideration the expected award in case of an injury requires too much rationality of consumers. However, although this argument may be true for individual consumers, for commercial purchasers (imagine a

\(^{116}\) Cf. Daughety & Reinganum, infra note 117, at 4 (citing the need to compensate society for the social harm as a basis for the split-award reform).
corporation regularly buying certain equipment) expected compensation in case of an accident may be a valid business consideration. If the product is regularly consumed, sometimes consumption results in a loss, and sometimes in damages awarded. To provide adequate compensation, in the long run the expected losses should be equal to the expected awards (taking into consideration the possibility of insurance, this means close-to-zero insurance premiums). Awarding less would be unjust, and awarding more is a windfall.

Besides resulting in low insurance costs, the stated rule is also in good correspondence with the natural notion of justice. On the balance, the proposed split-award rule seems to be economically reasonable.

To summarize, providing economically optimal consumption the whole amount of compensatory and supercompensatory damages should be awarded to the victim, while the whole amount of truly punitive damages should be kept by the state.

IX. LITIGATION COSTS

I now turn to the issue of litigation costs, both in terms of direct litigation expenses, including legal fees, and in terms of the personal effort of the plaintiff.

117. On the negative side, if any split-award statute is in place, both parties are induced to cut out the state by way of settlement, which somewhat spoils the ideal picture presented here. See Andrew F. Daughety & Jennifer F. Reinganum, Found Money? Split-Award Statutes and Settlement of Punitive Damages Cases, (Dep't of Econ., Vanderbilt Univ., Working Paper No. 00-W01R, 2001) available at http://www.vanderbilt.edu/Econ/wp-archive/workpaper/vu00-w01R.pdf (last visited Oct. 1, 2006); for a published version see 5 AM. L & ECON. REV. 134 (Oxford U. Press 2003). Arguably, for avoiding the distortion the state should have a share in the settlement amount. However, in the proposed model I completely disregard the possibility of a settlement for the sake of simplicity. On why the parties fail to settle, see Waldfogel, infra note 122.

118. An alternative idea, due to Hylton, is that the whole amount of supercompensatory damages (the “social compensation” portion) could go to a special fund from which other victims could collect damages, while truly punitive damages (the “gain elimination” portion), if any, could go to the state. Email from Keith N. Hylton, Professor of Law, Boston University to Sergey Budylin, Article in Punitive Damages (Nov, 20, 2006) (on file with author). Economically, this is similar to passing supercompensatory damages to insurers. Of course, implementation of this idea would require an even more radical statutory reform.

119. On the issue of litigation costs and optimal damages, see Keith N. Hylton, Welfare Implications of Costly Litigation under Strict Liability, 4 AM. L. & ECON. REV. 18, 19-20
Here I do not question the "American rule" stating that each party bears its own litigation costs. That is, the issue of whether or not the injurer should pay victims' costs for enhancing economic efficiency is not discussed. However, I argue that a part of the costs should be borne by society, also a beneficiary in the suit, to achieve economic optimality.

Direct litigation expenses are important for the victim's cost-benefit analysis when she decides whether to sue the injurer. To begin, assume the litigation expenses are constant, that is, they do not depend on the amount of the award. The victim sues only if her expected award is greater than the litigation expenses.\textsuperscript{120} On the other hand, from the social point of view it is efficient to sue if the total social loss related to the injury is greater than the litigation expenses; otherwise, avoiding litigation is economically efficient. If only compensatory damages are awarded to the victim, it results in a suboptimal number of suits filed, because the primary social loss attributable to others, as well as secondary social harm, is not taken into consideration by the victim. The problem is only partly solved if the award, as suggested above, includes both a compensatory part and a supercompensatory one, representing primary social losses, but not a truly punitive part, representing secondary harm. To cure the problem, the state should bear a part of the victim's litigation expenses proportional to the state's award, that is, to the truly punitive damages. It means that a relevant part of the plaintiff's litigation expenses should be additionally awarded to the plaintiff, and the same amount should be deducted from the truly punitive damages awarded to the state.

Assume now, perhaps more realistically for large awards, that the litigation expenses are not constant but rather proportional to the amount of the award.\textsuperscript{121} This means that the victim sues in any case, as long as the coefficient of proportionality is less than unity. Therefore, one should take into consideration some finer factors to define the optimal policy. Up to now we presumed that a court always draws the right decision, that is, holds the injurer liable provided that sufficient evidence is presented. Perhaps now it is time to recollect that real courts are

\footnotesize{(Oxford U. Press 2002) (showing in part that when litigation is costly, the optimal award is less than or equal to damages plus plaintiff's litigation cost; the model ignores secondary costs and low detection probabilities).}

\textsuperscript{120. }Cf. Polinsky & Che, supra note 110, at 6.

\textsuperscript{121. }Cf. Choi & Sanchirico, supra note 113, at 325.
imperfect and occasionally err in favor of one of the parties.\textsuperscript{122} The decision may substantially depend on parties’ efforts, mostly represented by their litigation expenses, including attorneys’ fees.\textsuperscript{123}

If an error in the plaintiff’s favor is as likely as an error in the defendant’s favor, the resulting uncertainty does not significantly spoil the picture of the expected losses being equal to the expected charges. Errors in favor of either party mutually compensate each other on the average. If, however, as suggested above, the parties’ litigation stakes are different, then so are their attorneys’ fees, and the court is more likely to err in favor of the injurer, whose stake is greater and attorney is better. To provide equality of chances, the plaintiff attorney’s fees should also be proportional to the total amount of damages, rather than to the amount of the plaintiff’s award. This again means that the state should bear a part of litigation expenses proportional to its share in the award.

In both approximations, fixed-cost and fallible-court ones, we arrive to the same result: the optimal victim’s award should include a part of her litigation expenses proportional to the state’s share in the award, that is, to truly punitive damages.\textsuperscript{124} This amount is covered by the state, being deducted from its part of the award rather than constituting an

\textsuperscript{122} In alternative (or in addition) to the fallible-court hypothesis, one can assume that the parties estimate their chances with a random mistake. See, e.g., Joel Waldfogel, \textit{Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation}, 41 J. L. \\ & ECON. 451 (1998). According to the author, the main economic puzzle of litigation is why parties fail to settle (basing on facts in hand) and instead chose to proceed to costly trial. \textit{Id.} at 451. There are two conflicting theories to explain this phenomenon. According to the theory of “diverging expectations” (DE), the parties estimate their chances to win at a trial with similar random mistakes. \textit{Id.} According to the theory of “asymmetric information” (AI), one of the parties is better informed about the facts and law related to the case, and the other estimates its chances with a considerable mistake. \textit{Id.} at 452. To reconcile the theories, the author points out that adjudication is not a single-moment event. The author analyzed 65,000 cases to arrive to the conclusion that, initially, parties are better described by the AI theory, but later tend towards the DE theory. \textit{Id.} at 474-75. This is because (1) most hopeless cases are adjudicated out early and (2) the parties become better informed, especially during the discovery. \textit{Id.} See also Keith N. Hylton, \textit{An Asymmetric Information Model of Litigation}, Law \\ & Econ. Working Paper No. 00-03, BOSTON U. SCH. OF L. (2000), \textit{available at} http://papers.ssrn.com/sol3/papers.cfm?abstract_id=239325 (last visited Nov. 16, 2006) (offering a “cradle-to-grave” mathematical model of tort liability).\textsuperscript{123} Cf. Bond, \textit{supra} note 113, at 8 (defining litigation expenses as including attorneys’ fees, fact-investigators’ fees, and bribes).

124. A different function for the litigation expenses (as depending on the award) may lead to a somewhat different proportion; however, the stated rule of thumb seems to be sufficient in most cases.
additional charge on the defendant. The result seems to be completely fair; apparently, fairness tends to be economically efficient.

In speaking about personal effort of the plaintiff, the reasoning is essentially the same. Apparently, the state should partly reward the plaintiff for her effort, in proportion to the state’s share in the award. This again seems to be fair: the plaintiff effectively plays the role of a state prosecutor, in part defending interests of the general public.\textsuperscript{125} The problem is how to evaluate the personal effort in monetary form. There does not seem to be a clear-cut economic way to calculate the optimal reward directly, unless various technical parameters are known, such as “the elasticity of the supply of claims.”\textsuperscript{126} Perhaps the best readily available estimation is an apportioned salary of a real state prosecutor (of course, only if we believe that her salary is economically optimal).

Anyway, this additional reward should not be too high. Besides overstimulating consumption, as discussed above, excessive awards stir litigation above the equilibrium level. The optimal award proposed above is intended to achieve economic neutrality for consumers: on the average, a consumer is fully compensated both for the losses associated with the consumption and for the costs associated with pursuing public interests, while bearing the personal part of the costs. If, however, the award includes an extra amount, then litigation \textit{per se} can produce net profit for plaintiffs on the average and, accordingly, has separate consumer value for them. Then plaintiffs are overmotivated to sue.

In the fixed-cost approximation such a judicial windfall means that the victim sues even where the total social loss, including her own loss, properly augmented, is less than the total litigation costs. This is obviously suboptimal. In the fallible-court approximation a potential windfall means that plaintiffs are induced to file poorly-grounded or even frivolous claims: even if the probability of success is low, the average expected net profit of the plaintiff is positive provided the excessive part of the award is sufficiently large. In addition to burdening the judicial system, plaintiffs may be willing to share their windfall with attorneys creating excessive demand for legal services.

To summarize, for providing an economically optimal litigation rate, the state should bear a share in litigation expenses proportional to its

\footnotesize
125. \textit{But see In re Exxon Valdez}, 270 F.3d 1215, 1228 (9th Cir. 2001) (warning that the “private attorney general” metaphor should be narrowly watched).

126. \textit{Cf.} Friedman, \textit{supra} note 74 (introducing “the elasticity of the supply of offenses”).
share in the award. In addition, the state should provide a fair compensation for the personal litigation effort of the plaintiff, attributable to the state share in the award. The amounts added to the victim’s award should be deducted from the state’s part of the award, rather than additionally charged on the defendant.

X. PRACTICAL APPLICATION

In this section I discuss consistency of the proposed recipe with the existing law and exemplify the approach application by discussing several paradigmatic cases.

The proposed formula for calculating the optimal charge on the injurer is in relatively good correspondence with the existing law. If the probability of liability is unity and no additional standard is breached, then it results in the common compensatory rule. The formula is also in direct correspondence with the duty-of-loyalty breach rule prescribing both compensation and disgorgement. On the face of it, the formula is in contradiction with the Restatement of Torts rule, authorizing either compensation or restitution. Note, however, that the disgorgement component arises in the optimal award only if a separate social standard, identifiable with the reprehensibility notion, is breached. Where the standard is breached, the formula results in imposition of truly punitive damages, which is in correspondence with the Restatement of Torts. However, where the overcompensatory part of the optimal award is quasi-punitive, it is not imposed for a breach of any additional norm, contrary to the Restatement of Torts approach. This implies that the existing law is suboptimal in this particular respect.

The formula seems almost in perfect correspondence with the common law factors cited by the Supreme Court in Haslip, discussed above: (a) the harm inflicted and the harm likely to occur enter into the first component of the formula, and the second component is an estimation of the secondary harm; (b) the degree of reprehensibility and the duration of the offensive conduct corresponds to the deviation from a social standard measured by the illicit profit contributing to the second component of the formula, and the frequency of similar past conduct

127. Unless one classifies as reprehensible a defendant’s attempt to conceal his offense, thus diminishing the probability of detection. See Polinsky & Shavell, supra note 8, at 898.
corresponds to the reciprocal of the probability of liability factor; (c) the profitability of the wrongful conduct directly contribute into the second component of the formula; (d) the total wealth of the defendant is relevant in intentional non-economic tort cases (as discussed below), and the total profit related to the wrongful activity is relevant in the economic cases, according to the formula; (e) litigation costs should be partly awarded to the plaintiff, according to the formula; (f) imposition of the criminal sanctions for the same offense has the same purpose as truly punitive damages and, therefore, should be taken into account when imposing punitive damages; and (g) and the same is true for civil sanctions.

The approach sheds some light on the punitive damages preemption issue. As argued above, a statute directly regulating potential injurers' behavior with a view to mitigate secondary social harms ought to preempt truly punitive damages, as serving the same purpose, but not quasi-punitive ones. If, however, the statute imposes only a fixed fine for a breach, insufficient to eliminate the whole illicit gain, truly punitive damages may still be applicable, with the amount of the fine deducted from the amount of damages.

In many practical cases, either the quasi-punitive or the truly punitive part is substantially less than the other one. Then, the lesser part can be safely neglected for the practical purposes of optimal damage estimation. Accordingly, most cases cited in support of their positions by either Polinsky and Shavell or Hylton may be also viewed as supporting the combined recipe.

Unfortunately, the formula contains parameters that are not necessarily directly measurable. Nevertheless, it provides a convenient framework for quantitative estimations, where the traditional common law factors leave one without any idea about the absolute amount of the punitive damages to impose. Importantly, the approach reconciles economic efficiency with ethics. If adopted, the formula makes reprehensible behavior economically inefficient, both in social and personal terms.

The proposed split-award formula is in correspondence with the ideas underlying existing split-award statutes. Arguably, those ideas are avoiding overmotivating plaintiffs, on the one hand, and compensating society for the harm inflicted, on the other hand. This is what the proposed formula does. However, the proposed formula is more fine-tuned that the existing statutes. They normally prescribe a fixed-
proportion split. In the absence of a so fine-tuned statute a court or rather a jury must choose between either properly influencing the injurer (the option that is normally preferred), properly motivating the plaintiff, or fixing upon somewhere in the middle. Achieving both goals within the fixed-split framework is generally impossible.

The proposed litigation-costs rule is only in rough correspondence with the cases and statutes either awarding litigation expenses to the plaintiff or awarding punitive damages with a view to reward the plaintiff for her socially useful effort. Below I discuss several hypotheticals, mostly based upon real court cases, to exemplify how the approach works.

To deviate from the explicitly economic product liability model, consider battery. Suppose that A hits B and inflicts harm resulting in pain and humiliation of, say, fifty dollars. A, in turn, feels pleasure as a result of the hitting, in the amount of one hundred dollars.128 If this is the end of the story, the result seems awkward: it is economically efficient for A to hit B and pay compensatory damages. However, most people would say something is wrong here; hitting other people should not be socially efficient.129

My explanation is that A hitting B not only harms B, but also inflicts indirect harm to society as a whole, breaching social peace and order. Living in a place where people regularly fight one another definitely reduces the quality of life of even those who are not hit personally. This secondary harm is difficult to calculate directly. However, it is straightforward to deter A from hitting; the worth of pleasure, one hundred dollars, should be extracted from him. As argued, this is an estimation of the social harm inflicted, and the whole amount of it should be passed to the state. The harm inflicted to the battered person, fifty dollars, should be charged additionally and passed to the victim. If the probability of liability is less than unity, then for achieving proper motivation both amounts should be inflated by the reciprocal of that probability. If the offender is held liable in, say, one case of two, then the victim should be awarded one hundred dollars, and the state two hundred dollars. Of them fifty dollars are compensatory damages, fifty dollars are supercompensatory damages, and two hundred dollars are

128. Cf. Friedman, supra note 74 (stating a similar hypothetical).
129. Arguably, in certain societies fights may be permissible (e.g., where both parties are pleased to fight (consider knight combats)).
truly punitive damages.

Note that the dollar amount of A's pleasure, although not directly measurable, arguably depends upon A's personal wealth: the richer a person is, the less the same amount of money is worth to him. This is a reason for making punitive damages dependent upon personal wealth.\(^{130}\) However, the dependence of damages upon wealth is reasonable only for non-commercial intentional torts (intentional torts committed for the reasons of personal pleasure).\(^{131}\) Applying the same principle in purely economic cases, such as product liability ones, is erroneous. A corporation does not injure consumers for pleasure; accordingly, the proper base for calculating punitive damages, if any, is not that corporation's net worth but rather its illicit economic profit attributable to the offense. If the net worth base is erroneously used for calculating damages, then high net worth individuals and corporations are sued disproportionally often, which results in decreasing economic efficiency.

Consider a case where a fast food restaurant visitor is seriously burned by hot coffee purchased in the restaurant while holding the open coffee cup between her legs in a running car.\(^{132}\) The plaintiff's argument is that the coffee was too hot, and therefore the restaurant is liable. In accordance to the discussion above, there are two separate questions to answer. First, whether the restaurant is liable under either the strict liability or negligence theory. If it is, the compensation amount may need augmentation to reflect the possibility of escaping liability.

---

130. A corollary, perhaps surprisingly, is that when a rich person hits a poor one for pleasure, more secondary social harm is generated than when a poor person hits a rich one for pleasure.


132. Cf. Martin E. Segal, Preventive Law for Business Professionals 12 (Thomson Learning 2005), available at www.swlearning.com/pdfs/chapter/0324225741_1.PDF (last visited Oct. 1, 2006) (citing Liebeck v. McDonald's, CV-93-02419 (D.N.M. 1994) (unpublished)). In that case, the jury awarded $160,000 in compensatory damages and $2.7 million in punitive damages. The coffee temperature was about 180 degrees Fahrenheit [which does not, in fact, seem to be "super-hot"]. The most damaging testimony to McDonald's was from their own witnesses: in particular, their quality-assurance manager testified that McDonald's had deliberately decided not to warn customers of the hot coffee danger, and their human-factors engineer testified that hot coffee burns were statistically insignificant when compared to the billions of cups of coffee McDonald's sells annually. Id. at 13. Indeed, in the preceding decade McDonald's received some seven-hundred reports of coffee burns and settled relevant claims for some $500,000 [which does not seem an extremely high loss for the restaurant]. Id. at 12.
However, where the restaurant escapes liability because of the victim’s unwillingness to sue, the respective amounts of harm are probably small. Accordingly, the multiplicator for the first component is arguably not much more than unity. The second question is whether the restaurant was reprehensible in serving coffee at the particular temperature. Obviously, this threshold should be different. Arguably, no temperature is reprehensible since most customers want coffee as hot as possible. Only if the restaurant was reprehensible, truly punitive damages should be charged. Then, they may be calculated as the restaurant profit received as a result of serving coffee that is too hot. Apparently, it is a certain portion of the restaurant’s total coffee sales; this was charged in the real case. However, the approach promoted in this article suggests awarding the truly punitive part of the damages to the state, partly with a view to avoid inducing frivolous claims.

One interpretation of the actual case is that the jury considered the restaurant’s balancing of their profits against customers’ possible burns as *per se* reprehensible. However, managers of all big corporations balance the costs of possible accidents against the costs of preventive measures on a regular basis. Under the utilitarian approach, such balancing is a part of legitimate cost-benefit analysis, absent additional social limitations. Accordingly, the correct question for the purposes of the punitive damages imposition is whether a separate pre-established social standard was breached. This breach could come from serving coffee at a temperature above a certain threshold, by not properly warning customers, or otherwise.

In the next case, a supertanker ran aground, spilling oil onto a beach causing an environmental disaster. The captain of the tanker was apparently drunk. As argued by Polinsky and Shavell, compensatory damages should not be augmented here because the probability of imposing damages is about unity (that is, in terms of their theory, no

133. The amount of punitive damages ($2.7 million) was apparently derived from the McDonald’s daily coffee sales ($1.35 million). *See* http://www.citizen.org/congress/civjus/tort/articles.cfm?ID=785. Both parties appealed; however, before appeals could be heard, the parties settled at an undisclosed amount. *Id.*

134. *Cf.* Polinsky & Shavell, *supra* note 8, at 903. In 1989, the supertanker Exxon Valdez ran aground and polluted some one thousand miles of Alaskan coastline. *Id.* at 903-04. The tanker’s captain was found to have violated regulations governing alcohol consumption. *Id.* at 904. In the district court the plaintiffs, including fishermen and Alaskan natives, were awarded several hundred million dollars in compensatory damages and $5 billion in punitive damages. *Id.*
punitive damages are applicable).\textsuperscript{135} But, should truly punitive damages be imposed? According to the proposed approach, it depends upon whether a recognized social standard has been breached. An arguably insufficiently safe tanker is not such a breach; no relevant standard was established at that time. However, drunk driving is clearly reprehensible in respect of supertankers. Further, the corporation was arguably reprehensible in not securing the sobriety of its captains.

Accordingly, punitive damages should be imposed: (1) upon the captain, depending on his wealth; and (2) upon the corporation, depending on total costs of ensuring sober captains for its tankers. Neither amount seems to be extraordinarily large compared to the compensatory part of the award. In the real case, the punitive damages the jury imposed on the corporation were incomparably greater. However, as the Ninth Circuit noted, the high clean-up costs borne by the corporation themselves provided a large incentive to avoid accidents.\textsuperscript{136}

\textsuperscript{135} Id.

\textsuperscript{136} In re Exxon Valdez, 270 F.3d 1215, 1244 (9th Cir. 2001). On appeal the Ninth Circuit vacated the $5 billion punitive damages award and remanded the case to determine a lower award. It was noted that Exxon had been separately punished for the environmental harm; accordingly, the case was "about commercial fishing" rather than "about befouling the environment." Id. at 1221. The plaintiffs were almost entirely compensated for their damages; therefore, the punitive damages was awarded to punish Exxon rather than to pay back the plaintiffs. Id. Punitive damages were not preempted by the Clean Water Act because the award vindicated only private interests, not the public interest in punishing Exxon for harm to the environment. Id. at 1231. It was agreed that Exxon's conduct was reprehensible because it knew that the captain was an alcoholic. Id. at 1242. However, the court held that in light of the constitutional standards recently established by the Supreme Court in \textit{Gore} and \textit{Cooper} (both came down after the district court ruled) the punitive damages award, equaling seventeen times the compensatory award, was excessive. Id. at 1246. On remand the district court, somewhat unwillingly, reduced the punitive damages award from $5 billion to $4 billion. In re Exxon Valdez, 236 F. Supp. 2d 1043, 1068 (D. Ala. 2002). The Ninth Circuit again vacated and remanded to reconsider in light of the new Supreme Court \textit{Campbell} decision. Sea Hawk Seafoods, Inc. v. Exxon Corp., No. 30-35166, 2003 U.S. App. LEXIS 18219 at *2 (9th Cir. Aug. 18, 2003). On remand the district court grudgingly reduced (or rather increased) the punitive damages from $5 billion to $4.5 billion. In re Exxon Valdez, 296 F. Supp. 2d 1071, 1110 (D. Ala. 2004). On appeal, the Ninth Circuit vacated the judgment and remanded the case again, with instructions to reduce the punitive damages award to $2.5 billion. In re Exxon Valdez, 472 F.3d 600 (9th Cir. 2006), amended by In re Exxon Valdez, 490 F.3d 1072 (9th Cir. 2007). The court indicated that because Exxon's conduct was not willful, and because Exxon promptly took steps to ameliorate the harm it caused, "a punitive damages award that corresponds with the highest degree of reprehensibility does not comport with due process when Exxon's conduct falls squarely in the middle of a fault continuum." In re Exxon Valdez, 490 F.3d 1072, 1073. Accordingly, the five-to-one ratio between punitive damages and the amount of
Suppose, now, that after the disaster society realized that the secondary social harm was underestimated, and chose to introduce a direct legal standard of safety, namely, double-hulled tankers. Now, if the same scenario repeats with an illegal single-hulled tanker, a new component arises in the punitive damages part; a punishment for a breach of the standard. As argued above, the optimal amount of punitive damages in such a case is equal to the augmented profit attributable to the illegal tanker. One idea of its estimation is the difference in the prices of a double-hulled tanker and a single-hulled one, multiplied by the number of illegal single-hulled tankers operated by the corporation. Another idea is to apportion the total profit of the corporation to the number of its illegal tankers. Accordingly, we have two potential calculatory bases for punitive damages: savings from cheaper tankers and profits from illegal tankers.

Which calculation base is more appropriate? The question does not seem to have a clear-cut answer. The solution depends on what amount of savings or profits is considered as attributable to the injurer’s offensive behavior. The ambiguity originates from the fact that, here, the applicable model is two-dimensional rather than one-dimensional. This means that to avoid offense, the injurer could either acquire double-hulled instead of single-hulled tankers or else cease their business altogether. Both options are perfectly legal. However, they result in different calculations of illicit gains attributable to exploitation of single-hulled tankers. Apparently, the question should be answered on a case-by-case basis, taking into consideration the real comparative importance of the offense, whether the injurer made illicit savings in an otherwise legal business or whether the business was illegal altogether.

---

138. Similar issues routinely arise in intellectual property law. If a trademark infringer is to account for his profits, 15 U.S.C. § 1117(a) (2004), should he account for all profits relating to goods bearing the infringing mark or only for the extra profit specifically originating from the infringing mark use? Here, the problem can be partly solved by burden of proof distribution. See, e.g., Mishawaka Rubber & Woolen Mfg. Co. v. S. S. Kresge Co., 316 U.S. 203, 206-07 (1942) ("The plaintiff of course is not entitled to
In the final example, a hotel knowingly rented out rooms heavily infested with bedbugs. Although in the words of Judge Posner, bedbug bites are “painful and unsightly,” the monetary measure of the harm inflicted to guests is arguably small. Even if properly augmented to incorporate compensation attributable to non-suining victims, compensatory damages may be insufficient to persuade the hotel’s management to treat the rooms against bugs. On the other hand, unsanitary conditions in hotels are arguably socially harmful, which means truly punitive damages are applicable. Note that they are imposed for breaching a separate standard, that is, for a willful refusal to treat the rooms, whereas the augmented compensation ought to be available even in the case of the hotel’s negligence. Truly punitive damages are calculated as the profit attributable to renting the infested rooms. Is the attributable profit equal to the whole profit from the rooms or just to the amount saved in anti- bug treatment costs? Since running a heavily

profits demonstrably not attributable to the unlawful use of his mark. The burden is the infringer’s to prove that his infringement had no cash value in sales made by him.” (citations omitted). Similarly, in a copyright infringement case the copyright owner may be entitled to only a relevant share of the infringer’s profit. 17 U.S.C. § 504(b)(2004). See, e.g., Sheldon v. Metro-Goldwyn Pictures Corp., 309 U.S. 390 (1940) (affirming apportionment of an award of profits against the studio so as to give the plaintiff only the portion of the profit attributable to the use of the copyrighted material).

139. Cf. Mathias v. Accor Econ. Lodging, Inc., 347 F.3d 672 (7th Cir. 2003). Bedbugs were discovered in a motel in 1998. The motel declined to pay $500 for spraying all rooms, although it charged upwards of $100 a day for a room. Id. at 674. By 2000 the infestation reached “farcical proportions.” Id. at 675. Two motel guests attacked by bedbugs brought a suit alleging that the motel was guilty of “willful and wanton conduct.” The jury awarded $5,000 as a compensation and $186,000 as punitive damages to each of the guests. Id. at 674. As the court notes, “[i]t is probably not a coincidence” that the award to each plaintiff is exactly equal to $1,000 per room in the motel. Id. at 678. The Seventh Circuit affirmed, having refused to apply the Campbell single-digit-ratio guideline, on the grounds that punitive damages must be proportional to the wrongfulness of the defendant’s action while the compensable harm done was slight and difficult to quantify. The court also cites the need to finance the lawsuit as a reason for the imposition of punitive damages. Id. at 676-77. The opinion by Judge Posner provides a concise but profound review of the history and theory of punitive damages law. The multiplicator approach is explicitly endorsed (“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.”). Id. at 677. An economic analysis of the case is provided by Steven Shavell, On The Proper Magnitude of Punitive Damages: Mathias v. Accor Economy Lodging, Inc., 120 HARV. L. REV. 1223, 1226 (2007) (arguing in part that “nonlegal economic incentives,” such as reputational concerns, may significantly reduce the need for punitive damages).

140. Id. at 675.
bedbug infested hotel is altogether illegal, the proper measure would be the entire profit of the hotel for the whole period of its infestation. Notably, the jury’s award in the real case seems to be based on the income from all rooms in the hotel for a certain period (two thousand dollars per room). Under the promoted approach, the whole amount of the truly punitive damages should go to the state, but the plaintiff should recover from that amount a large part of her litigation expenses and fair compensation for her personal effort.

To summarize, the proposed optimal-charge formula is in relatively good correspondence with the existing law, but the proposed split-award and share-cost formulas differ from the existing statutory rules, the proposed formulas being more economically fine-tuned.

XI. Philip Morris USA v. Williams

As mentioned above, the most recent U.S. Supreme Court punitive damages case was Philip Morris USA v. Williams. In a badly split (5-4) decision, the majority of the Court arrived to a conclusion that seems to completely strike down all punitive damages theories offered by all economic schools. Specifically, the Court held that “the Constitution’s Due Process Clause forbids a State to use a punitive damages award to punish a defendant for injury that it inflicts upon non-parties.” Apparently, the economically-minded analysts are left with the only hope that this is not the last word of the Court on the subject matter. In this section I discuss the case in some detail and confront the resolution suggested by the above economic analysis with the actual outcome based on due-process considerations.

For more than forty years, Jesse Williams smoked Marlboro cigarettes, up to three packs a day. When Williams’ family told him smoking was dangerous, he referred to publications asserting it was not. In 1997, Williams died of lung cancer caused by his smoking. Williams’ widow, Mayola, filed a lawsuit in an Oregon court against Philip Morris, Marlboro’s parent company, claiming negligence and fraud. She based her fraud claim on the tobacco company’s forty year publicity campaign. According to the claim, the company knew all along or at least for most of the forty years that smoking was dangerous, but tried to persuade the.

142. Id. at 1063.
public that the danger was unproven. The alleged reason was to give smokers a justification for continuing to smoke. Accordingly, the company as well as the whole tobacco industry intended to deceive smokers like Williams, and in fact deceived them. Since the jury ruled for the plaintiff, the facts here are stated in the light most favorable to the plaintiff, following the Oregon Supreme Court opinion.\textsuperscript{143}

The jury found for the plaintiff on both the negligence and fraud claims. It awarded more than $820,000 in compensatory damages of which only about $20,000 were economic damages, and $79.5 million in punitive damages.\textsuperscript{144} The trial court capped the non-economic compensatory damages at $500,000 and, most importantly, reduced the punitive damages award to $32 million.\textsuperscript{145} Note that even the reduced award is far beyond the \textit{Campbell} “single-digit-ratio” guidepost. On the plaintiff’s appeal, the Oregon Court of Appeals reversed the trial court and reinstated the punitive award.\textsuperscript{146} The U.S. Supreme Court granted certiorari and remanded the case to the Oregon Court of Appeals to be reconsidered in light of \textit{Campbell}.\textsuperscript{147} The Court of Appeals did so, but came to the same conclusion as before.\textsuperscript{148} The Oregon Supreme Court affirmed, noting in part that “the guideposts are only that — guideposts.”\textsuperscript{149} It also noted that Philip Morris’s conduct was “extraordinarily reprehensible” since it “engaged in a massive, continuous, near-half-century scheme to defraud” smokers (again, looking at it in the light most favorable to the plaintiff).\textsuperscript{150} The U.S. Supreme Court granted certiorari.

The review was limited to the following questions:
(1) Whether, in reviewing a jury’s award of punitive damages, an appellate court’s conclusion that a defendant’s conduct was highly reprehensible and analogous to crime can “override” the constitutional requirement that punitive damages must be reasonably related to the plaintiff’s harm; (2) Whether due process permits a jury to punish a defendant for the effects of its

\textsuperscript{143} Williams v. Phillip Morris, 340 Or. 35 (2006).
\textsuperscript{144} \textit{Id.} at 44.
\textsuperscript{145} See Williams v. Phillip Morris, 48 P. 3d 824 (Or. Ct. App. 2002).
\textsuperscript{146} \textit{Id.} at 828.
\textsuperscript{147} See Phillip Morris USA, Inc. v. Williams, 124 S.Ct. 56 (2003).
\textsuperscript{149} Williams v. Phillip Morris, 127 P.3d 1165 (Or. 2006).
\textsuperscript{150} \textit{Id.}
conduct on non-parties.\textsuperscript{151}

Before discussing the ultimate decision of the Court, let us take a
look at what various economic theories has to say about the case.
Arguably, in this case the large punitive damages award was due either
to the wealth of the tobacco company or the amount of its profits from
the cigarette business. If so, this may correspond to the gain-elimination,
but not to the loss-internalization, ideology. Notably, Professors
Polinsky and Shavell and Professor Hylton participated in amicus curiae
briefs, filed on opposing sides. Unsurprisingly, the Polinsky and Shavell
brief is in support of the petitioner (Philip Morris), while Hylton’s brief
is in support of the respondent (Williams).

In their brief Polinsky and Shavell are concentrated on demonstrating
that corporate wealth should play no role in determining the amount of
punitive damages.\textsuperscript{152} Although not completely disagreeing with Polinsky
and Shavell on this particular point, Hylton’s brief argued that, because
the Oregon Supreme Court did not rely on Philip Morris’s wealth, there
was not “any reason in law and economics theory and practice to
overturn the decision below.”\textsuperscript{153} The brief provides an excellent review
of the economic theory of deterrence and punitive damages. In
particular, it indicates that under both existing economic approaches the
existence of a tortfeasor’s other victims should be taken into account.\textsuperscript{154}

The case seemed to be a perfect opportunity for applying economic
to principles of damages calculation. In fact, Polinsky and Shavell
considered a hypothetical, strikingly similar to this case, in their
landmark article, concluding that compensatory damages should be
multiplied where “the victim may have difficulty determining that the
harm was the result of some party’s act – as opposed to simply being the
result of nature, of bad luck.”\textsuperscript{155} On the other hand, under the Hylton
approach, where the tobacco company’s conduct is reprehensible, the
proper measure of punitive damages is the company’s illicit profit or

\textsuperscript{151} Phillip Morris USA v. Williams, 126 S. Ct. 2329, see pet. writ cert. at i (Mar.
2006) 2006 WL 849860 (stating the questions presented).
\textsuperscript{152} Brief for Phillip Morris USA et al as Amici curiae Supporting Petitioners, Phillip
\textsuperscript{153} Brief for Williams et al as Amici curiae Supporting Respondent, Phillip Morris
USA v. Williams, 127 S. Ct. 1057 (2006) (No. 05-1256) [hereinafter Brief Supporting
Respondent].
\textsuperscript{154} Id.
\textsuperscript{155} See id. at 17 (quoting Polinsky & Shavell, supra note 8, at 887).
gain.\textsuperscript{156}

Given the comprehensive analysis of the case provided by both schools, the application of this article’s approach is rather straightforward. First of all, the award should include the compensatory part, augmented to reflect the existence of other Oregonians who were similarly harmed by the company but failed to sue. A conservative estimation of their number, offered by the Oregon Court of Appeals, is one hundred.\textsuperscript{157} This suggests $82 million or perhaps $52 million, taking into consideration the statutory cap on non-economic compensatory damages and estimation of quasi-punitive damages. This does not depend on whether the company’s conduct was reprehensible. Further, as discussed above, and consistent with the Polinsky and Shavell’s analysis, the wealth of Philip Morris should not be taken into account when calculating punitive damages. However, consistent with Hylton’s analysis, where the conduct is reprehensible, it is proper to take into account the illicit gains of the company. The hard task is to determine what part of the tobacco’s company gains is related to the fraud in question (received from deceived smokers), and what part can be said to be earned properly (received from risk-assuming smokers). Apparently, the number of deceived smokers is much more than the number of seriously harmed smokers.\textsuperscript{158} Although the materials of the case do not provide the needed estimation, it would be straightforward to design a survey for determining the number of deceived smokers. Then, the applicable penalty, truly punitive damages, could be calculated as the gains received from those smokers; only Oregonian smokers should be taken into account by an Oregon court.\textsuperscript{159} The amount of truly punitive damages is therefore a certain percentage of the company’s profits received in Oregon; this may appear to be comparable with the amount of quasi-punitive damages, or even dwarf it, given the domestic tobacco operating income of the company is in the billions of dollars annually.\textsuperscript{160}

\textsuperscript{156} \textit{Id.} at 18.


\textsuperscript{158} \textit{Cf. id.} at 55-56 (taking into consideration both Oregonians actually injured by the defendant and a broader class of Oregonians put at risk of injury by the defendant).

\textsuperscript{159} \textit{But see} BMW of N. Am. v. Gore, 517 U.S. 559 at 574, n.21 (while the number of sales in other states should not be used as a multiplier in computing the amount of the punitive sanction, it still may be relevant to the determination of the degree of reprehensibility of the defendant's conduct).

\textsuperscript{160} \textit{See} Altria Group, Inc., 2005 \textit{Annual Report},
An alternative way to calculate the illicit gain is multiplying the income received from an average smoker by the number of deceived smokers. Williams, in particular, bought from the company more than 40,000 packs of Marlboro's over the period of forty years. Multiplied by, at least, thousands of deceived Oregonian smokers, this could result in a much larger award than in the real case.

Under the approach of this article, to optimally influence the defendant and others like it, quasi-punitive and truly punitive damages should be charged together. Of them, truly punitive damages, imposed for reprehensible behavior, depend upon the illicit gains of the defendant rather than the harm inflicted. Highly reprehensible behavior often means high illicit gains. This suggests an answer to the first question presented to the U.S. Supreme Court: where the conduct of the defendant is highly reprehensible, punitive damages lose their relation to the harm inflicted.

This still leaves open several problems. First, a due process issue: charging this large amount of punitive damages can be devastating for the tobacco company, whereas the standard of proof in a civil case is relatively low. As mentioned above, this may require a reduction of the theoretical punitive damages amount.

Further, under the approach of this article, awarding the whole amount of punitive damages to the plaintiff is suboptimal. In fact, to achieve optimality, the whole amount of truly punitive damages should be kept by the state, except for a share of the plaintiff's litigation expenses proportional to the state's award and fair compensation for the plaintiff's personal effort. Awarding more to the plaintiff is a sure way of motivating both cigarette consumption and frivolous tobacco lawsuits. Without a finely tuned split-award statute the jury must choose between under-motivating the defendant and over-motivating the plaintiff; no award is optimal in either sense.

Finally, there is a problem of double imposition: if the same amount of punitive damages is charged in successive suits, the defendant will be repeatedly punished for the same conduct. Possible solutions, offered by Hylton, include informing the jury about earlier awards or authorizing the court to provide offsets for successive awards.\textsuperscript{161}

\textsuperscript{161} Brief Supporting Respondent, \textit{supra} note 153, at 25.
In short, the theoretical amount of truly punitive damages may require some reduction on due process and consumer-motivation related grounds. As for the actual award in this case, in light of the above analysis it looks like a very conservative estimation, only slightly, if any, exceeding the amount of compensatory and quasi-punitive damages.

As for the second question presented to the Court, since apparently there was no serious dispute between the two economic schools,\(^{162}\) I can only reiterate that the whole purpose of punitive damages is to account for the effect of the defendant's conduct on non-parties.

However convincing economic arguments might sound, the U.S. Supreme Court has chosen a completely different approach, based on due-process considerations. According to the Court, a punitive damages award based upon a jury's desire to punish a defendant for harming persons who are not before the court would amount to a taking of property from the defendant without due process.\(^{163}\) Accordingly, such an award is not allowed. This prohibition to punish for the harm inflicted to non-parties is a novel, not previously known, provision of constitutional law.\(^{164}\)

The reasoning offered by the Court is as follows. First, "a defendant threatened with punishment for such injury has no opportunity to defend against the charge."\(^{165}\) Second, "to permit punishment for injuring a nonparty victim would add a near standardless dimension to the punitive damages equation."\(^{166}\) Finally, the Court found "no authority supporting the use of punitive damages awards for the purpose of punishing a defendant for harming others."\(^{167}\) However, the Court agreed that a plaintiff may show harm to others in order to demonstrate reprehensibility.\(^{168}\)

Since reprehensibility is determinative for punitive damages, the last statement is in considerable tension with the Court's main holding. Indeed, if harm to others determines reprehensibility, and reprehensibility determines punishment, then punishment is based upon

---

162. See id. at 15.
163. Williams, 127 S. Ct. at 1060.
164. Id. at 1065 ("We did not previously hold explicitly that a jury may not punish for the harm caused others. But we do so hold now.").
165. Id. at 1063.
166. Id.
167. Id.
168. Id. at 1064.
harm to others, which is prohibited. This contradiction is emphasized by Justice Stevens in a dissenting opinion: “When a jury increases a punitive damages award because injuries to third parties enhanced the reprehensibility of the defendant’s conduct, the jury is by definition punishing the defendant—directly—for third-party harm.” 169 Unlike the majority, Justice Stevens sees “no reason why an interest in punishing a wrongdoer for harming persons who are not before the court should not be taken into consideration when assessing the appropriate sanction for reprehensible conduct.” 170 According to Justice Stevens, “punitive damages are a sanction for the public harm the defendant’s conduct has caused or threatened.” 171 (This formula is of course the essence of the economic approach discussed throughout this article.) As for the majority’s due-process-related apprehensions, “[t]here is little difference between the justification for a criminal sanction, such as a fine or a term of imprisonment, and an award of punitive damages.” 172

Justice Thomas went even further, opining that “the Constitution does not constrain the size of punitive damages awards” at all. 173 Justice Ginsburg, with whom Justice Scalia and Justice Thomas joined, dissented on procedural grounds, finding no reason to vacate the Oregon Supreme Court’s judgment when the Court agrees that a jury is allowed to consider the extent of harm suffered by others as a measure of reprehensibility. 174

Noteworthy, the Court declined to elaborate its quantitative Campbell guidelines, refusing to decide whether the award was constitutionally “grossly excessive.” 175 In the end, the Court vacated the Oregon Supreme Court’s judgment and remanded the case for further proceedings. 176

To summarize, Williams considerably undermined all economic justifications of punitive damages, holding that a jury is not allowed to “punish” a defendant for the harm inflicted to nonparties. The essence of most economic approaches to punitive damages is that such damages are

169. ld. at 1067 (Stevens, J., dissenting).
170. ld. at 1066 (Stevens, J., dissenting) (citation and internal quotation marks omitted).
171. ld.
172. ld.
173. ld. at 1067 (Thomas, J., dissenting) (internal quotation marks omitted).
174. ld. at 1068 (Ginsburg, J., dissenting).
175. ld. at 1065.
176. ld.
a measure of public harm, that is, harm to nonparties. However, the Court allowed to "consider" the harm inflicted to nonparties when assessing reprehensibility; and reprehensibility is determinative for punishment. This leaves economically-minded scholars with some hope that further elaboration of these visibly contradictory novel provisions of constitutional law may appear to be less than completely inconsistent with the existing economic theories of punitive damages.

XII. CONCLUSION

Economic theories of punitive damages include gain-elimination approach and loss-internalization approach. This article offers a theory that has a purpose of combining benefits of both approaches. Ultimately, it interprets punitive damages as a social harm measure.

In terms of proper deterrence, the economically optimal amount of damages consists of two components. The first is the amount of primary losses inflicted to the victim. The second is the amount of secondary losses inflicted to society as a whole. The best available estimation for the second component is normally the amount of gain attributable to the deviation of the injurer from a socially approved standard of reprehensibility. In economic terms, the behavior is reprehensible if it results in excessive secondary social losses, as compared to the injurer’s benefit. If the behavior is not reprehensible, the second component of the optimal damages is zero.

To reflect the non-unity probabilities of holding the injurer liable, both components should be inflated by their respective multipliers. The first component is divided by the probability of holding an injurer liable where a victim is injured. The second component is divided by the probability of holding the actor liable where his behavior is reprehensible, or, equivalently, divided both by the probability of injury where a socially approved standard is breached and the probability of holding an injurer liable where a victim is injured.

The overcompensatory part of the first component may be called supercompensatory or quasi-punitive damages, since it is not related to deviation from any standard; the second component may be called truly punitive damages. Both contribute to the overcompensatory part of the total optimal damages amount. If a separate regulation is adopted with a view to mitigate secondary social harms, it ought to preempt truly punitive damages but not supercompensatory damages.
In terms of proper motivation of consumers, the economically optimal award is equal to the sum of the compensatory and supercompensatory damages. The whole amount of truly punitive damages should be awarded to the state for achieving optimality. Taking into consideration litigation costs, the plaintiff should receive a certain additional compensation from the amount of truly punitive damages awarded to the state. First, the state should bear a part of the plaintiff’s litigation expenses in proportion to the state’s share in the award. Second, the plaintiff should be fairly compensated for her personal efforts.

While the proposed optimal-charge formula is in relatively good correspondence with the existing law, the proposed split-award and share-cost rules would require a radical statutory reform for their implementation.

The recent U.S. Supreme Court decision in Williams considerably undermined all economic justifications of punitive damages, holding that a jury can not punish for nonparty harm. However, since the Court allowed a jury to “consider” such nonparty harm when awarding punitive damages, the decision is not completely fatal for the economic theory of punitive damages.