Maryland Law Review

Volume 57 | Issue 1 Article 6

Evidence, Economics, and Ethics: What Information Should Jurors Be Given to Determine the Amount of a Punitive Damage Award?

David Crump

EVIDENCE, ECONOMICS, AND ETHICS: WHAT INFORMATION SHOULD JURORS BE GIVEN TO DETERMINE THE AMOUNT OF A PUNITIVE-DAMAGE AWARD?

DAVID CRUMP*

---

* Newell H. Blakely Professor of Law, University of Houston. A.B., Harvard College; J.D., University of Texas.

With the customary disclaimer that all deficiencies remain the responsibility of the author, I acknowledge with thanks that I am indebted to my colleagues, Professors Seth Chandler and Bruce Hay, for valuable comments on an earlier draft of this Article.

INTRODUCTION

Punitive damages are an uncertain battleground today. In the legislatures, tort reform associations joust with plaintiffs' lawyers over a dizzying variety of proposals for change.[1] In the courts, the cases feature fact patterns ranging from overheated coffee[2] to embezzlement,[3] and the legal issues are as diverse as constitutional due process[4] and concerns of federalism.[5] As for the public, it seems ready,

---

1. *See infra* notes 9-10, 12-17, 26-27 and accompanying text.

2. *See Damage Award Cut in McDonald's Case*, HOUS. CHRON., Sept. 15, 1994, at A4, *available in* 1994 WL 4215906 (reporting a jury verdict of $2.9 million to a woman who suffered third-degree burns from spillage of extremely hot coffee purchased from McDonald's, as well as the judge's remittitur of all but $480,000, which was accompanied by the judge's statement that McDonald's behavior had been "willful, wanton, reckless and . . . callous"). The case has prompted a significant volume of writing. *See, e.g.*, Bruce Feldthusen, *Punitive Damages in Canada: Can the Coffee Ever Be Too Hot?*, 17 LOY. L.A. INT'L & COMP. L.J. 793, 794-97, 806 (1995) (exploring the Canadian punitive-damage system and suggesting that "[t]he Canadian regime best reflects the theoretical underpinning of tort as understood in our [American] culture"); Susanah Mead, *Punitive Damages and the Spill Felt Round the World: A U.S. Perspective*, 17 LOY. L.A. INT'L & COMP. L.J. 829, 829-57, 860 (1995) (discussing the $3.5 million punitive damages in a hypothetical "coffee spill" case and concluding that, "[i]n spite of the efforts of tort reformers and business interests, . . . it is unlikely that the award of punitive damages in product liability will ever be eliminated completely").

3. *See* Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 6-8 (1991) (affirming a $1.04 million punitive-damage award against an insurance company for embezzlement committed by its agent).

4. *See id.* at 19-24 (discussing constitutional due process); *see also id.* at 24-40 (Scalia, J., concurring in the judgment) (detailing the history of due process).

5. *See* BMW of N. Am., Inc. v. Gore, 116 S. Ct. 1589, 1597-98 (1996) (raising concerns of federalism).

perhaps even eager, to impose whopping penalties on deep-pocket tortfeasors,[6] but at the same time, cherished horror stories make many people skeptical about the misuse of such a potent tort weapon.[7]

It is not surprising, therefore, that the jury still is out on the question whether all of the "reforms" in punitive damages really are reforms. Some of the changes seem quite unlikely to produce improvements.[8] While one state finagles with procedural concerns such as separate trials[9] or enhanced burdens of proof,[10] others rewrite jury instructions[11] or redefine the rules of evidentiary relevance.[12]

---

6. *See infra* note 268 and accompanying text.

7. *See* Dick Thornburgh, *America's Civil Justice Dilemma: The Prospects for Reform*, 55 MD. L. REV. 1074, 1075, 1085-86 (1996) (proposing caps, proof standards, and an enhanced burden of proof, and citing examples of "runaway" punitive awards that make people skeptical of the misuse of these awards).

8. *See infra* Part III.B.1-2 (critiquing laws providing for net-worth evidence and for certain kinds of caps). Another example is the widespread support for an enhanced burden of proof, requiring jurors to find "clear and convincing evidence" of the conditions authorizing punitive damage. Jan Woodward Fox & Kate McConnico, *Punitive Damages in Texas 1995: Chapter 41 of the Texas Civil Practice & Remedies Code*, 21 T. MARSHALL L. REV. 21, 22-27 (1996). Assuming such a change would make any difference at all, experienced trial lawyers may find reason to suspect that it might work contrary to its proponents' intention by actually *lowering* the burden:

> Though "clear and convincing" is a higher *legal* burden than proving liability by a preponderance of the evidence, the new standard may work in a plaintiff's favor for common-sense reasons. The term "by a preponderance of the evidence" is often explained to the jury as 51% or the majority of the evidence. It could raise in the jurors' minds a mathematical requirement, a ratio, or an amount that must outweigh the opponent's evidence. "Clear and convincing" on the other hand, is soley [sic] ends oriented. The law mentions nothing of a specific amount of evidence required, but rather focuses on the feelings of the jury . . . . The standard thus becomes *subjective* in orientation and *focuses on quality rather than quantity*. Whether this will make any difference at all as a practical matter remains to be seen.

Id. at 27.

9. *E.g.*, GA. CODE ANN. § 51-12-5.1(d) (Supp. 1997) (providing for a bifurcated trial whenever punitive damages have been requested); MINN. STAT. ANN. § 549.20(4) (Supp. 1997) (providing for a bifurcated trial at the request of any of the parties); TEX. CIV. PRAC. & REM. CODE ANN. § 41.009 (West 1997) (stating that a bifurcated trial would be provided only on a motion by the defendant).

10. *E.g.*, ALA. CODE § 6-11-20(a) (1993) (requiring proof by clear and convincing evidence); OHIO REV. CODE ANN. § 2315.21(D)(2) (Anderson Supp. 1996) (same); *see also supra* note 8.

11. *E.g.*, Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 43 (1991) (O'Connor, J., dissenting) (advocating this approach as a constitutional requirement).

12. *E.g.*, TEX. CIV. PRAC. & REM. CODE ANN. § 41.011 (West 1997) (listing six relevant factors to be used in determining exemplary damages); *see also* Fox & McConnico, *supra* note 8, at 26, 35 (analyzing relevant evidence under this Texas statute).

Still other reforms involve caps,[13] ratios,[14] limits,[15] new thresholds of evidentiary sufficiency,[16] or intensified scrutiny by trial and appellate judges.[17] The 1996 *Model Punitive Damages Act* (*Model Act*)[18] adopts several of these features, including an enhanced burden of proof,[19] a tightened definition of gross negligence,[20] a nine-factor checklist for dollar amounts,[21] a requirement for bifurcation of trials in some cases,[22] special review by the trial judge,[23] and a hearing to reduce awards that are "unfairly duplicative."[24]

The picture that emerges is that of a struggle in search of a purpose.[25] For example, Ohio's 1996 reform legislation featured rigorous caps,[26] and so did the federal tort reform bill of

---

13. *E.g.*, GA. CODE ANN. § 51-12-5.1(g) (Supp. 1997) (setting a general cap of $250,000); IND. CODE ANN. § 27-12-14-3 (Michie 1994) (setting a cap of $750,000 for both actual and punitive damages in certain wrongful death cases).

14. *E.g.*, NEV. REV. STAT. ANN. § 42.005 (Michie 1996) (setting a cap of three times the compensatory damages if such damages are more than $100,000, and setting a cap of $300,000 if such damages are less than $100,000).

15. *E.g.*, VA. CODE ANN. § 8.01-38.1 (Michie Supp. 1997) (setting a $350,000 limit on medical malpractice punitive awards).

16. *E.g.*, TEX. CIV. PRAC. & REM. CODE ANN. § 41.003 (West 1997) (replacing gross negligence with higher thresholds).

17. *E.g.*, *id.* § 41.013 (requiring justification "with specificity" when a punitive award is upheld or overturned); Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 19-20 (1991) (approving Alabama's approach of intensified judicial scrutiny).

18. MODEL PUNITIVE DAMAGES ACT (1996); *see also* Richard C. Reuben, *This Model Sports No Caps: Proposal by Uniform Law Commissioners Tightens Punitive Procedures*, A.B.A. J., Oct. 1996, at 22, 22 (analyzing the *Model Act* and summarizing different commentators' positions on the punitive-damages debate).

19. MODEL PUNITIVE DAMAGES ACT § 5.

20. *Id.* § 5 cmt.

21. *Id.* § 7.

22. *Id.* § 11.

23. *Id.* § 8.

24. *Id.* § 10.

25. *See* Bruce Chapman & Michael Trebilcock, *Punitive Damages: Divergence in Search of a Rationale*, 40 ALA. L. REV. 741, 761-826 (1989) (explaining rationales based on compensation, retribution, and deterrence); Andrea A. Curcio, *Painful Publicity: An Alternative Punitive Damage Sanction*, 45 DEPAUL L. REV. 341, 346-51, 358-93 (1996) (describing rationales and advocating publicity as an alternative to monetary punitive damages); David D. Haddock et al., *An Ordinary Economic Rationale for Extraordinary Legal Sanctions*, 78 CAL. L. REV. 1, 8-50 (1990) (distinguishing the authors' proposed bargaining-based model from illicit-gains and court-error models); Sandra N. Hurd & Frances E. Zollers, *State Punitive Damages Statutes: A Proposed Alternative*, 20 J. LEGIS. 191, 195-203 (1994) (surveying a variety of statutes and critiquing reform as "misguided"); David F. Partlett, *Punitive Damages: Legal Hot Zones*, 56 LA. L. REV. 781, 792-802 (1996) (describing the "Search for a Rationale"); W. Lee Pittman & Bert S. Nettles, *Debate: What Is the Role or Function of Punitive Damages?*, 24 CUMB. L. REV. 453, 455-77 (1994) (debating differing rationales).

26. *See* OHIO REV. CODE ANN. § 2315.21(A)-(D) (Anderson 1996). The 1996 Ohio Act provides a limit of the lesser of $100,000 or three times compensatory damages for businesses with fewer than twenty-five employees, and the lesser of $250,000 or three times

1995.[27] The *Model Act*, however, has no caps at all,[28] and the drafters' commentary refers to caps as "arbitrary."[29] "There was no real controversy on that, interestingly enough," reported one of the advisors to the drafters.[30] But the *Model Act* relies on other factors, such as jury consideration of the defendant's wealth,[31] which other commentators have rejected.[32] The hefty debate means that continuing change is likely.[33] In summary, courts and legislatures lack a coherent vision of the function they want punitive damages to serve, and therefore, they have no clear goal toward which they can target these new standards.

---

compensatory damages for those with more than twenty-five employees. *Id.*; *accord Ohio Enacts Landmark Tort Reform Bill*, THE ADVOCATE (Tex. Civ. Just. League), Dec. 1996, at 6 (explaining the Ohio Act).

27. *See* H.R. 10, 104th Cong. § 103(c) (1995) (proposing limiting punitive damages in a product liability action to three times economic injury or $250,000, whichever is greater); *accord To Reform the Federal Civil Justice System; To Reform Product Liability Law: Hearing on H.R. 10 Before the House Comm. on the Judiciary*, 104th Cong. 66 (1995) (statement of Richard K. Willard, Esq., Steptoe & Johnson) (explaining the cap); Thornburgh, *supra* note 7, at 1085-86 (same). The President vetoed the bill, calling the cap "arbitrary." Nancy Mathis, *Clinton Vetoes Bill Limiting Suits over Defective Products*, HOUS. CHRON., May 3, 1996, at A2, *available in* 1996 WL 5596310.

28. *See supra* note 18 and authorities therein cited.

29. MODEL PUNITIVE DAMAGES ACT prefatory note para. 20 (1996). President Clinton made the same statement about the proposed cap in the federal bill, which he vetoed. *See supra* note 27.

30. *See* Reuben, *supra* note 18, at 22.

31. MODEL PUNITIVE DAMAGES ACT § 7(a)(4) (listing as a factor "the defendant's present and future financial condition and the effect of an award on each condition").

32. *E.g.*, Chapman & Trebilcock, *supra* note 25, at 777-78, 801-04, 822-25 (concluding that wealth should not be a factor under either the compensatory, retributive, or deterrence rationales). This Article is among those rejecting the wealth factor. *See infra* Part III.B.1. Ironically, Alabama formerly excluded net-worth evidence as unduly prejudicial but changed this approach after the Supreme Court in *Haslip* upheld, but exposed flaws in, Alabama's system, which did not provide jury guidance on the factors considered by judges. Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 19-24 (1991); Life Ins. Co. v. Johnson, 684 So. 2d 685, 688-702 (Ala.), *vacated*, 117 S. Ct. 288 (1996).

33. Thus, for example, a spokesperson for the Association of Trial Lawyers of America (ATLA) criticized the *Model Act* on the ground that higher standards were not needed, because "recent studies . . . indicate that punitive damages are extremely rare and tend not to be that high." Reuben, *supra* note 18, at 22. But a leading proponent of tort reform criticized the same *Model Act* by arguing that the substitution of "a bunch of fluffy factors" for a clearly defined cap was a denial of "[e]lementary due process." *Id. Compare* Thornburgh, *supra* note 7, at 1085 (asserting that punitive damages are "now almost routinely claimed in tort litigation") *and* STEVEN HAYWARD, THE ROLE OF PUNITIVE DAMAGES IN CIVIL LITIGATION: NEW EVIDENCE FROM LAWSUIT FILINGS 1, 4, 7-8 (Pacific Research Inst. for Pub. Policy Briefing, Feb. 1996) (finding that "[p]unitive damages are demanded in 27 percent of all cases where they are conceivably recoverable," and concluding that "[l]awsuits that include punitive damage demands take one-third longer to resolve than suits without these demands") *with* Marc Galanter, *Real World Torts: An Antidote to Anecdote*, 55 MD. L. REV. 1093, 1126-40 (1996) (analyzing surveys to conclude that punitive awards occur with "low frequency" in personal injury and product liability cases but concluding from anecdotal evidence that "the threat of punitive damages can be a significant factor in settlement").

This Article attempts to find a unifying principle by undertaking an economic analysis of tort remedies and applying it to the evidentiary principles that determine punitive damages. The first Part of the Article sets up an economic model of both compensatory and punitive awards. Briefly put, this analysis shows that the core function of punitive damages, at least where economically motivated actors are concerned,[34] is to fill a gap in the deterrence of accident costs in a market economy. This gap comes about because the externalization of accident costs is not fully addressed by compensatory damages, owing to failures of detection, prosecution, proof, or remedy. Economic analysis also shows that there is a theoretically appropriate level of deterrence, so that the tort system becomes dysfunctional if our legal rules lead to the imposition of either too much punitive-damage liability or too little.

Economic calculations, however, cannot properly form the exclusive basis for analysis of punitive damages. Information deficiencies sometimes mean that types of evidence theoretically correlated with economic efficiency are inferior to other types of evidence that ostensibly seem less accurate. Furthermore, the consequentialist approach that underlies economic reasoning may depart so severely from moral considerations that the result is unacceptable in deontological terms. The second Part of the Article therefore deals with these criticisms and attempts a synthesis of economic analysis with the differing concerns that the criticisms raise.

The third Part of the Article examines how the correct amount of punitive damages is to be determined in a given case. Tort reform has brought forth an array of different evidence rules, some of which, when evaluated by an economic model, can be exposed as ill-suited to the purpose. For example, although many jurisdictions[35] and the *Model Act*[36] admit evidence about the defendant's net worth, an economic analysis shows that this evidence actually has little relevance to the proper function of punitive damages. Nor is net-worth evidence

---

34. The analysis in this Article applies principally to economically motivated actors. A different problem arises in the event that injurers obtain socially illicit utility from the infliction of loss itself, as in the case, say, of hate crimes, or the liability imposed by the civil jury upon O.J. Simpson, or "[w]here, for instance, a man kicks his neighbor's dog because he positively enjoys the unhappiness his neighbor will feel over it." STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW 147 (1987). In this event, the analysis is fundamentally altered, because society does not value the basic activity in which the actor is engaged, or in more technical terms, "we may not want to count the man's utility as an addition to social welfare." *Id.* In such a case, liability may be more efficient if it exceeds the sum of all losses by an amount great enough to remove all utility from the actor's conduct. *Id.*

35. *See infra* note 248 and accompanying text.

36. *See supra* note 31 and accompanying text.

defensible in deontological terms. Indeed, this Article concludes that net-worth evidence tends to divert the jury from more appropriate considerations to such a degree that it should be excluded, pursuant to Federal Rule of Evidence 403,[37] as excessively prejudicial or misleading.

The goal, instead, should be the formulation of evidence rules that guide juries toward computation of punitive awards that relate to the deterrence gap, referred to above, and fall within a proportional system of moral blameworthiness. The difficulty is that measuring this gap is uncertain and complex in concrete cases—so much so, that an insistence upon purity in economic theory would conflict with considerations of clarity, predictability, and reasonable trial duration. The final Part of the Article therefore contains the author's conclusions, which include the concept that evidentiary standards should be crafted so that punitive damages will serve their economic and moral purposes, with the caution that perfect congruence between theory and practice will be unattainable if the governing law is to allow for efficient jury trials. Appropriate standards can be found in cases that consider such factors as profitability of the wrongful conduct, actual and expected harm, and "reprehensibility." Despite the apparent crudity of these concepts, they are likely to provide better jury guidance and more practical evidence than current standards or other theoretical constructs.

Before the analysis, it is necessary to distinguish several ancillary issues that this Article does not consider in depth. First, it is possible to argue that punitive damages do not and indeed cannot serve their claimed economic function.[38] The argument rests upon theoretical assumptions such as perfect information as well as accurate and full recoveries of all damages by all injured persons, and is therefore

---

37. FED. R. EVID. 403.

38. The argument posits that at any care level below the optimum due-care level, the firm experiences incentives that force it to increase care. *See infra* fig.6. According to Figure 6, the economic pressures on any firm operating with a combination of resources that is to the left of point *O* (the due-care level) will force it toward point *O*, without the addition of punitive damages. A punitive award, by this argument, is superfluous. *Cf.* Alan Calnan, *Ending the Punitive Damage Debate*, 45 DEPAUL L. REV. 101, 109-22 (1995) (advocating the abolition of punitive damages in favor of restorative remedies for dignitary injuries); Robert D. Cooter, *Economic Analysis of Punitive Damages*, 56 S. CAL. L. REV. 79, 79 (1982) (arguing that "[i]f fault is unintentional, then imposing punitive damages . . . is both unnecessary for deterrence and undeserved as punishment," although punitive damages may be justified for intentional wrongdoers or knowingly negligent conduct known to be likely to injure). This reasoning arguably is incomplete, however. *See infra* Part I.B.

treated here only in passing.[39] Second, this Article is concerned primarily with behavior based upon economic exchange in recognizable markets and therefore does not necessarily fit the actor who derives perverse utility from breaking the law or gratuitously injuring others.[40] Civil suits based upon hate crimes, vandalism, or the conduct for which O.J. Simpson was found liable, therefore, would require adaptation of this theory, although it probably could be fitted to the purpose.[41]

Finally, there is the question whether an analysis of punitive damages really even matters. Some commentators suggest that punitive awards are so rare that their effect, economic or otherwise, is of little significance. Others disagree.[42] In any event, it appears that interjurisdictional differences are such that the issue is a cognizable problem in some locations,[43] and furthermore, that the punitive factor undoubtedly affects some of the vast majority of cases that settle.[44] This Article therefore assumes that the issue is not without significance.

---

39. *See, e.g.*, John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965, 966, 994-97 (1984) (demonstrating that if all relevant information is not perfectly known to actors, classical economic conclusions, presumably including those underlying punitive damages, may not hold, but asserting that damage multipliers may be justified in some cases, if a degree of uncertainty is assumed); *see also infra* Part II.A.

40. *See supra* note 34.

41. The adaptation would include a sufficient increase in the damage award to equal the perverse enjoyment the actor experiences, thus eliminating the illicit utility. *See supra* note 34.

42. *See supra* note 33 and authorities therein cited. *Compare* A. Mitchell Polinsky, *Are Punitive Damages Really Insignificant, Predictable, and Rational? A Comment on Eisenberg et al.*, 26 J. LEGAL STUD. 663 (1997) (critiquing conclusions about the rationality of punitive damages from empirical results while assuming results are correct) *with* Theodore Eisenberg et al., *The Predictability of Punitive Damages*, 26 J. LEGAL STUD. 623 (1997) (reporting an ambitious empirical study of punitive-damages judgments and attempting to draw conclusions, which Polinsky's article critiques, about the rarity and size of punitive damages).

43. *See* Partlett, *supra* note 25, at 816-23 (summarizing punitive-damage studies and showing that the problem varies geographically and topically); George L. Priest, *Punitive Damages Reform: The Case of Alabama*, 56 LA. L. REV. 825, 825 (1996) (describing punitive awards of such frequency and magnitude as to "dramatically affect the entire civil dispute process" in that state). *See generally* STEPHEN DANIELS & JOANNE MARTIN, CIVIL JURIES AND THE POLITICS OF REFORM (1995) (presenting evidence against the existence of a nationwide problem but showing jurisdictional variations); HAYWARD, *supra* note 33 (demonstrating demands for punitive damages in 27% of cases where conceivably recoverable and noting extended length of suits demanding punitives).

44. *See supra* note 33 and authorities therein cited.

I. THE ECONOMIC FUNCTION OF PUNITIVE DAMAGES

One possible model considers punitive damages as compensation.[45] The theory is that injured plaintiffs do not recover certain losses, such as the amounts they pay for attorneys' fees or the entrepreneurship involved in bringing suit, and that punitive damages supply a corrective to the omission.[46] Alternatively, punitive damages are sometimes considered as a substitute for compensation in those instances, such as defamation cases, where injury is apparent but difficult to value.[47] This concept is better termed "presumed" damages, in that it permits rough estimation of what really are compensatory rather than punitive damages.[48]

In any event, the compensation rationale for punitive damages is dubious,[49] because normally it should be more economically efficient to permit recovery of all losses by the best possible estimate, including attorneys' fees if they are deemed items of loss, rather than by imposing a penalty in excess of actual losses.[50] This concept arguably is carried out by recent statutes that divert part of punitive awards to the state.[51] The better rationale is not compensation but deterrence. The economist would explain that what is important is that the defendant pays an appropriate level of punitive damages, not that the plaintiff receives them.

---

45. *See* Chapman & Trebilcock, *supra* note 25, at 761-78 (analyzing the compensation rationale).

46. *Id.*

47. *Id.* at 763 (analyzing punitive damages in the context of dignitary torts, which the authors say might better be termed "aggravated" damages because they are not really punitive).

48. *See infra* notes 74-75 and accompanying text.

49. *See* Chapman & Trebilcock, *supra* note 25, at 763 (critiquing the rationale on the ground that it fails to explain why a higher degree of fault should be required for punitive damages than that required for compensation for actual losses, except in cases such as malicious injury to dignity, where the compensation is for "aggravated," rather than punitive, damages).

50. *See id.* (suggesting that the result is a "windfall" rather than compensation); *see also infra* Part I.B.2. (describing the economic inefficiency of excessive punitive damages).

51. *E.g.*, FLA. STAT. ANN. § 768.73(2)(a)-(b) (West 1997) (requiring that 35% of the punitive-damages award be paid to a medical assistance trust fund, if the action was based on personal injury or wrongful death, or to state general revenue, if it is another type of action); MO. REV. STAT. § 537.675 (1997) (requiring, after the deduction of attorneys' fees and expenses, that 50% of the punitive award be contributed to a tort victim compensation fund); OR. REV. STAT. § 18.540 (1988) (requiring that one-half of the punitive-damages award be paid to the criminal injuries compensation fund).

A. *A Summary: Punitive Damages and Economic Theory*

To an economist, the most important aspect of the price system is the way in which it works to allocate resources.[52] Money has no intrinsic worth,[53] whether it comes in the form of payments for purchases, compensatory damages, or for that matter, punitive damages. Instead, the costs of various resources that a firm uses to provide a product or service are an inducement to economic efficiency. The price system forces the firm to produce in the most efficient manner possible, and it also induces the firm to provide those products or services that the consumers of a society want.[54] For the most part, the market functions in this manner independently of the tort system, but tort remedies cannot be understood except in the context of the market system.[55]

1. Compensatory Damages as Economic Deterrents.—Against this background, the function of compensatory damages can be seen as that of addressing market imperfections.[56] The theory of market efficiency breaks down if the firm can avoid costs by "externalizing" them.[57] If cheaper means of production result in environmental pollution or in an unacceptably large proportion of accidents and injuries, for example, the price system counterproductively forces firms to chose this privately-desirable-but-socially-dysfunctional method of production. The economist would see the imposition of damage liability as a corrective to these externalities.[58] The desirable level of damage liability would be reached when it precisely balanced the consumers' desire for readily available products against the desire to avoid the

---

52. *See* PAUL A. SAMUELSON & WILLIAM D. NORDHAUS, ECONOMICS 741-54 (13th ed. 1989) (discussing markets and economic efficiency); MARK SEIDENFELD, MICROECONOMIC PREDICATES TO LAW AND ECONOMICS 21-60 (1996) (discussing producer decisionmaking, market dynamics, and efficiency); *see also* A. MITCHELL POLINSKY & STEVEN SHAVELL, PUNITIVE DAMAGES: AN ECONOMIC ANALYSIS pt. II (John M. Olin Cent. for Law, Econ. and Bus. Discussion Paper No. 212, 1997) (forthcoming 111 HARV. L. REV. (1998)) (presenting a general economic theory and applying it to punitive damages).

53. SAMUELSON & NORDHAUS, *supra* note 52, at 226-27.

54. *See supra* note 52 and authorities therein cited.

55. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 773-75 (discussing market imperfections and regulatory or liability responses to them); *cf.* SEIDENFELD, *supra* note 52, at 61 (discussing market imperfections).

56. *See supra* note 55 and authorities therein cited.

57. "Externalizing" costs refers to the reduction of one's own production costs by methods that impose costs or losses upon other firms or individuals. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 972 (defining "externality" as "[a]n activity that affects others for better or worse, without those others paying or being compensated for the activity"); *see also id.* at 773-75 (discussing market imperfections and regulatory responses to them).

58. *See supra* note 55 and authorities therein cited.

harmful consequences, such as pollution or injuries, that result from cheap production.[59] In this view, deviations from the optimal level of damage liability are undesirable, irrespective of whether they are upward or downward. Too much damage liability results in an undesirable suppression of the production of goods or services; too little results in undesirably high levels of external effects.[60]

Usually, the tort system provides the requisite deterrence through compensation to the plaintiff.[61] This compensatory orientation, however, obscures the economic function of damage remedies. Although compensation is the measure the courts have adopted, and it obviously is the most important aspect of the remedy to individual plaintiffs, the more significant aspect of compensatory damages to an economist is the deterrence of the defendant that results from the defendant's having to add the amount of potential damage verdicts to its production costs.[62] If these damages are correctly computed, they precisely measure the degree to which society values the corrective, or deterrent, that it chooses to administer.[63] The theory requires that the compensation be complete, including all of the costs or losses suffered by the plaintiff, whether explicitly monetary or not, and whether readily measurable or not.[64] If this level is reached, the tort system provides exactly the level of deterrence that is socially desirable.[65]

---

59. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 191-92 (3d ed. 1992) (discussing reasons why damages should be equal to, but limited to, the amount of the loss); POLINSKY & SHAVELL, *supra* note 52, at pt. II.A (same).

60. *See* POSNER, *supra* note 59, at 191-92. It has been argued that the direct health-care costs of job-related injuries and illnesses are far greater than has been realized. *See* J. Paul Leigh et al., *Occupational Injury and Illness in the United States: Estimates of Costs, Morbidity, and Mortality*, 157 ARCHIVES INTERNAL MED. 1557, 1557 (1997) (concluding that "[o]ccupational injuries and illnesses are an insufficiently appreciated contributor to the total burden of health care costs in the United States"). Personal losses to the injured workers and economic losses for their employers add to these costs. This consideration arguably supports greater deterrence of accident-producing behavior.

61. *See infra* notes 66-75 and accompanying text.

62. *See* POSNER, *supra* note 59, at 191 (noting that the real point is that the defendant should pay damages equaling the loss, that the remaining issue is whether they should be paid to the injured person rather than the state, and that the reasons for compensating the plaintiff are, first, to maintain an incentive for injured persons to sue, and second, to avoid overspending on precautions by potential victims); *see also* A. Mitchell Polinsky & Yeon-Koo Che, *Decoupling Liability: Optimal Incentives for Care and Litigation*, 22 RAND J. ECON. 562, 563 (1991) (arguing that the amount received by the plaintiff need not and in some instances should not be the same as that paid by the defendant).

63. *Cf.* POSNER, *supra* note 59, at 164 (discussing the importance of weighing the costs and benefits of accident prevention when determining the optimal level of liability).

64. *See id.* at 196-201 (illustrating losses in efficiency that result when damages fail to include the value of some kinds of harm, such as hedonic injuries, and demonstrating the difficulty, yet importance, of correct valuation).

65. *See id.*

Sometimes the courts expressly have analyzed this relationship between the compensation standard and the deterrence function. For example, in *Memphis Community School District v. Stachura*,[66] the Supreme Court reviewed a damage verdict under 42 U.S.C. § 1983 based upon the First Amendment claims of an allegedly wrongfully suspended public school teacher.[67] The district court instructed the jury to award the plaintiff (1) compensatory damages, plus (2) punitive damages, *plus* (3) an amount of *additional* damages based upon the value or importance of the abstract constitutional right that the defendant allegedly violated.[68] Specifically, the district court stated:

> "The precise value you place upon any Constitutional right which you find was denied to Plaintiff is within your discretion. You may wish to consider the importance of the right in our system of government, the role which this right has played in the history of our republic, [and] the significance of the right in the context of the activities which the Plaintiff was engaged in at the time of the violation of the right."[69]

The district court entered a large judgment based upon a jury verdict that included this supplementary element.[70] The court of appeals affirmed.[71] Owing to the fundamental importance of the right at issue, these lower courts evidently concluded that a dollop of extra deterrence was required in addition to compensatory and punitive damages.

The Supreme Court reversed, expressly tying the function of compensatory damages (provided that all losses were fully compensated) to the deterrence purpose:

> Punitive damages aside, damages in tort cases are designed to provide "*compensation* for the injury caused to plaintiff by defendant's breach of duty." To that end, compensatory damages may include not only out-of-pocket loss and other monetary harms, but also such injuries as "impairment of reputation . . . , personal humiliation, and mental

---

66. 477 U.S. 299 (1986).

67. *Id.* at 301-02. Section 1983 creates a private cause of action in favor of individuals who are deprived of their constitutional rights. *See* 42 U.S.C. § 1983 (1994); *accord Stachura*, 477 U.S. at 305-06. The relevant portion of the First Amendment provides: "Congress shall make no law . . . abridging the freedom of speech." U.S. CONST. amend. I; *accord Stachura*, 477 U.S. at 301-02.

68. *Stachura*, 477 U.S. at 305.

69. *Id.* at 303 (alteration in original) (quoting the jury instructions of the district court).

70. *Id.*

71. *Id.*

> anguish and suffering." Deterrence is also an important purpose of this system, but it operates through the mechanism of damages that are *compensatory*—damages grounded in determinations of plaintiffs' actual losses.[72]

The Court refused to allow the challenged instructions to authorize a form of "presumed" damages, a traditional remedy when no realistic measure of damages can be found.[73] Presumed damages, said the Court, "are a *substitute* for ordinary compensatory damages, not a *supplement* for an award that fully compensates the alleged injury."[74] Here, "no rough substitute for compensatory damages was required . . . , since the jury was fully authorized to compensate [plaintiffs] for both monetary and nonmonetary harms caused by [defendant's] conduct."[75]

In *Stachura*, the proper level of deterrence was approximated by compensatory damages, which provided a level of deterrence consistent with society's valuation of avoidance of the harm in question.[76] Excessive deterrence, such as that which would have resulted from the district court's gratuitous addition of a supplement to protect the abstract right in question, would have been dysfunctional. It would have caused school districts to become excessively risk-averse, motivating them to tilt the balance toward retention of incompetent teachers and disruptive students.[77] The Supreme Court's reversal was correct, and its analysis of the function of compensatory damages was sound.

2. Punitive Damages as Gap-Fillers in the System of Economic Deterrence.—The economist would see complete compensatory damages, at least in theory, as fulfilling this function of deterring accident cost externalization.[78] The price system, in this view, induces the firm to produce goods and services efficiently, while the tort reparations system confronts the firm with the precise cost of losses to victims in the form of compensatory damages.[79] But the compensation-based theory is accurate only if *every* tort victim recovers *fully* for *all* losses.[80] Perhaps for this reason, the modern law of remedies avoids the pejo-

---

72. *Id.* at 306-07 (alteration in original) (citations omitted) (quoting 2 F. HARPER ET AL., LAW OF TORTS § 25.1, at 490 (2d ed. 1986) and Gertz v. Robert Welch, Inc., 418 U.S. 323, 350 (1974)).

73. *Id.* at 310.

74. *Id.*

75. *Id.* at 312.

76. *See supra* notes 62-63 and accompanying text.

77. *Cf. supra* note 60 and accompanying text.

78. *See supra* notes 59-60 and accompanying text.

79. *See supra* notes 52-55 and accompanying text.

80. *See supra* notes 61-65 and accompanying text.

rative labeling of uncertain damages as "speculative" or "guesswork," and it tends to accept imperfect modeling as a substitute for exact calculation. As one court of appeals has said, "[C]ompensation for undisputed injury should not be denied merely because the amount of damages cannot be precisely and exactly determined."[81]

Even with this corrective, however, the theory of deterrence through compensation breaks down. All victims still will not always recover their full losses. Transaction costs (such as attorneys' or experts' fees) and proof difficulties are such that not all injured persons will sue for, let alone recover, exactly what they have lost.[82] It is for this reason that punitive damages are useful. The economic function of these damages, according to this argument, is not so much that of "punishing" an individual based upon "wrongful intent" as that of adjusting the level and locus of damage liability to take account of undervaluation of external costs by the tort system through compensatory damages alone.[83]

At the same time, there is a need to limit punitive damages. Economic theory suggests that they can be harmful rather than helpful, if they are imposed in excessive amounts.[84] Therefore, there is a need to restrict their availability to cases in which compensatory damages are an insufficient deterrent and to compute their amounts so that they fit their respective deterrence gaps. The economist would see a threshold requirement of gross negligence, for example, as a means of limiting the availability of punitive damages to cases of underdeterrence.[85] A more economically precise measurement would inquire

---

81. Hawthorne Indus. v. Balfour MacLaine Int'l, Ltd., 676 F.2d 1385, 1388 (11th Cir. 1982) (construing U.C.C. § 2-715 cmt. 4 (1977), which rejects any requirement of mathematical precision and allows "any manner which is reasonable under the circumstances").

82. *See* DAVID CRUMP ET AL., CASES AND MATERIALS ON CIVIL PROCEDURE 1041-42 (2d ed. 1992) (illustrating the difficulty for injured persons to recover losses that are not easy to prove or to value); *cf.* POSNER, *supra* note 59, at 220 (comparing punitive damages and penal sanctions that exceed losses because of factors such as failures of detection and transaction costs); SHAVELL, *supra* note 34, at 146-51 (discussing reasons for imposing liability in excess of loss).

This analysis applies principally to economically motivated actors. When the tortfeasor derives illicit utility from the tort, as in the case of a hate crime, the analysis must be adjusted to eliminate this utility. *See supra* note 34.

83. *See* POLINSKY & SHAVELL, *supra* note 52, at 8-9 ("In summary, *punitive damages should ordinarily be awarded if, but only if, an injurer has a chance of escaping liability for the harm he caused.*"); *see also infra* text accompanying notes 100-102 (describing economic reasoning by the majority and the dissent in a Supreme Court opinion in which the standard of fault is discussed as a means of influencing actors to optimize their prevention of harm rather than in terms of wrongfulness).

84. *See supra* notes 59-60 and accompanying text.

85. *See supra* note 82 and authorities therein cited.

about the deterrent adequacy of actual damages.[86] Thus, the gross negligence threshold is a crude yardstick. Nevertheless, it approximates this function.[87] If a state government were to determine that this standard did not properly calibrate the desired reduction in injuries with the desired level of production of goods and services, the economist might advise shifting to a different standard. For example, the economist might require a showing of intentional injury if punitive damages were too high,[88] or an automatic doubling or trebling of compensatory damages based on simple proof of liability (as in antitrust cases) if they were too low.[89] In addition, various means of measuring the proper amount of punitive damages have evolved in the law, such as requiring proportional relationships to actual damages (or imposing absolute limits, as some states have done).[90]

Again, the courts sometimes have referred expressly to the deterrent function of punitive damages in economic terms. For example, in *Smith v. Wade*,[91] the Supreme Court considered a 42 U.S.C. § 1983 suit against Smith, a prison guard, for having placed in Wade's cell two other inmates who beat and sexually assaulted him.[92] Wade voluntarily sought administrative segregation because of prior assaults, and a vacant cell was available.[93] He alleged that Smith violated the Eighth Amendment's prohibition on cruel and unusual punishment because the guard knew or should have known that assault was likely.[94] The trial court instructed the jury that Smith could be liable for compensatory damages only if he acted with gross negligence, "defined as 'a callous indifference or a thoughtless disregard for the consequences of one's act or failure to act.'"[95] Smith thus could not be

---

86. This is so, because if actual damages provide all the deterrence necessary, there is no economic reason for greater damages. *See supra* notes 61-65 and accompanying text.

87. *Cf.* POSNER, *supra* note 59, at 209-10 (justifying punitive damages in cases of indifference by the tortfeasor because of a probable correlation between this factor and the deterrence gap).

88. *Cf. infra* notes 91-102 and accompanying text (discussing a Supreme Court opinion that approved of a standard for assessing punitive damages that required a showing of "gross negligence" but not of actual intent).

89. *See* POSNER, *supra* note 59, at 315-16 (discussing the computation of antitrust damages).

90. *See supra* notes 9-28 and accompanying text.

91. 461 U.S. 30 (1983).

92. *Id.* at 32, 34-35; *see also supra* note 67 (explaining actions arising under 42 U.S.C. § 1983 (1994)).

93. *Smith*, 461 U.S. at 32.

94. *Id.* The Eighth Amendment provides: "Excessive bail shall not be required, nor excessive fines imposed, *nor cruel and unusual punishments inflicted.*" U.S. CONST. amend. VIII (emphasis added).

95. *Smith*, 461 U.S. at 33 (quoting the trial judge's jury instructions).

liable even for compensatory damages on a finding of only simple negligence.[96] The trial judge also instructed the jury on punitive damages in a way that required no greater culpability than that required for compensatories:

> "[I]f the conduct of one or more of the defendants is shown to be a reckless or callous disregard of, or indifference to, the rights or safety of others, then you may assess punitive or exemplary damages in addition to any award of actual damages.
>
> ". . . The amount of punitive or exemplary damages assessed against any defendant may be such sum as you believe will serve to punish that defendant and to deter him and others from like conduct."[97]

After the jury found Smith liable and awarded $25,000 in compensatory and $5000 in punitive damages, the Supreme Court affirmed a judgment based upon these instructions.[98]

Smith argued that the punitive award should be reversed because the instruction should have limited punitive damages to situations involving "'ill will, spite, or intent to injure.'"[99] The Supreme Court rejected this argument:

> Smith's argument . . . is that an actual-intent standard is preferable to a recklessness standard because it is less vague. . . . He concedes, of course, that deterrence of future egregious conduct is a primary purpose of both § 1983 and of punitive damages. But deterrence, he contends, cannot be achieved unless the standard of conduct sought to be deterred is stated with sufficient clarity to enable potential defendants to conform to the law and to avoid the proposed sanction.
>
> . . . .
>
> . . . The need for exceptional clarity in the standard for punitive damages arises only if one assumes that there are substantial numbers of officers who will not be deterred by compensatory damages; only such officers will seek to guide their conduct by the punitive damages standard. The presence of such officers constitutes a powerful argument *against* raising the threshold for punitive damages.[100]

---

96. *Id.*

97. *Id.* (quoting the trial judge's jury instructions) (emphasis omitted).

98. *Id.* at 33, 56.

99. *Id.* at 37 (quoting Brief for Petitioner at 9).

100. *Id.* at 49-50 (citations omitted).

This reasoning, although it does correctly characterize the deterrence function, is flawed. All corrections officers now will need to trim their performance to avoid the prospect of vague punitive-damage liability.[101] This liability can be inappropriately overimposed just as it can be underimposed, and it can be imposed even upon officers who believe they are acting reasonably, if a jury later disagrees. Thus, in his dissent, Justice Rehnquist, joined by Chief Justice Burger and Justice Powell, pointed out that "the uncertainty resulting from largely random awards of punitive damages will have serious effects upon the performance by state and local officers of their official duties."[102]

Despite this disagreement, the Justices seem to have correctly understood the basic economic purpose of punitive damages.[103] The majority properly made the choice of a threshold dependent not on anything inherent in formalisms such as "intent" or "spite," but rather on the use of these labels to communicate the optimal level of deterrence. The dissent—which seems to have had the better understanding of the indifference of economic deterrents to purity of heart—accepts the deterrence principle, but it argues that the level chosen by the majority will result in overkill.

B. Developing the Economic Theory from Basic Principles

1. Adjusting the Defendant's Cost Curves Through Tort Remedies.— The economic theory sketched above can best be developed by an examination of short-term marginal cost curves. Marginal cost is the cost associated with each additional unit of output.[104] In the short run, marginal cost increases with high levels of output, as expensive factors of production are pressed into service; the firm pays overtime to workers, uses worn machinery, and buys expensive materials from less efficient firms on the fringe.[105] Therefore, the marginal cost curve slopes upward.[106] Figure 1 shows a private firm's marginal cost

---

101. *Cf.* SHAVELL, *supra* note 34, at 79-83 (explaining the effects of uncertainty); *see also infra* Part II.A.2-3 (discussing irrationality and uncertainty).

102. *Smith*, 461 U.S. at 88 (Rehnquist, J., dissenting).

103. *See supra* note 83 and accompanying text.

104. SAMUELSON & NORDHAUS, *supra* note 52, at 514 (explaining that marginal cost "denotes the extra or additional cost of producing [one] extra unit of output").

105. *See id.* at 514-15 (discussing the calculation of marginal cost).

106. *Id.* Actually, the marginal cost curve is U-shaped when considered as a whole, because at low output the cost of each unit is high, reflecting scale economics. *Id.* The marginal cost curve slopes upward, however, at the point where it intersects the average cost curve at the point of lowest average cost. *Id.* This represents efficient production. *See id.* at 515-20 (discussing the relationship between cost and production).

curve (indicated by the dotted line), when the producer is not motivated to consider externalized accident costs.[107]

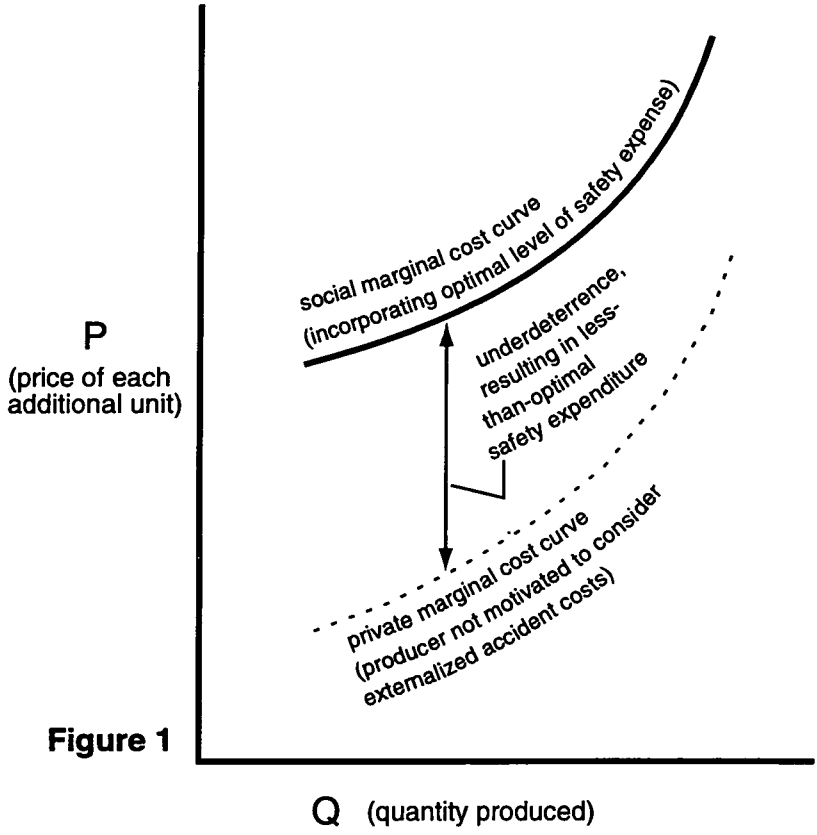

Figure 1

In such a situation, the society wishes that the producer actually would use a more expensive method of production, one that does not externalize such losses as those caused by accidents.[108] Therefore, the socially desirable level of production cost is higher, even though this cost means higher product cost to the consumer.[109] The social marginal cost curve (indicated by the solid line) is shaped somewhat like the private firm's marginal cost curve, but it reflects a higher level of dollar cost per unit of production. The cost level is higher because it includes safety-related behaviors that reduce accidents. Thus, Figure

107. *See infra* fig.1. This figure is adapted from POSNER, *supra* note 59, at 176-77 fig.6.2.

108. *See* POSNER, *supra* note 59, at 180-81 (discussing a consumer's willingness to pay more for a safer product in terms of a balancing of possible accident costs against the costs of a safer product).

109. *See id.*

1 shows the social marginal cost curve, which incorporates the optimal level of safety expense, at a higher cost level than the private firm's marginal cost curve. The difference between the curves is the level of underdeterrence, or the degree to which the producer will externalize. The producer's lack of motivation to consider accident costs results in this amount of shortfall in safety expenditures from an optimal level.

The economic goal of the tort system, then, is quite simple: to force the producer to internalize accident costs by shifting the firm's cost curve. In general, the law targets this goal through actual, compensatory damages, as we saw in the preceding subpart of this Article.[110] Inevitably, however, there is a gap in deterrence, which may be small or large, caused by failures in detection, prosecution, proof, or remedy.[111] Figure 2 illustrates this set of circumstances.[112] The firm's private marginal cost curve (dotted line) again is below the social marginal cost curve (solid line). The imposition of actual damage liability, however, causes the firm to shift its marginal cost curve upward, because the firm is motivated to make expenditures for safety features by the possibility that the courts may force it to compensate accident victims.[113] The shifted marginal cost curve (dashed line) is not socially optimal, however, because the failure of all accident victims to recover their costs fully results in a level of underdeterrence, as is indicated.

Figure 3 illustrates the way in which punitive damages may force the proper shift.[114] In the absence of a tort system, the producer

---

110. *See supra* notes 61-65 and accompanying text.

111. *See supra* notes 82-83 and accompanying text.

112. *See infra* fig.2. This figure, like Figure 1, is adapted from POSNER, *supra* note 59, at 176-77 fig.6.2, with the addition of the deterrence gap.

113. It should be added that a shifted demand curve for the consumer is an alternative possibility. *See* SHAVELL, *supra* note 34, at 52 (explaining that potential customers would go elsewhere if a firm were to take less than optimal care). That is to say, if the producer externalizes injuries, the consumer responds by exhibiting a demand curve that is lower than the demand curve for a safer product. *Id.* If information is good, the consumer may take precautions such as selecting a safer product, purchasing ancillary equipment that reduces risk, and using additional care in connection with the product. *See* A. MITCHELL POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS chs. 9, 13 (2d ed. 1989) (discussing the allocation of risk between an injurer and a victim, and describing the effect of an informed consumer on the allocation of risk). Additional care by the consumer may be more efficient than additional care by the producer, and there is a theoretically optimal mix of care by both. *See id.* Efficiency of consumer care, however, is impaired by informational deficiencies. *See* SHAVELL, *supra* note 34, at 53 (describing the impact of imperfect customer knowledge on the care taken by the customer).

114. *See infra* fig.3. Once again, this figure is adapted from POSNER, *supra* note 59, at 176-77 fig.6.2. The concept of punitive damages as a gap-filler is added. *See generally* Keith N. Hylton, *A Missing Markets Theory of Tort Law*, 90 Nw. U. L. REV. 977 (1996) (explaining

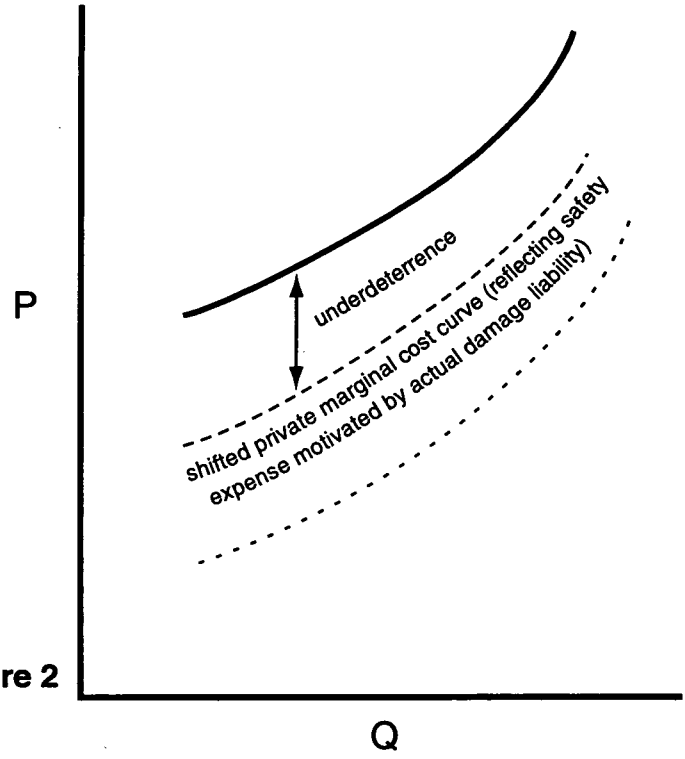

Figure 2

would confront only the private cost of producing the product (dotted line), without incurring any safety expense. The social marginal cost curve (solid line) incorporates the level of safety expenditures that the society desires. The goal of deterrence is to force the producer to consider external accident costs by incorporating a proper level of safety expense, a level at which the shifted cost curve would approach the social curve. The firm's shifted marginal cost curve (dashed line) now includes expenses for safety features induced by the *combination* of compensatory and punitive damages. This combination of compensation with optimal punitive damages, measured to fit the deterrence gap, forces the firm to produce with the socially desired mix of safety and non-safety expenses.

In summary, there is an economically optimal level for punitive damages. This optimal level is reached, the economist would say, when the fully shifted private marginal cost curve becomes identical to

---

tort law with reference to the incompleteness of markets as allocators of risks); Jason S. Johnston, *Punitive Liability: A New Paradigm of Efficiency in Tort Law*, 87 COLUM. L. REV. 1385 (1987) (describing a theory of optimal punitive liability based upon the redressing of deficient damages due to uncertainty in recovery).

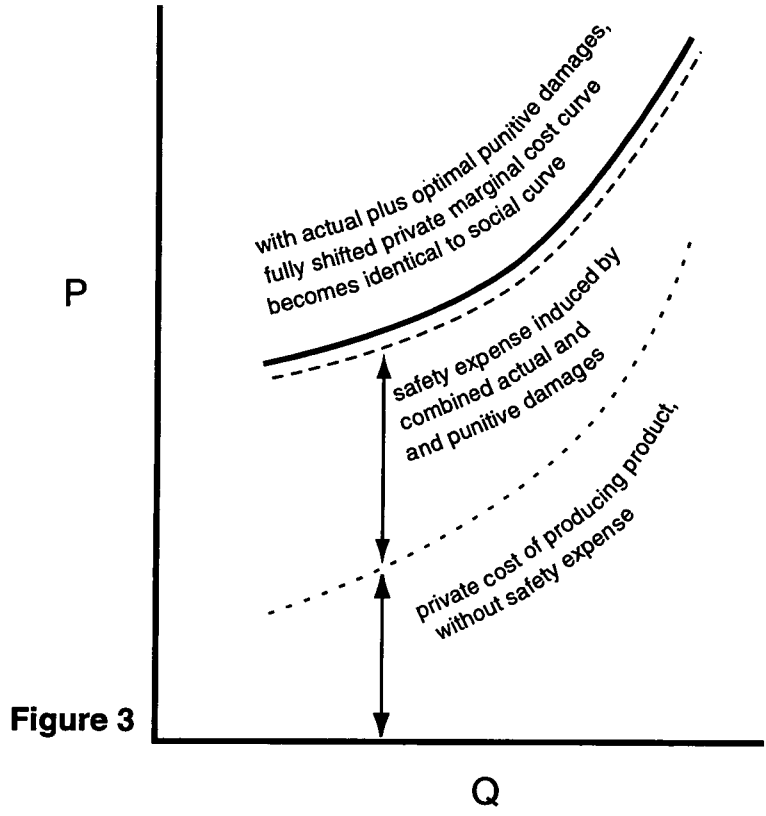

Figure 3

the social marginal cost curve. At this level, the firm is producing with precisely the mix of efficient production expense plus safety expense that represents the society's desired allocation of goods and services. The combination of punitive and compensatory damages, at this level, forces firms to reduce their externalization of accident costs by spending the optimal amount on accident prevention.

2. The Economic Limits: Costs, Benefits, Supply, and Demand.—Punitive damages are difficult to value, however, and therefore, they are difficult to keep in check.[115] One reason is that cost curves of this kind are exceedingly difficult to derive empirically.[116] Therefore, levels of underdeterrence are difficult to measure. Furthermore, there is a natural tendency to assume that if one dollar of accident-prevention expense is a good idea, two dollars are twice as good; con-

115. *See* CRUMP ET AL., *supra* note 82, at 1024.

116. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 9-10, 595-602 (discussing the roles of risk and uncertainty in economics).

sequently, there is a human urge to exaggerate the punitive award with the idea of perfectly deterring accidents. A decisionmaker in the form of a judge or juror may reason (erroneously) that the proper level of punitive damages is "whatever level is necessary to make these companies stop this behavior"—i.e., to stamp out the offending behavior completely.[117]

This situation is depicted in Figure 4.[118] Again, the cost curve of the private firm externalizing accident costs (dotted line) is below the social marginal cost curve (solid line). In this case, the imposition of a large punitive-damage liability actually shifts the private marginal cost curve (dashed line) above the social curve. These circumstances result in a level of punitive-damage "overkill," represented by the difference between these two curves.[119]

The disadvantages of this overkill are illustrated in Figure 5.[120] In an atomistically competitive market,[121] the firm's supply curve is identical to its marginal cost curve.[122] Figure 5 therefore labels the marginal cost curves as supply curves and couples them with a demand curve. The supply curves are sloped upward, because producers are eager to produce greater quantities when price increases; the demand curve, on the other hand, slopes downward, because consumers are willing to purchase greater quantities when price decreases.[123] The intersection of the supply and demand curves is the market price.[124] Higher prices bring about a disequilibrium in which sellers wish to

---

117. *Cf. infra* note 268 and accompanying text (illustrating popular attitudes toward punitive awards imposed on economic actors).

118. *See infra* fig.4. This figure is adapted from POSNER, *supra* note 59, at 176-77 fig.6.2, with the addition of overdeterrence considerations. *See id.* at 224-25.

119. *See* SHAVELL, *supra* note 34, ch. 6 (discussing the optimal level of liability and the results of overdeterrence).

120. *See infra* fig.5. This figure is adapted from POSNER, *supra* note 59, at 176-77 fig.6.2, with the addition of overdeterrence considerations.

121. "Atomistic competition," also called "perfect competition," refers to a market in which the number of sellers is so large that none can exert any individual influence on price (i.e., each firm occupies an atomistically small market share), and the product is undifferentiated, homogeneous, and generic (i.e., brand differentiation does not permit price fluctuation). *See* SAMUELSON & NORDHAUS, *supra* note 52, at 968 (defining perfect and imperfect competition). Theories of competitive efficiency typically are fashioned for atomistically or perfectly competitive markets on the one hand, or for oligopoly or monopoly on the other. *See id.*

122. This is so in that portion of the marginal cost curve where marginal cost is rising. *Cf. supra* note 106 and accompanying text (explaining the slope of the marginal cost curve). This is the area where the firm is efficient, which it must be in a competitive market. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 542-43 (discussing the derivation of supply curves from marginal cost curves).

123. *See* SEIDENFELD, *supra* note 52, at 36-37.

124. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 557.

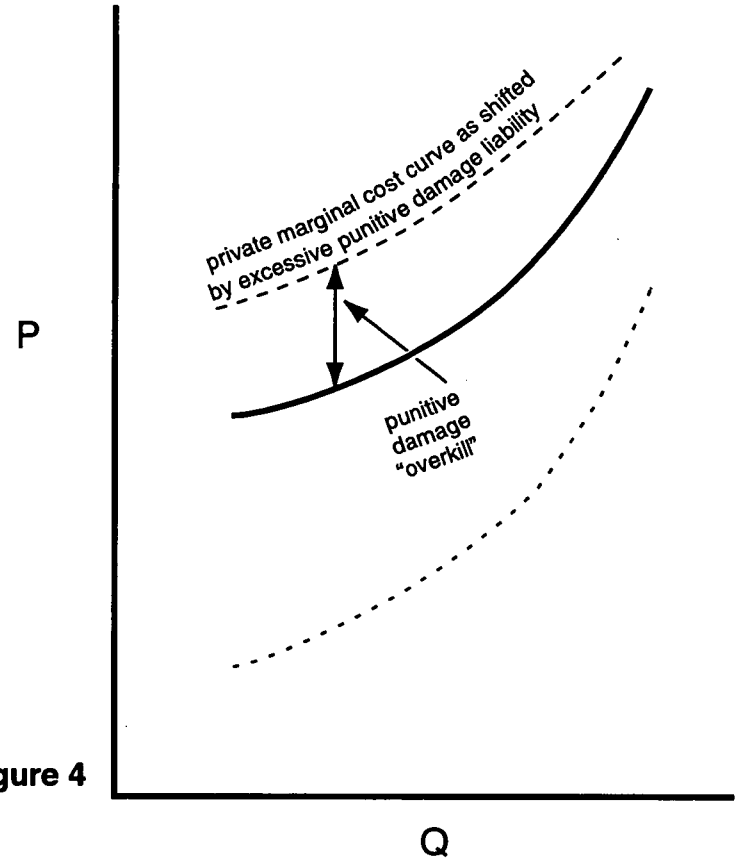

Figure 4

enter the market, while buyers wish to exit; a lower price results in an undersupply, with consumers demanding more units than sellers are willing to furnish.[125] The market price allocates resources efficiently,[126] as a general proposition,[127] because it correctly matches consumption to output.[128]

Once again, the dotted line represents the firm's supply curve when the firm externalizes accident costs.[129] This would be the supply curve if there were no tort system and no prospect of damage liability. In such circumstances, producers are willing to supply great quantities at low prices, but there is a catch: They cause a socially unacceptable

125. *See id.*

126. *See id.* at 741-54 (discussing markets and economic efficiency).

127. *See id.* This generalization is subject to exceptions, when equilibrium either is not efficient or is not reached (as by cobweb-type divergence). *Cf. id.* at 557-64 (describing markets with peculiar dynamics).

128. *See id.* at 742-45 (discussing markets and economic efficiency).

129. *Cf. supra* fig.1 (illustrating private supply curve as compared to social cost curve).

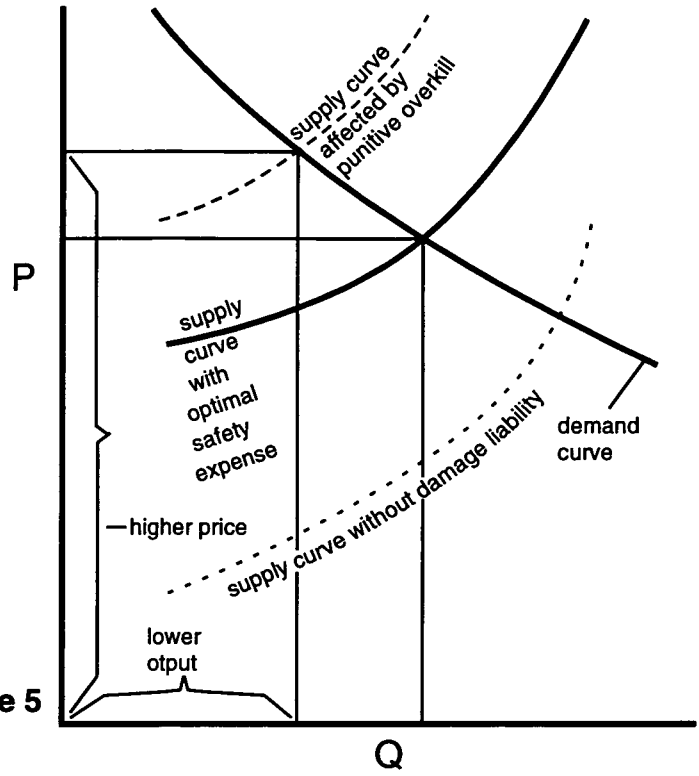

Figure 5

number of accidents.[130] The higher-level supply curve (solid line) is the social supply curve, reflecting optimal safety expense. The supply curve affected by punitive-damage overkill (dashed line) represents the supply curve that will result if firms are affected by an excess of punitive liability.

When firms face this sort of punitive-damage overkill, the tort system sends dysfunctional messages to producers. As a result, the market price is higher, and output is lower, as Figure 5 shows.[131] Consumers find that their purchases are more difficult and more expensive to complete, even in the case of crucially necessary products. If we apply an excess of punitive damages to producers of life-saving medicines, ambulance manufacturers, or health-care providers, for example, we may find that we have both restricted the output of these goods and services and priced them out of the reach of some consum-

130. *See supra* note 108 and accompanying text.

131. *See supra* fig.5; *cf.* SEIDENFELD, *supra* note 52, at 44-46 (discussing the effects of government intervention on pricing). Although Seidenfeld's examples concern the imposition of taxes and price caps, the analysis would be similar for a dysfunctional cost, such as an excessive amount of punitive damages, imposed by government.

ers.[132] The judge or juror who wants to impose crushing punitive liability, so as to "stop" accidents completely, may be acting with the best of intentions and may be following natural human urges, but the economic result of this choice can be disastrous shortages of essential commodities.

C. Economics and Tort Law: Finding the Proper Occasions for Deterrence—and the Right Level

It should not seem peculiar that this analysis of punitive damages is done in economic terms. Economic analysis of tort law is a useful aid to its proper development, from age-old problems of trespass to modern issues of product liability. Even when the reasoning of the common-law courts has been ruthlessly formalistic, the results almost always have made sense in economic terms.[133] The economic goal, and usually the approximate result of the common law, is to administer the deterrent effect of a damage remedy in the appropriate circumstances and at roughly the proper level.[134] This level of deterrence is that at which the last dollar spent on accident prevention exactly equals the loss in manufacturing value that it causes, so that greater expenditures will cost more than they are worth.[135]

Occasionally, common-law judges have used explicit formulae that recognize this economic goal. Perhaps the most famous case is *United States v. Carroll Towing Co.*,[136] in which Judge Learned Hand explained the basic negligence calculus.[137] There, the issue was

---

132. *Cf.* SEIDENFELD, *supra* note 52, at 46 & fig.27 (illustrating shortages resulting from a price cap). *See generally* THOMAS J. CAMPBELL ET AL., THE CAUSES AND EFFECTS OF LIABILITY REFORM: SOME EMPIRICAL EVIDENCE (National Bureau of Econ. Research Working Paper No. 4989, 1995) (showing that states that reduced liability through tort reform experienced increases in measured productivity and employment); GENERAL AVIATION MANUFACTURERS ASS'N, REPORT TO THE PRESIDENT AND CONGRESS: THE RESULTS OF THE GENERAL AVIATION REVITALIZATION ACT (1996) (describing the increase in production and employment resulting from the General Aviation Revitalization Act of 1994, 49 U.S.C. §§ 40101-40120 (1994)). In libel cases, it appears that punitive damages disproportionately discourage political speech, unpopular viewpoints, and minority speakers. *See* Nicole B. Cásarez, *Punitive Damages in Defamation Actions: An Area of Libel Law Worth Reforming*, 32 DUQ. L. REV. 667, 682-88 (1994) (discussing the chilling effect of punitive damages on protected speech and the use of punitive damages to punish unpopular viewpoints and speakers).

133. *See* POSNER, *supra* note 59, at 23 (pointing out that although judicial opinions of common-law courts rarely make explicit reference to economics, they often "bear the stamp of economic reasoning").

134. *Cf. id.* (characterizing the economic goal of the common-law system as the maximization of society's wealth).

135. *See id.* at 163-67 (discussing the economics of accidents).

136. 159 F.2d 169 (2d Cir. 1947).

137. *See id.* at 173; *see also* POSNER, *supra* note 59, at 163-67 (explaining the Learned Hand formula). Judge Posner also demonstrated that the formula underlies other deci-

whether a barge owner was negligent in failing to have an attendant on board while the barge was being towed, in case she broke away.[138] Judge Hand's analysis was "a function of three variables: (1) [t]he probability that she will break away; (2) the gravity of the resulting injury, if she does; [and] (3) the burden of adequate precautions."[139] Judge Hand observed that it might help to state the formula in algebraic terms, and he proceeded to do so: "[I]f the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P: i.e., whether $B < PL$."[140]

Stated as an equation in modern mathematical notation, Judge Hand's formula means that negligence liability begins to appear at the point where $B < PL$. This is an economic formula. Judge Posner explains the significance of Judge Hand's reasoning as follows:

> If the cost of safety measures . . . exceeds the benefit in accident avoidance to be gained by incurring that cost, society would be better off, in economic terms, to forgo accident prevention. . . . [O]verall economic value or welfare would be diminished rather than increased by incurring a higher accident-prevention cost in order to avoid a lower accident cost. If, on the other hand, the benefits in accident avoidance exceed the costs of prevention, society is better off if those costs are incurred and the accident averted, and so in this case the enterprise is made liable, in the expectation that self-interest will lead it to adopt the precautions in order to avoid a greater cost in tort judgments.[141]

Graphically, this reasoning can be illustrated by a diagram such as Figure 6.[142] The burden, B (or cost due to accident prevention expenses) increases as the degree of care, or the amount of money spent on safety, increases. Therefore the line representing B slopes upward. The probable cost due to losses from accidents, or the probability of an accident multiplied by the likely amount of the loss ($P \times L$), produces the downward-sloping line PL, because the value of these losses

---

sions, such as *Adams v. Bullock*, 125 N.E. 93, 93 (N.Y. 1919), which held that because of the relation of burden, probability, and damages, the defendant was not liable for injuries to a boy who touched a trolley wire with a metal cable, and *Hendricks v. Peabody Coal Co.*, 253 N.E.2d 56, 60-61 (Ill. App. Ct. 1969), which imposed damages because the cost or the burden was slight as compared to the risk. POSNER, *supra* note 59, at 166-67.

138. *Carroll Towing*, 159 F.2d at 171.

139. *Id.* at 173.

140. *Id.* (emphasis added).

141. Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 29, 32-33 (1972).

142. *See infra* fig.6. This figure is adapted from POSNER, *supra* note 59, at 164-65 fig.6.1. It is modified in some respects, such as by the addition of features that illustrate overdeterrence.

declines as the degree of care increases. The intersection of the two curves, where $B = PL$, is the optimal level of care, or "due care." At lesser degrees of care, the tort system threatens the producer with negligence liability, thereby inducing greater social efficiency through negligence liability for damages.[143] At greater degrees of care, the producer wastefully invests safety expenses that cost more than they are worth in accident reduction.[144] Here, the market system will induce efficiency.

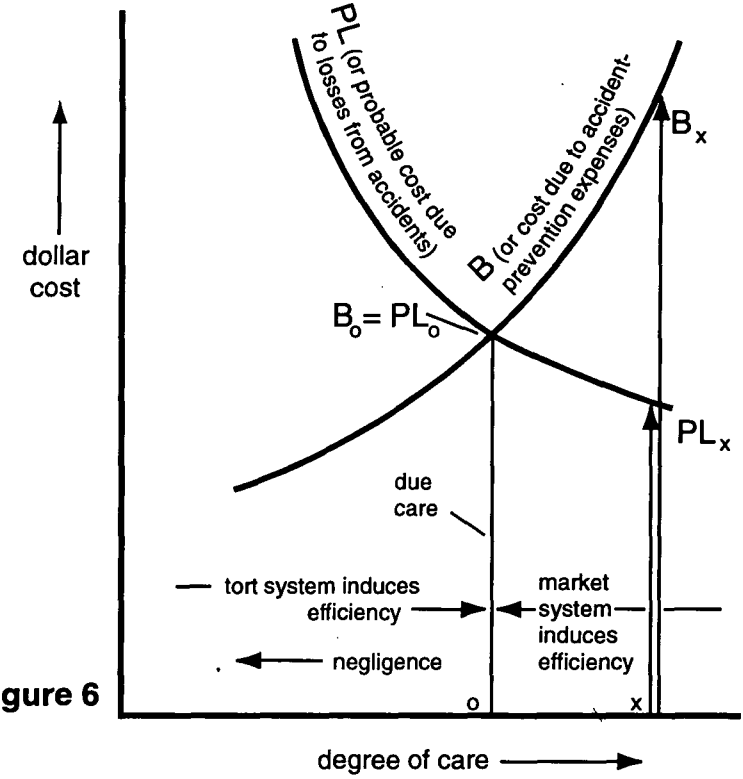

Figure 6

Figure 6 allows us to illustrate the effect of excess liability. At safety level X, for example, the expense attributable to the degree of care is disproportionately high, well above the level of due care. At safety level X, the total safety expense (B_x) added to the probable accident loss (PL_x) greatly exceeds the total at the due care level (point O), $B_o + PL_o$. (That is, $B_x + PL_x > B_o + PL_o$.) At the higher level, an automobile producer is forced to manufacture a product that resem-

143. *See supra* notes 61-65 and accompanying text.
144. *See supra* notes 59-60 and accompanying text.

bles a Sherman tank in order to avoid an accident that is the economic equivalent of a hangnail. This dysfunctional effect results from the response of a rational manufacturer to excess damage liability.

This analysis dovetails with the theory of punitive damages set out in earlier sections of this Article.[145] The goal of the tort system is to create precisely that level of deterrence at which the producer is motivated to match safety expenses and accident costs so that $B = PL$.[146] If, owing to underprosecution or nonrecovery, compensatory damages fail to achieve this level, the economic theory means that the law should impose punitive damages to close the gap. Inadequate use of punitive damages means a market that produces cheap cars, which are undesirable even though inexpensive, because they are made without seat belts and with exploding gasoline tanks. Punitive-damage overkill, on the other hand, means that the tort system forces production of clumsy, armored vehicles with little enhancement in safety. Consumers of products affected by this overkill will discover that the products are more expensive and more difficult to find, even if the products are useful and desirable. Once again, an excess of punitive damages applied to producers of life-saving medicines, ambulances, or services of physicians will both restrict the output of these goods and services and price them out of the reach of some consumers.

II. ARGUMENTS AGAINST THE ECONOMIC APPROACH: IMPERFECTIONS IN THE MESSAGE

Any effort to apply these principles to actual decisions about evidence and punitive damages, however, must take account of the limits of the economic theory upon which they rest. This Part of the Article considers two kinds of limits that may cause the theory to send the wrong message to actors in the marketplace. First, difficulties with obtaining necessary information about the economic conditions may produce the wrong level of punitive damages.[147] Second, the economic theory set forth here is largely divorced from moral concerns. A deontological analysis of punitive damages—an approach emphasizing moral duty, "oughtness," or right and wrong—may give different answers about the kinds of evidence rules that are desirable.[148] This Part concludes that the economic theory must be considered in light of these limits, but that it still is a useful tool for analysis of punitive damages and the evidence rules that influence them.

---

145. *See supra* Part I.B.
146. *See supra* fig.6.
147. *See infra* Part II.A.
148. *See infra* Part II.B.

A. *The Consequences of Lack of Information: Incommensurability, Irrationality, and Uncertainty*

Diagrams such as Figures 1 through 6 conceal some fearsome difficulties. They are useful because they clarify economic relationships. But they achieve this clarification precisely because they remove other problems. For example, a real firm in a real marketplace cannot consult any handbook or web site that will show it the precise contours of the social marginal cost curve. The firm is more likely to be able to derive its own private marginal cost curve, but even this is arrived at with a great deal of guesswork.[149] In summary, principles of evidence and jury control must take account of incommensurability, irrationality, and uncertainty, if they are to work in the real world.[150]

1. Incommensurability of Values.—When we compare accident losses and safety expenses, we are in a sense comparing apples to oranges. Justice Scalia stated the problem well when he observed, in *Bendix Autolite Corp. v. Midwesco Enterprises*,[151] that the Supreme Court often is called upon to decide "whether a particular line is longer than a particular rock is heavy."[152] Thus, in the abortion cases, the Court has undertaken to "balance" the societal interests in preserving potential life against individual abortion-related interests such as care of an unwanted child.[153] There is no common measure for these values, just as there is no common unit for apples and oranges.[154] It might be supposed that the balancing by the Court would be more reliable when one factor seems far more prominent than another, as if we compared a six-inch line with a thousand-pound rock. But this appearance of accuracy is illusory. What if the line is a chain of plati-

---

149. *Cf.* SAMUELSON & NORDHAUS, *supra* note 52, at 595-600 (discussing the economics of uncertainty, particularly with regard to risk premiums); SEIDENFELD, *supra* note 52, at 66 (treating imperfect information as a market deficiency).

150. *See* SHAVELL, *supra* note 34, at 79-83 (discussing the effects of uncertainty, error, and misperception). *See generally* Mark Geistfeld, *Placing a Price on Pain and Suffering: A Method for Helping Juries Determine Tort Damages for Nonmonetary Injuries*, 83 CAL. L. REV. 773 (1995) (using market methods to provide alternatives to the arbitrariness of tort awards that result from deficient information, compensation, and enforcement).

151. 486 U.S. 888 (1988).

152. *Id.* at 897 (Scalia, J., concurring in judgment).

153. *Cf.* Planned Parenthood v. Casey, 505 U.S. 833, 845-49 (1992) (reaffirming the essential holding of *Roe v. Wade* but recognizing that its boundaries are not susceptible to the expression of a simple rule); Roe v. Wade, 410 U.S. 113, 154 (1973) (holding that the right of personal privacy includes the abortion decision but that it is qualified by certain interests of the state).

154. *Cf.* LAURENCE TRIBE, ABORTION: THE CLASH OF ABSOLUTES 27-41 (1990) (demonstrating the incommensurability of the two competing values implicated in the abortion debate—a fetus's right to live and a woman's right to personal autonomy).

num, while the rock is only a half-ton of cracked limestone?[155] In the tort arena, the risk of loss of life or paraplegic injuries may be weighed against the consumer's ability to choose an attractive convertible in the automobile market.[156]

The economist assumes that each of these values can be translated into dollar amounts.[157] This is one of several assumptions that enables the economist to draw clear diagrams, and although the resulting clarity is useful in illustrating economic tradeoffs,[158] the fact that the tradeoffs are of incommensurate values must not be forgotten. In some instances, the law recognizes the Herculean difficulty of placing dollar amounts on values—in fixing damages for the wrongful loss of a human life, for example—and the governing legal principles normally must recognize the roughness of the result.[159] Still other values, such as love, achievement, or pleasure, are similarly difficult to quantify, although again the law does so when it is necessary.[160] Plac-

---

155. *Cf.* Cass R. Sunstein, *Incommensurability and Valuation in Law*, 92 MICH. L. REV. 779, 818-20 (1994) (discussing approaches by the law to valuation of, or refusal of the law to value, intangible goods such as life, love, etc.).

156. This is so because the product must be "unreasonably dangerous," meaning that its risk must be disproportionate to the utility derived from it. RESTATEMENT (SECOND) OF TORTS § 402A cmt. i (1965). "Utility" in this context includes aesthetic concerns: "Good whiskey is not unreasonably dangerous merely because it will make some people drunk, and is especially dangerous to alcoholics; but bad whiskey, containing a dangerous amount of fusel oil, is unreasonably dangerous." *Id.*

157. *See* POSNER, *supra* note 59, at 165 (graphing, on the same dollar scale, damages from injuries against production costs or burdens).

There are at least two arguments countervailing against a concern about incommensurability. First, the purpose of punitive damages is to put a price on the activity, and so there is no incommensurability for the defendant, who weighs dollars against dollars. The law merely speaks in the same language as the defendant. The trouble with this reasoning, however, is that it does not solve the problem of the lawgiver or adjudicator, who must fix the amount of the sanction by considering incommensurate values and making them, somehow, commensurate. Second, it is likely that any other remedy (e.g., imprisonment) will exhibit the same problems of incommensurability, so that the alternative is to do nothing.

158. Thus, for example, Gary S. Becker has demonstrated the usefulness of economic analysis in predicting a wide variety of human behavior that otherwise might be thought of as non-economic, including crime, racial discrimination, and marriage and divorce. *See generally* GARY S. BECKER, A TREATISE ON THE FAMILY (1981) (demonstrating an economic approach to aspects of family life); GARY S. BECKER, THE ECONOMIC APPROACH TO HUMAN BEHAVIOR (1976).

159. *See, e.g.*, Seffert v. Los Angeles Transit Lines, 364 P.2d 337, 345 (Cal. 1961) (recognizing the difficulty of valuing pain and suffering, but nevertheless recognizing the right to a limited recovery); Sullivan v. Old Colony St. Ry. Co., 83 N.E. 1091, 1092 (Mass. 1908) ("The rule of damages is a practical instrumentality for the administration of justice. . . . Its object is to afford the equivalent in money for the actual loss caused by the wrong of another.").

160. *See, e.g.*, Sanchez v. Schindler, 651 S.W.2d 249, 251, 252, 254 (Tex. 1983) (overruling past decisions, limiting wrongful-death recovery to pecuniary amounts, and allowing

ing a dollar value on the abstract right to the freedom of speech based upon its importance to our republic, as the district judge charged the jury to do in the *Stachura* case discussed above,[161] takes the problem into another dimension. In any event, this problem of incommensurability is an important consideration in crafting real-world rules of evidence and jury control for punitive damages.

2. *Irrationality of Enforcement.*—Even if we could derive clear, accurate curves by perfectly translating into dollars all accident losses, from hedonic pain to loss of life, and comparing them to accurate totals of all safety expenses, we would face a second and related difficulty. Marginal cost curves are meaningless unless decisionmakers understand them. It would be necessary to educate every judge, and for that matter every juror, in the economic theory that begins this Article.[162] We would need to explain social marginal cost curves and their interrelationships with damage principles in jury instructions that could be perfectly understood by lay jurors, who then would apply them without confusing them with concepts such as moral blame.[163] Furthermore, and most importantly, because the economic purpose of our tort system is to consider inducements for firms to produce efficiently while internalizing accident costs, we would need to ensure that the meaning of every verdict, present or future, would be adequately known to every firm. If a very high punitive award, for example, were to be based upon the jury's perception of a peculiar fact-bound instance of underdeterrence, the firm would need to react to the meaning of the verdict by perceiving the limits on ways in which the verdict could be applied to that firm's (presumably different) circumstances.[164]

---

recovery for intangibles such as companionship). *But cf.* Hogan v. Santa Fe Trail Transp. Co., 85 P.2d 28, 33-34 (Kan. 1938) (disallowing the recovery of $4000 awarded by the jury as compensation for the loss of enjoyment of playing the violin, but recognizing that pain and suffering were compensable).

161. *See supra* notes 66-75 and accompanying text.

162. *Cf.* SAMUELSON & NORDHAUS, *supra* note 52, at 8 ("After you have studied and learned a body of economic principles, you comprehend reality in a new and different way." (emphasis omitted)). For an interesting effort to provide better jury guidance, see Oscar G. Chase, *Helping Jurors Determine Pain and Suffering Awards*, 23 HOFSTRA L. REV. 763, 777-78 (1995), which advocates that jurors be furnished with grids or charts containing median, high, and low sums awarded in previous cases for each of the widely accepted nine-point severity levels.

163. Economics is a science of rational choice and of the effects of incentives. *See* POSNER, *supra* note 59, at 3-4. Deontological reasoning is described in *infra* Part II.B.

164. *See* POSNER, *supra* note 59, at 3-4 (explaining the assumption of rational self-interest). It is not necessary, however, that all participants act consciously for a message of this

In practice, judges do not mathematically follow either economic principles or rules of law.[165] This failure creates slippage in the theory of deterrence. Furthermore, jurors cannot be screened for understanding of economic theory; indeed, they often fail to apprehend relatively simple instructions.[166] This failure produces more slippage. And finally, the producer ordinarily can make only the crudest guess concerning the relevance of a novel jury verdict, even assuming the firm or its insurer knows of it.[167] These elements of irrationality are another factor influencing evidence rules and jury control of punitive damages.

3. Uncertainty of Liability in the Future.—Learned Hand's *BPL* formula assumes that factors such as the "probability" of accidents are known,[168] but if so, they are known in only a very general way. Sometimes, recurring situations can be subjected to actuarial analysis, but even this science is imperfect, because it depends upon evaluation of past circumstances and their comparison to a different future.[169] The firm considering production efficiency against safety expenses cannot precisely know either the probability that an additional dollar of safety expense will avoid an accident or the mathematically expected loss

---

kind to be communicated. *Id.* For example, the result might occur because of specifications by an insurer who is better informed than individual insureds.

165. *Cf.* Clayton P. Gillette, *Rules, Standards and Precautions in Payment Systems*, 82 VA. L. REV. 181, 185-86 (1996) (describing the effects of uncertainty in commercial law and the means of dealing with it); Louis Kaplow & Steven Shavell, *Accuracy in the Assessment of Damages*, 39 J.L. & ECON. 191, 201-03 (1996) (demonstrating systematic factors that distort damage awards and defeat the public understanding of those factors); Rolando F. Peláez, *Higgledy-Piggledy Awards for Lost Earnings*, 36 JURIMETRICS J. 325, 336 (1996) (demonstrating that "the below-market method forecloses accuracy in the discounting process and opens the way to spectacular award errors"); Mark V. Tushnet, *Following the Rules Laid Down: A Critique of Interpretivism and Neutral Principles*, 96 HARV. L. REV. 781, 814-21 (1983) (cataloguing reasons why stare decisis is ineffectual and why resulting legal rules are indeterminate); Richard Whisnant & Diane DeWitt Cherry, *Economic Analysis of Rules: Devolution, Evolution and Realism*, 31 WAKE FOREST L. REV. 693, 728 (1996) (describing the economics of enforcement uncertainty in the regulatory environment).

166. *See, e.g.*, Leonard V. Sand & Steven Alan Reiss, *A Report on Seven Experiments Conducted by District Court Judges in the Second Circuit*, 60 N.Y.U. L. REV. 423, 456 (1985) (citing experiments that assist jurors' understandings of lengthy charges); J. Alexander Tanford, *The Law and Psychology of Jury Instructions*, 69 NEB. L. REV. 71, 79 (1990) (stating that psychological studies conclude that jury instructions are often incomprehensible to jurors).

167. Even legal advice from a knowledgeable professional is often crude. *Cf.* CRUMP ET AL., *supra* note 82, at 1104-07 (documenting the reaction of banking attorneys to a complex decision, which imposed liability on a lender after foreclosure).

168. *See* POSNER, *supra* note 59, at 165; *see also supra* notes 140-141 and accompanying text.

169. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 595-602 (explaining the difficulty of predicting the future).

from the accident.[170] For that matter, it cannot predict the evidence rules and legal principles that a future jury will apply to such a hypothetical accident years from now.[171]

Evidence and jury-control principles depend heavily upon these formidable difficulties. The goal of real-world trials cannot be the precise punitive award that exactly balances safety expenses and accident costs; rough estimation is all that can be attempted. In many instances, a court will be forced to sacrifice accuracy for a clarity of expression that reduces the problems of incommensurability, irrationality, and uncertainty that are inherent in the tort system.[172] It probably is not worth the candle, for example, to have expert economists routinely testify about supply and demand or to fix the dollar level of appropriate deterrence in car wreck cases. We usually do not instruct juries in terms such as marginal costs, transaction costs, or the law of diminishing returns, and we do not expect manufacturers to have to use these concepts to understand the verdicts that result.[173]

Nevertheless, the economic analysis of punitive damages should be influential. Although we should not choose an ostensibly accurate principle if it is incomprehensible or misleading in the context of a jury trial, we should not make the opposite error either. That is to say, we should not adopt a given principle, or rely upon a given type of information, merely because it is clearly known. The principle or information may, in fact, be inaccurate and misleading.[174] This delicate balance between accuracy and clarity should animate the derivation of evidence rules and jury control for punitive damages. It is to this subject, therefore, that Part III of this Article will turn later. Presently, this Article will consider a second limit on the economic analysis: the deontological approach.

---

170. *See id.* In theory, uncertainty by itself should not create this problem unless errors cumulate. *Cf. id.* at 9-10. Pure uncertainty, if unaccompanied by bias, sometimes will underdeter and sometimes will overdeter, and if the two possibilities are equal, the firm's calculus will not be disadvantaged by it. But this theory assumes perfectly unbiased uncertainty, which is an improbable state of affairs given the plaintiff's burden of proof. Furthermore, the theory assumes no effect from risk aversion.

171. *See* SHAVELL, *supra* note 34, at 82 (describing the uncertainty of how courts will evaluate future evidence).

172. *See* CRUMP ET AL., *supra* note 82, at 873. This is the thrust of the movement for "plain language" in jury instructions. *Id.* The difficulty is that so-called plain language is rarely clear and precise for every use to which it might be put in an actual case. *Id.*

173. *See infra* Part III.B.3 (discussing advantages and disadvantages of admitting these kinds of evidence in some punitive-damages cases).

174. An example is the admissibility of net-worth evidence as a guide to the determination of punitive damages. *See infra* Part III.B.1 (concluding that net-worth evidence should be excluded).

B. Evidence Relating to Moral Duties: Deontological Factors in Punitive Damages

1. Comparing Consequentialist and Deontological Analyses of Punishment.—The economic arguments sketched above are utilitarian or consequentialist. They seek to maximize utility by using a less costly deterrent to minimize costly behavior.[175] But there are ways to evaluate punitive damages by measures that in some instances have arguable claims to superiority over the marketplace view with which they may conflict. Specifically, one may consider punitive damages as a deontological device,[176] concerned less with maximizing utility than with moral duty.[177] In this alternate view, the justification for exemplary damages is not their function in correcting market deficiencies.[178] Instead, it is an expression of rightness and wrongness, of just deserts, or of moral blameworthiness.[179]

Whereas the consequentialist views the deterrent message perceived by uninvolved actors in the marketplace as the principal purpose of a sanction (even to the exclusion of compensation of persons unjustly injured), the deontologist emphasizes the individual moral positions of the participants in the event.[180] The plaintiff, in this view, deserves compensation not because compensation serves an economic function, but because it is "right."[181] Similarly, the actor who is involved in what we refer to as causation of the injury is to be punished in accordance with the degree of that person's moral blameworthi-

---

175. *Cf.* Partlett, *supra* note 25, at 795-800 (describing deterrence theory and distinguishing it from retributive or moral justifications of punitive damages). *See generally* Avery Wiener Katz, *Positivism and the Separation of Law and Economics*, 94 MICH. L. REV. 2229 (1996) (contrasting the positivist or descriptive approach of economists to the normative or prescriptive orientation of lawyers).

176. Deontology is the study of "ethics dealing esp[ecially] with duty, moral obligation, and right action." RANDOM HOUSE WEBSTER'S COLLEGE DICTIONARY 362 (1995).

177. *See* Partlett, *supra* note 25, at 800-02 (explaining the moral basis of punitive damages); *see also* Jane B. Baron & Jeffrey L. Dunoff, *Against Market Rationality: Moral Critiques of Economic Analysis in Legal Theory*, 17 CARDOZO L. REV. 431, 431-32 (1996) (arguing the superiority of moral justifications over market theories for certain of our most important values); Heidi M. Hurd, *The Deontology of Negligence*, 76 B.U. L. REV. 249, 251-52 (1996) (justifying negligence law by reference to moral theory).

178. *See generally* Partlett, *supra* note 25 (arguing that punitive damages are used for social reasons).

179. *See id.* at 801.

180. *See* Hurd, *supra* note 177, at 252-54 (distinguishing consequentialism and deontology); *see also* Katz, *supra* note 175, at 2241 (describing the orientation of economists toward descriptive, as opposed to normative, approaches).

181. *See* Hurd, *supra* note 177, at 252-54.

ness, and not in a greater or lesser amount calculated by the need to communicate a corrective to the marketplace.[182]

These premises, in the punitive-damages context, lead to an emphasis upon the accomplishment of retributive justice against morally blameworthy injurers. Closely connected to this retributive principle is the idea that penalties should be proportionally graded.[183] To take an example from recent news stories, a school principal might seek to explain a severe punishment, such as expulsion, for a child who possesses Advil or Midol, on the ground that it carries out the school's policy of zero tolerance toward drugs.[184] The consequentialist would ask whether this sanction produces more cumulative benefit in terms of its deterrent purpose than the total of costs or disadvantages attributable to it. The deontologist, on the other hand, would inquire whether the punishment has been triggered by a sufficient violation of moral duty and whether it is proportional to this individual's blameworthiness. It is possible that the consequentialist would see the means (i.e., the severe consequences borne by one individual) as justified by the end (i.e., the cumulative benefits to all others in the society). The deontologist would be troubled, however, by the sacrifice of the individual for the allegedly greater welfare of the many. Deontological reasoning thus includes asking whether the end justifies the means.[185]

Still another deontological concept is that of expiation.[186] Punishment serves the function of discharging the actor's moral debt and fits the wrongdoer for full acceptance back into the society.[187] Punishment also may serve the purpose of condemnation.[188] It thereby pro-

---

182. *Cf.* H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 231 (1968) (positing three conditions for punishment: a voluntary act that is morally wrong, a punishment that is in some way equivalent in severity to the wrong, and that such punishment is just or morally good).

183. *See id.* at 160-70; *see also* TEX. PENAL CODE ANN. § 1.02(3) (West 1994) (including among the objectives of the code "to prescribe penalties that are proportionate to the seriousness of offenses").

184. *See Student Suspended for Carrying Advil; Girl Says Punishment "Too Severe"; Texas School Defends Policy*, WASH. POST, Oct. 10, 1996, at A13, *available in* 1996 WL 13425589.

185. The foremost proponent of this view, still, is Immanuel Kant. *See* IMMANUEL KANT, THE METAPHYSICAL ELEMENTS OF JUSTICE 99-107 (John Ladd trans., Bobbs-Merrill Co. 1965) (1797); *see also* Hurd, *supra* note 177, at 250 (citing Kant for the deontologist's belief in punishment "[w]hen and only when [persons] deserve retribution").

186. *See* FRANZ ALEXANDER & HUGO STAUB, THE CRIMINAL, THE JUDGE AND THE PUBLIC 212-14 (rev. ed. 1956).

187. *See id.* at 212 (stating that rehabilitation, even if effectual, "will not remove the emotional demand that crime must be expiated").

188. *See* HART, *supra* note 182, at 169-73 (discussing the denunciatory theory of punishment); *see also* Partlett, *supra* note 25, at 803 (dealing with the related but distinct elements of retribution and denunciation).

duces in morally upright people the satisfaction that justice is intact,[189] expresses solidarity with innocent injured persons so as to assuage their suffering,[190] and compensates for losses so that nonresponsible persons do not undergo the unfair burden of bearing them.[191] Each of these considerations is related to deontological concepts or to retributive justice, although these ideas also may be related to consequentialist reasoning.[192]

Few people are exclusively consequentialists, and few are pure deontologists. The deontologist, for example, cannot claim to use purely moral principles to derive all of the answers to production or pricing problems facing a firm in the workplace. Likewise, the consequentialist cannot pretend that every decision can be based upon individualized determination of costs and risks, as opposed to relying for some decisions upon rules or norms.[193] Moreover, the consequentialist cannot persuasively defend a regime that unfairly imposes crushing burdens on random individuals to achieve general goals. In contemplating the need to deter drunk drivers, for example, even a person who predominantly is a consequentialist would be unlikely to advance an argument favoring the death penalty for slightly intoxicated of-

---

189. *See* EMILE DURKHEIM, THE DIVISION OF LABOR IN SOCIETY 108-09 (George Simpson trans., Collier-MacMillan Ltd. 1933) (1893) ("We can thus say without paradox that punishment is above all designed to act upon upright people . . . ."); JOEL FEINBERG, DOING AND DESERVING 98, 100-05, 115-16 (1970) (developing subsidiary purposes of this condemnation aspect, including authoritative disavowal, symbolic nonacquiescence, vindication of the law, and absolution of others); 2 SIR JAMES FITZJAMES STEPHEN, A HISTORY OF THE CRIMINAL LAW OF ENGLAND 81 (1883) (stating that the "close alliance between criminal law and moral sentiment is in all ways healthy and advantageous to the community").

190. One theory of this kind is that punishment annuls the wrong done. *See* GEORG WILHELM FRIEDRICH HEGEL, THE PHILOSOPHY OF RIGHT 38 (T.M. Knox trans., Encyclopedia Britannica 1952) (1821) ("The annulment of the crime is retribution."). A different way of putting it is that punishment enables the injured person to overcome the mastery asserted by the wrongdoer and to regain the equality and dignity that existed before the act. *See* Jean Hampton, *Correcting Harms Versus Righting Wrongs: The Goal of Retribution*, 39 UCLA L. REV. 1659, 1685-86 (1992) (arguing that retribution must acknowledge the damage to the victim's worth and repair the damage to the victim's ability to realize the value of her worth).

191. This is so especially for compensatory damages, but it also has been advanced in support of punitive damages to the extent that they fill gaps in compensation (as distinct from deterrence) left by attorneys' fees, rules of nonrecovery, and incomplete damages. *See* Partlett, *supra* note 25, at 793-95.

192. *See infra* notes 195-196 and accompanying text (discussing the interrelatedness of the two theories).

193. Thus, the deontologist cannot plausibly pretend that a distinct moral imperative should govern every question of economically efficient production, nor can the consequentialist credibly maintain that every moral question should be controlled by an ad hoc calculus of cost and benefit that dispenses with all other rules or principles. *See* Hurd, *supra* note 177, at 253-54.

fenders on the ground that theoretical economic calculations proved that the execution of the first unfortunate arrestee would produce deterrence powerful enough to result in a net benefit.[194]

In fact, it is possible to argue that moral retributivism is really a disguised form of utilitarianism.[195] In this view, for example, proportional justice is beneficial because it adjusts the deterrent to avoid hidden costs such as public resistance or rebellion against punishments perceived as unfair. Likewise, condemnation and compensation produce utility in terms of better performance by the generally law-abiding population. Perhaps it is equally possible to view some kinds of consequentialism as disguised forms of deontology.[196] In this view, the retributive correctness of punishments is related to the marketplace, because moral blameworthiness is related to the calculus of risk, or to the excusability (or lack thereof) of the actor's particular resolution of the balancing of the likelihood of harm against the burden of precautions.

2. Deontological Considerations Governing Evidence for Determining Punitive Damages.—It follows that considerations of retributive justice may be useful measures for evaluating punitive damages, even in those cases in which the damages can be justified by market considerations. A purely economic view might support the imposition of punitive damages on a few randomly identified actors in amounts disproportionate to their fault, on the ground that their losses will be exceeded by the benefits of deterrence to others.[197] A deontological view will serve to limit this effect.[198]

BMW of North America, Inc. v. Gore[199] is the Supreme Court's most recent pronouncement on the constitutional limits of punitive damages, and it shows the heavy influence of deontological considerations.[200] Gore bought an automobile that had been repainted to

---

194. *Cf.* POSNER, *supra* note 59, at 224-25 (analyzing the disadvantages of using severe penalties in an attempt at perfectly deterring minor crimes).

195. *Cf.* Partlett, *supra* note 25, at 806 (arguing that retributive theory "looks to consequences and is to be so judged").

196. For a hierarchical view of the two conceptions, see Hurd, *supra* note 177, at 253-54. Professor Hurd sees consequentialist limits as general guidelines trumped by the more-specific principles of deontology: "[T]he principal payoff of deontological maxims is their ability to define and patrol the borders of consequential justification." *Id.* at 254.

197. *See supra* note 194 and accompanying text.

198. *See supra* note 182 and accompanying text.

199. 116 S. Ct. 1589 (1996).

200. *See id.*; *see also* Lee v. Edwards, 101 F.3d 805, 809-13 (2d Cir. 1996) (applying *BMW* to reduce the award of $200,000 in punitive damages against a police officer in a civil rights case on the ground that it was out of line with awards in more egregious cases); Continental Trend Resources, Inc. v. OXY USA Inc., 101 F.3d 634, 635-43 (10th Cir. 1996) (applying

correct minor pre-delivery damage, but BMW did not inform him of this arguable defect.[201] The repainting caused a diminution in value that the jury compensated by a $4 thousand award.[202] The jury also awarded $4 million in punitive damages, which the Alabama Supreme Court later reduced to $2 million.[203] The United States Supreme Court's opinion, rejecting this punitive award as violative of due process,[204] combined consequentialist and deontological reasoning. It did so by recognizing the state's affirmative interest in deterring undesirable conduct by economic means, a consequential goal, and by also crediting the interest in punishing unlawful conduct, a deontological purpose.[205] The Court also limited the state's power to punitive amounts that were not "grossly excessive."[206] The Court defined this limitation with three evidentiary factors, which in turn were partly consequentialist and partly deontological.

The first factor, which the Court identified as perhaps the most important, was the "degree of reprehensibility" of the defendant's conduct.[207] The Court's reasoning made this apparently deontological consideration depend upon several subsidiary factors, which were related more closely to moral blameworthiness than to utilitarian concerns. For example, the Court pointed out that BMW's conduct did not impair the car's performance and that the car manufacturer evinced no indifference to life, health, or safety; Gore's loss, in the eyes of the Court, was "purely economic."[208] Similarly, BMW acted neither in bad faith nor deliberately, and it avoided such blameworthy conduct as false statements, acts of affirmative misconduct, and concealment of evidence.[209] A consequentialist, in contrast, would not have concentrated upon the blameworthiness of BMW's mental processes or upon any categorical difference between "health and safety" and "economic" interests. Instead, the consequentialist would have inquired into the cumulative gains of the society from the deterrent and would have evaluated whether they exceeded the total costs. The gains would have been measured in economic terms, irrespective

---

BMW in a commercial case to reduce a $30 million award to $6 million), *cert. denied*, 117 S. Ct. 1846 (1997).

201. *BMW*, 116 S. Ct. at 1593.

202. *Id.*

203. *Id.* at 1594-95.

204. *Id.* at 1604.

205. *Id.* at 1595.

206. *Id.*

207. *Id.* at 1599.

208. *Id.* at 1592.

209. *Id.* at 1601.

of their origin in safety or aesthetics, and the costs would have been evaluated similarly in economic terms, independently of BMW's mens rea.[210] The Court's reasoning on the basis of "reprehensibility"[211] thus was largely deontological.

The second factor identified in the *BMW* case was the ratio between the plaintiff's actual loss and the punitive award.[212] This factor can be viewed in consequentialist terms: It is an effort, albeit an imperfect one, to increase the likelihood that the costs of the sanction will not exceed the benefits that it is expected to produce. Strictly speaking, the reasoning would not be valid unless all societal benefits were compared to all social costs,[213] but perhaps informational difficulties make this degree of accuracy seem so ambitious that they justify the rougher measure.[214] In the alternative, consideration of the actual-damages-to-punitive-damages ratio can be defended in deontological terms, because actual damages provide a measure—albeit an exceedingly rough measure—of the actor's moral blameworthiness. This conclusion rests on the assumption that just deserts depend upon a combination of guilty mind, wrongful conduct, and an injurious result that correlates, in turn, with the harm suffered by a randomly identified victim.

Finally, the Court considered the difference between this punitive award and civil or criminal sanctions available for comparable misconduct.[215] The Court apparently considered these sanctions both as a yardstick against which the tribunal should measure the present amount and as a means of notice that would enable the actor to predict the consequences of chosen courses of conduct.[216] Again, this factor is capable of being analyzed in deontological terms, in that the comparison to like conduct correlates with a concern for retributive proportionality. The existence of notice is evidence of a morally blameworthy mental state. However, the analysis also can be undertaken in consequentialist terms, in that notice is related to the effectiveness of the deterrent. Furthermore, the Court used the comparison to demonstrate that there was no reason to assume "that a more modest sanction would not have been sufficient,"[217] and this

210. *See supra* Part II.B.1.
211. *BMW*, 116 S. Ct. at 1599.
212. *Id.* at 1601.
213. *See supra* Part I.
214. *See supra* Part II.A.
215. *BMW*, 116 S. Ct. at 1603.
216. *See id.*
217. *Id.*

logic is consistent with economic reasoning, which seeks to assure that the cost does not exceed the expected gain.

BMW is a constitutional decision, and as such, it only sets limits upon the states' abilities to assess punitive damages.[218] It does not specify how the amount must be determined. Its deontological reasoning, therefore, is important as a means of restricting punitive damages at the margin, but it does not tell us the best way for a state to go about the specific business of computing them. In particular, it does not negate the possibility that consequentialist reasoning, such as the economic theory set out in this Article, may be the preferred method for calculating the award, with deontological considerations serving as a limit.[219]

C. *Synthesis: An Economic Approach, Limited by Informational and Deontological Concerns*

In summary, informational deficiencies and deontological criticisms arguably mean that an economic perspective on punitive damages should not be the exclusive focus. And yet there are justifications for heavily emphasizing economic analysis in formulating evidence rules for trials involving punitive damages.[220] In the first place, the sanction is itself meant to be experienced in economic terms. It often is imposed upon actors immersed in economically motivated activity, under circumstances in which the countervailing considerations are those of the marketplace.[221]

Furthermore, considerations of moral blameworthiness are readily arguable from the evidence of conduct at issue and are made palpable by the evidence of individual injuries. What are less tangible in such cases are the larger consequences upon the society as a whole, largely in economic terms, of which evidence is more diffuse and difficult to apply.[222] Precise deontological rules are likely to be lacking

---

218. *See id.* at 1602-03.

219. Perhaps this reasoning dovetails with the general relationship between the two approaches: Arguably, deontology best serves to limit consequential reasoning. *See supra* note 196.

220. The most carefully reasoned argument for consequentialism in the form of economic deterrence as the "principal or only basis" for punitive damages is probably Dan B. Dobbs, *Ending Punishment in "Punitive" Damages: Deterrence-Measured Remedies*, 40 ALA. L. REV. 831, 858-63 (1989). For a contrary view, advancing retributive theory as superior to both compensation and deterrence, see Partlett, *supra* note 25, at 800-06.

221. As has been indicated above, the theory in this Article best fits the economically motivated actor, as opposed to the actor who derives illicit utility from the infliction of harm itself. *See supra* note 34.

222. *See infra* Part III.B.3-4 (discussing the feasibility of admitting evidence about the deterrence gap or about profitability).

when it comes to production decisions, and the effort to derive them on an ad hoc basis in individual cases involving severe injuries has high potential to mislead. Thus, for example, it often is possible for an advocate to deride an efficient production decision as motivated by profits at the expense of lives,[223] so that a powerfully appealing deontological principle appears to be implicated.[224] In fact, however, the result of such reasoning may be erroneous. At some point, efficient production decisions are proper even if they carry unavoidable dangers of accident and injury. This must be the case if we are to avoid shutting down production altogether, everywhere, with the result that we would inflict on ourselves losses far exceeding those of the accidents the questioned activity could ever cause.[225] A consequentialist theory based on deterrence would distinguish this case from one in which punitive damages are appropriate.[226] Many of the wide variety of deontological claims that might be invoked, however, would not.[227] And yet, like the unqualified people-over-profits principle described above, deontological claims may resemble the song of the Sirens in their appeal.

A synthesis of these concepts, therefore, may warrant consideration of rules for punitive damages expressed primarily in terms of their economic validity. But the economic conclusions must be evaluated, in turn, against considerations of information deficiency and consistency with deontological concerns. For example, a given type of evidence may be correlated only roughly with measures of economic efficiency, but it may be superior to other types of evidence that might theoretically be considered more accurate, if the evidence can be

---

223. For a scholarly analysis of this and other simplistic logical constructs that furnish time-honored and effective (but often misleading) jury arguments, see generally Neal R. Feigenson, *The Rhetoric of Torts: How Advocates Help Jurors Think About Causation, Reasonableness, and Responsibility*, 47 HASTINGS L.J. 61 (1995).

224. The defendant, too, can easily mislead by false deontology. An example, which is just as far off the point as the plaintiff's denunciation of "profits," is the defendant who argues that a plaintiff claiming intangible or exemplary damages is "greedy." *Cf. In Summation, Simpson's Lawyer Decries "Character Assassination,"* HOUS. CHRON., Jan. 23, 1997, at A11, *available in* 1997 WL 6536379 (reporting O.J. Simpson's defense lawyer as arguing that the Goldman family's civil suit "isn't a fight for justice, it's a fight for money").

225. Thus, for example, the law of product liability recognizes that it is not tortious to sell dangerous products if they are not "unreasonably dangerous." *See* RESTATEMENT (SECOND) OF TORTS § 402A cmt. i (1965). Many products are unavoidably unsafe. *See id.* cmt. k.

226. *See id.* cmt. i. The concept of "unreasonable" dangerousness invokes the risk-utility calculus that permits the distribution of products that are unavoidably unsafe but nevertheless beneficial. *See id.*

227. *See generally* Feigenson, *supra* note 223 (describing jury strategies for finding liability based on the following notions: there must be a person to blame for every injury, bad acts are committed by bad people, and people who cause severe injuries are "more responsible" than those who cause lesser injuries).

readily determined, understood by juries, and communicated to other actors. Likewise, an economically efficient approach to evidence, but one that achieves deterrence without due regard to distribution of punishments according to moral blameworthiness, may be less acceptable than one that is less immediately efficient, but that squares better with the criterion of proportional justice. Thus, as some commentators have argued, economic-deterrent analysis should be the principal ingredient,[228] but as others have said, information deficiencies[229] and deontological concerns[230] should limit the calculation.

III. EVIDENCE TO DETERMINE THE AMOUNT OF A PUNITIVE AWARD: THE RULES AND THEIR APPLICATIONS

A. The Governing Rules of Evidence

There is no specific rule that governs evidence about the amount of punitive damages. For the most part, the decisions are interpretations of the most general principles of evidence contained in Federal Rules 401 through 403 or their state-law equivalents.[231] The structure of these rules gives each a particular function. First, Rule 401 defines relevant evidence in terms so broad that they arguably are all-inclusive. Evidence is relevant if it has "any tendency" to make a fact that is "of consequence to the determination of the action" more or less likely.[232]

Rule 402, in turn, makes relevant evidence admissible, unless it is excluded by another rule or principle.[233] It also provides that evidence that is not relevant is not admissible.[234] The breadth of the definition of relevance, however, is such that Rule 402 will admit almost anything in any case. Elsewhere, in fact, the author of this Article has argued that the literal terms of Rules 401 and 402 make them perfectly transparent, so that, if interpreted literally, they never ex-

---

228. *See supra* note 220.

229. *See supra* Part II.A and authorities therein cited. This reasoning should not, however, cause us to substitute an inaccurate measure for an accurate one merely because it can be calculated more readily. *See infra* Part III.B.1 (critiquing the use of net-worth evidence).

230. Professor Hurd's formulation of this role for deontology is well phrased. *See supra* note 196.

231. FED. R. EVID. 401-403. For a striking example, see *John Deere Co. v. May*, 773 S.W.2d 369 (Tex. App. 1989, writ denied), in which the court interpreted the Texas equivalents of these rules to uphold the admissibility of thirty-four separate accidents, with varying degrees of similarity to the event on trial, as relevant to punitive damages. *Id.* at 374.

232. FED. R. EVID. 401.

233. FED. R. EVID. 402.

234. *Id.*

clude any arguable proposition.[235] Even if the judge applies a pragmatic gloss, as most courts do, Rules 401 and 402 provide an exceedingly loose filter, which admits even evidence of marginal relevance.[236]

This conclusion, however, does not mean that all marginally relevant evidence will be admitted. It means only that the task is done instead by exclusionary rules. Foremost among these exclusionary principles is Rule 403, which provides that relevant evidence is not admissible if its probative value is "substantially outweighed" by certain counterweights, which include prejudice, confusion, and waste of time, among others.[237] Even this rule, however, is "loaded": It is biased in favor of admissibility. The mere existence of prejudice will not exclude. Even the presence of a substantial amount of prejudice will not exclude. Furthermore, prejudice will not exclude even if it counterbalances or even exceeds the probative value. It is only when prejudice or other counterweights "substantially" outweigh the probative value of an item of evidence that it is excludable under Rule 403.[238]

Other rules may come into play with regard to punitive damages, but only as to narrow classes of evidence. Rule 407, for example, excludes subsequent remedial measures offered to prove negligence or culpable conduct.[239] Although such evidence may be admissible for other purposes,[240] this exclusion probably disallows the use of subsequent remedial measures to determine liability for punitive damages, and probably the amount as well.[241] Rule 411 excludes evidence of liability insurance to prove fault.[242] Again, however, the exclusion does not apply if the evidence is offered for another purpose.[243] There have been suggestions, for example, that the jury should know

---

235. *See* David Crump, *On the Uses of Irrelevant Evidence*, 34 HOUS. L. REV. 1, 9-14 (1997).

236. *See id.* at 14-17.

237. *See* FED. R. EVID. 403.

238. *See id.*; *see also* Crump, *supra* note 235, at 17-20.

239. *See* FED. R. EVID. 407.

240. *See id.* (allowing evidence of subsequent remedial measures for other purposes, including "proving ownership, control, or feasibility of precautionary measures, if controverted, or impeachment").

241. This arguably is so, because the tendency of the evidence to demonstrate "culpable conduct" would overwhelm any other use in connection with punitive damages. *See id.* (excluding evidence of subsequent remedial measures proving "negligence" or "culpable conduct"); *cf.* FED. R. EVID. 403 (excluding relevant evidence that is "substantially outweighed" by factors such as "unfair prejudice").

242. *See* FED. R. EVID. 411.

243. *See id.* (allowing evidence of liability insurance coverage for other purposes, including "proof of agency, ownership, or control, or bias or prejudice of a witness").

about insurance when assessing the amount of punitive damages, because punitive damages should hurt, or punish, the defendant.[244]

Application of all of these rules depends upon identification of the underlying issues. Relevant evidence depends upon an inference that is tied to an issue that is "of consequence to the determination of the action."[245] Prejudice, in turn, is best defined as the degree to which the evidence "suggest[s] decision on an improper basis."[246] It is here that the analysis of evidence in punitive-damage cases is weakest because of the lack of consensus about the reasons for imposing them. As observed at the beginning of this Article, courts and legislatures lack a coherent vision of the function they want punitive damages to serve, and therefore, they have no clear goal toward which they can target their standards.[247]

B. Common Evidentiary Approaches to Determining Punitive Damages: A Critique

1. Net-Worth Evidence: Is It, in Fact, Worthless?—Many jurisdictions admit evidence of the defendant's wealth as a basis for determining the amount of a punitive award.[248] It is illustrative of the confusion of goals that some other jurisdictions have excluded this evidence on the ground that it is misleading and prejudicial.[249] Strangely, the National Conference of Commissioners on Uniform State Laws has approved a *Model Punitive Damages Act* that specifically invites the jurors to consider the defendant's "financial condition," among other factors.[250]

This kind of evidence, however, ordinarily has little to do with either economic goals of punitive damages or with deontology, and it should be excluded. Net-worth evidence does not support the economic function of punitive damages when viewed from an industry-wide perspective. This function, as illustrated in Figure 3 above, is to

---

244. *See* PAUL F. ROTHSTEIN, EVIDENCE IN A NUTSHELL: STATE AND FEDERAL RULES 16 (2d ed. 1981) (stating that a defendant's ability to pay is relevant where punitive damages are requested "because such damages should have sting").

245. FED. R. EVID. 401.

246. FED. R. EVID. 403 advisory committee's note.

247. *See supra* notes 25-33 and accompanying text.

248. *See, e.g.*, Life Ins. Co. v. Johnson, 684 So. 2d 685, 702 (Ala. 1996), *vacated on other grounds*, 117 S. Ct. 288 (1996); Wayte v. Rollins Int'l, Inc., 215 Cal. Rptr. 59, 72 (Ct. App. 1985); Lunsford v. Morris, 746 S.W.2d 471, 476 (Tex. 1988).

249. *Cf.* Annotation, *Punitive Damages: Relationship to Defendant's Wealth as Factor in Determining Propriety of Award*, 87 A.L.R.4TH 141, 181-84 (1991 & Supp. 1996) (listing cases in which courts have held that the wealth of the defendant could not be considered by the finder of fact in arriving at its award of punitive damages).

250. *See* MODEL PUNITIVE DAMAGES ACT § 4(a) (1996).

combine with compensatory damages to provide the proper level of deterrence for harmful conduct. Punitive damages should not be either so small that they leave a deterrence gap, as in Figure 2, nor so excessive that they result in overkill, as in Figure 4. In the latter situation, consumers find that purchases are more difficult and more expensive to complete, even in the case of crucially necessary products.

Perhaps an argument favoring net-worth evidence can be constructed concerning the risk-adverseness of individual actors in the marketplace. A person's degree of risk aversion may change with levels of wealth. Thus, we tend to suspect that the marginal utility of wealth is inversely related to its absolute amount, or in other words, that the utility of an additional dollar declines as a person becomes more wealthy.[251] To illustrate this particular case, Professor Shavell provides a diagram identical to Figure 7, in which there is a concave graph of utility versus wealth.[252] In such a case, the argument could be asserted that a greater sanction must be imposed upon a wealthy individual to result in a degree of deterrence equivalent to that imposed by a smaller penalty upon a person of modest means.[253]

This argument has more potential to mislead, however, than to justify such a conclusion. If punitive damages are assessed against a publicly held corporation or other distinct business entity, as seems typical of such cases,[254] the graph may not have this concave shape. Given their specifically economic purpose, such entities may be more consistent throughout the range of wealth in their attitude toward risk, or risk neutrality.[255] Insurance, even for individuals, may have a

---

251. *See* SHAVELL, *supra* note 34, at 187 & n.2.

252. *See infra* fig.7. This figure was reprinted from SHAVELL, *supra* note 34, at 188 fig.8.1.

253. Professor Shavell does not make this argument, and his reasoning explains its fallacy. He points out that the graph of marginal utility must be determined by knowing a person's actual degree of risk aversion at every level. *See* SHAVELL, *supra* note 34, at 187 & n.2. In other words, the graph must be derived, rather than simply assumed to be concave. *See id.*; *accord* Chapman & Trebilcock, *supra* note 25, at 802 (pointing out that we cannot with any confidence assume that a $10,000 loss means less to a wealthy person than to a poor individual (citing Milton Friedman & L.J. Savage, *The Expected-Utility Hypothesis and the Measurability of Utility*, 60 J. POL. ECON. 463 (1952))).

254. *See infra* note 268 and accompanying text; *see also* POLINSKY & SHAVELL, *supra* note 52, pt. IV (discussing the punishment of firms in deontological terms). Polinsky and Shavell raise questions about the appropriateness of punishing firms, as opposed to punishing culpable individuals within them; about the difficulty of showing that the punishment of a firm actually will punish the individuals who are, in fact, culpable; and about the likelihood that the punishment actually will hurt innocent individuals, such as shareholders or customers. *Id.* at 95-96.

255. *See* SHAVELL, *supra* note 34, at 189 (reasoning that a tendency toward this attitude is likely, but noting that risk aversion for an individual firm depends upon communication among shareholders, managers, and employees, regarding their respective attitudes).

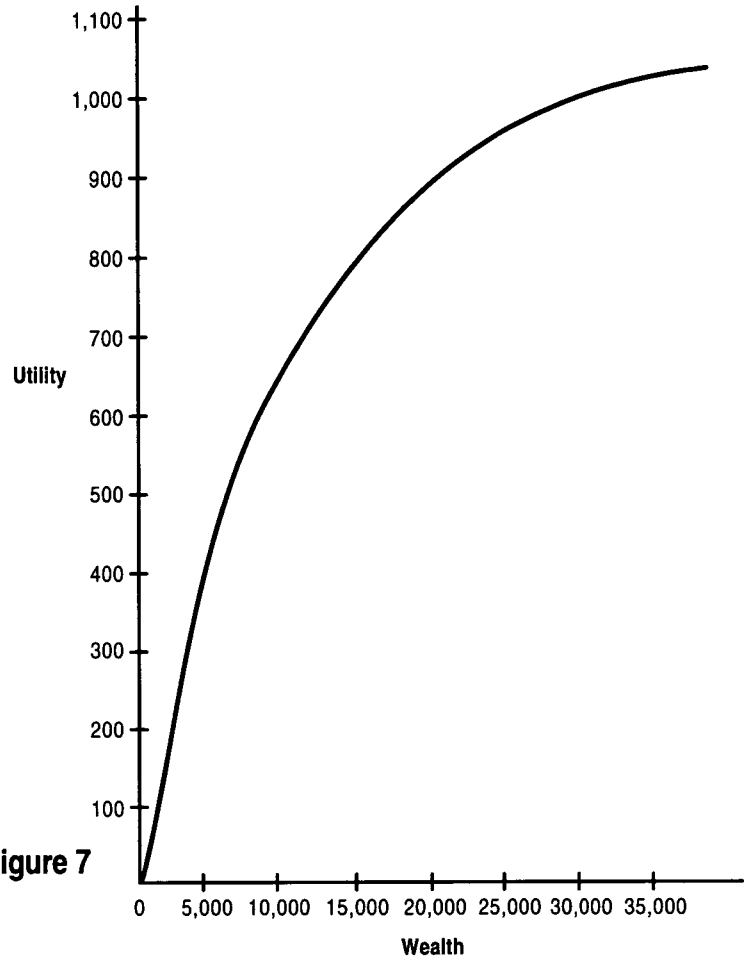

Figure 7

from ECONOMIC ANALYSIS OF ACCIDENT LAW by Steven Shavell

similar theoretical effect. Furthermore, the declining-utility argument does not take into account the possibility that wealthy persons or firms might engage in more transactions that would subject them to risk.[256] Thus, if a $2 million corporation is subjected to twice the penalty for each violation as its $1 million competitor, but the penalty is assessed in twice as many cases because the $2 million firm has engaged in twice as many transactions creating liability, the result may be an excessive reaction to the decline in marginal utility of wealth. This situa-

256. *See* Chapman & Trebilcock, *supra* note 25, at 803 (explaining that "radically inequitable results are likely to be generated by weighing corporate defendants' wealth in assessing punitive damages," and giving a similar example).

tion, then, would produce not only the classic kind of punitive-damage overkill in which the defendant is given multiple punishments, but also an overkill due to the misleading use of wealth in determining punitive awards.

More importantly, the argument based on the marginal utility of wealth attempts to prove too much. The graph shows a person of zero net worth as having the highest marginal utility of wealth, and this is true in the sense that the pauper is more excited about receiving a dollar than is the millionaire.[257] But if the argument were extended indiscriminately to draw conclusions about responsiveness to a large punitive-damage award, it would produce absurd results. The dubious conclusion, then, would be that the judgment-proof defendant fears a large liability more than the person of means. In this topsy-turvy scenario, a person with nothing to lose has the most to lose.

There is no reason to suppose, however, that the net worth of an individual or firm has any fixed relationship to the frequency of its misconduct in the marketplace, or to the deterrence gap. Such a correlation would depend upon the theory that those who succeed in achieving a positive financial condition tend to do so by dishonest or predatory practices. It seems more likely that some successful persons and firms are successful because they operate properly in the marketplace, while others do not.[258] Indeed, one might as easily infer that marginal competitors (i.e., those in poor financial conditions) are the most likely to cut corners by externalizing costs. This reasoning would lead one to assess punitive damages in inverse proportion to net worth, or in exactly the opposite way from the Uniform Commissioners' approach.[259] Both of these approaches are irrational.

In fact, one way to view the irrationality of reliance on net worth is to consider a very large entity that becomes insolvent. In recent years, large entities such as national department stores have filed petitions in bankruptcy.[260] Imagine that such a bankrupt, but huge, corporation were to engage deliberately in a fraudulent or extremely

---

257. *See supra* fig.7. This actually is the meaning of Professor Shavell's diagram. *See supra* notes 251-253 and accompanying text.

258. Thus, for example, it seems possible that a well-to-do individual whose upbringing stressed risk aversion, honesty, frugality, and respect for others would be more heavily deterred by the prospect of a given punitive-damage award than a person of marginal means, especially if the latter person could not begin to pay the award.

259. *See* MODEL PUNITIVE DAMAGES ACT § 7 (1996) (allowing the admission of net-worth evidence).

260. *See, e.g.*, Richard D. Hylton, *A Crisis Built on Debt; Despite Its Strong Reputation in Retailing, Macy's Choices in the 80's Left It Crippled*, N.Y. TIMES, Jan. 27, 1992, at A1, *available in* LEXIS, News Library, Nyt File (reviewing management missteps that led Macy's to bankruptcy).

dangerous course of conduct as a means of ameliorating its insolvency. It would be counterproductive, then, to enable the defendant to escape punitive-damage liability (or for that matter, to impose "negative" punitive damages, reducing the compensatory award),[261] but this is exactly what the theory of net-worth evidence ostensibly calls for.

In deontological terms, also, the admittance of net-worth evidence seems irrational. Evidence of the defendant's wealth would be excluded if it were offered in virtually any other kind of case.[262] The usual rule is that wealth or poverty, like such factors as ethnicity, gender, and age, should not make a difference in adjudication of substantive rights.[263]

Perhaps it can be argued that punitive damages are different from other punishments.[264] By this theory, the wealthy defendant who misbehaves would be less severely punished by a given fine than the marginal one who engages in the same conduct. It would follow, then, that a larger fine must be imposed on a wealthy defendant to achieve the same level of punishment. Alternatively, the wealthy defendant may be considered more reprehensible than the marginal defendant. But once again, it is unclear why a bankrupt (but very large) department store, for example, should be treated more favorably when it engages in the same conduct as an industry leader.

Finally, there is the argument that punitive damages should suffice to "stop" the defendant's conduct or "get the defendant's attention." This, however, is an economic argument masquerading as deontology; it is about deterrence. And the deterrent function of punitive damages occurs due to the underdeterrence, or gap, that is created when compensatory damages are not perfectly recoverable for all

---

261. This reasoning assumes that if large positive net worth mandates adding to the compensatory award, large negative net worth equivalently justifies subtracting from it. This odd result is insupportable but follows the same logic.

262. *Cf.* Brokopp v. Ford Motor Co., 139 Cal. Rptr. 888, 899 (Ct. App. 1977) (holding that "[a]ppeals to the sympathy of the jury based on the size or corporate status of a defendant are improper," but denying reversal because of the absence of an objection).

263. *See id.*

264. *See* Chapman & Trebilcock, *supra* note 25, at 801-02 (recognizing a "plausible argument" for differentiating punitive damages from other retributive sanctions that are objectively equal despite wealth—e.g., incarceration—on the ground that monetary penalties, in order to be equivalent, require distinctions on the basis of subjective attitudes toward money). It should be noted that Chapman and Trebilcock reject the argument as valid "in principle" but unworkable "as a practical matter." *Id.* at 802.

losses.[265] Again, there is no reason to suppose that this gap has any close correlation with net worth.[266]

Instead, net-worth evidence seems to inject a significant amount of prejudice, particularly if prejudice is defined as the "suggest[ion of] decision on an improper basis."[267] If net-worth evidence is received in a unitary trial on liability, damages, and punitive damages, there is likely to be a tendency toward jury misuse of this evidence to impose liability and actual damages upon a defendant perceived as wealthy.

Recently, the California research firm Metricus conducted a nationwide study, polling people eligible for jury duty. The findings include the following: (1) Even before knowing anything else about the dispute, seventy percent were more likely to favor an individual over a corporation; (2) respondents were "much more likely" to believe an accusation that a defense lawyer was not telling the truth than the same accusation about a plaintiff's lawyer; and most significantly, (3) "60% of [potential] jurors deemed a $1 million judgment 'just a slap on the hand' for a corporation."[268]

Some jurisdictions have addressed this problem by bifurcating the trial: If, in the first trial, the jury finds liability for punitive damages (for example, with a gross negligence finding), a second trial is held in which the defendant's net worth is exposed.[269] Normally, however, jury selection and opening statements must forecast the issues for the jurors. It would be difficult, in other words, to avoid having the jurors focus upon net-worth issues early in the trial. The bifurcation procedure, in fact, has resulted in some trials in which the trial lawyers have concluded that the jurors speculated about the defendants' net worths erroneously, believing that they were much larger than they actually were; the result was prejudice just the same.[270]

And even if the prejudice could be cabined off, so that it does not affect the findings of liability and actual damages, there remains the problem that assessing punitive damages on the basis of wealth is itself

---

265. *See supra* Part I.B-C.

266. *See supra* notes 253, 258 and accompanying text.

267. *See supra* note 246 and accompanying text (explaining the definition of prejudice according to the Federal Rules of Evidence).

268. Stephen J. Adler, *Corporations Face Uphill Struggle in Jury Trials Involving Individuals*, WALL ST. J., Nov. 13, 1991, at B5, *available in* 1991 WL-WSJ 584290.

269. *See supra* note 9 and accompanying text.

270. *See, e.g.*, Mary Hull, *Early Verdict: Moriel Playing No Favorites*, TEX. LAW., Sept. 12, 1994, at 1, *available in* LEXIS, News Library, Txlawr File (describing the largest amount of actual damages awarded in Travis County, Texas, in a case which prohibited evidence of the defendant's net worth being heard until after a finding of gross negligence).

a "decision on an improper basis."[271] For the reasons given above, neither economic nor deontological reasoning supports the use of this factor. Furthermore, this factor seems dubious in light of the reasoning engaged in by the sixty percent of potential jurors who would consider a $1 million judgment against a corporation as a "slap on the hand."[272] Instead, whether it is a slap on the hand is more closely related to the size of the deterrence gap.

2. Caps, Limits, and Ratios.—Some jurisdictions limit punitive damages by absolute caps: The award may not exceed a defined dollar amount.[273] Others impose limits by ratios: The award may not exceed an arithmetic multiple, defined by statute, of the actual damages.[274] Still other jurisdictions have enacted combinations of these two approaches or have used other factors, such as the lesser of a fixed dollar amount or the defendant's gross annual income, as a cap.[275]

These caps and ratios have the potential to be irrational for reasons similar to the irrationality of net-worth evidence, but in the opposite direction. There is no reason to suppose that the deterrence gap will correspond to a particular ratio of actual damages. If the claim is readily detectable and easily prosecutable, then the gap will be a fraction of the damages; if the defendant's conduct, on the other hand, is difficult to discover, and involves expensive, high-risk litigation, then a ratio of two to one, for example, may be inadequate to address the deterrence gap. An absolute-dollar figure may be equally nonsensical, if not more so. Clandestine conduct that causes significant damage may result in a deterrence gap that exceeds the absolute-dollar cap.[276]

Perhaps, however, ratios are not as irrational as either absolute caps or net-worth evidence. They at least relate punitive damages to actual harm, even if they do so only in specific cases. Perhaps econometrics could tell us whether a multiple of two, three, or four would be appropriate in the economy overall as an approximation of

---

271. *See supra* note 246 and accompanying text.

272. *See supra* note 268 and accompanying text.

273. *See supra* note 13.

274. *See supra* note 14.

275. *See* Hurd & Zollers, *supra* note 25, at 195 (describing different variations of caps and multipliers used in legislative tort reform).

276. Thus, the prefatory note to the *Model Act* implies that these devices may be "arbitrary." *See* MODEL PUNITIVE DAMAGES ACT prefatory note para. 20 (1996) ("The Drafting Committee felt that it could improve upon the procedure, burden of proof, judicial review, and similar matters so that *arbitrary* monetary limitations may not be necessary." (emphasis added)).

the deterrence gap.[277] Statutes capping punitive damages by multiples may be based upon guesstimates of this computation. Even so, a ratio approach would undercompensate for the gap in some kinds of cases, such as those involving repetitive conduct producing small damages,[278] while overcompensating for the gap in others, particularly those where the conduct is discoverable and results in high actual-damage awards.[279]

Perhaps ratios or caps make more sense in deontological terms. The most cherished horror stories about punitive damages have to do with punitives so excessive that they seem unrelated to actual harm (e.g., hundred-to-one or greater ratios).[280] Even if the deterrence gap is large because the conduct is repetitive and causes small injuries, hundred-to-one awards engender disrespect for the law.[281] They offend the concepts of proportionality and retributive justice.[282]

In summary, punitive-damage caps, particularly by way of ratios, can be defended on deontological grounds and, to a lesser degree, by economic reasoning. Their defensibility, however, depends upon their functioning as upper limits, rather than as routine caps. In other words, these figures should be set high enough so that they do not impair the function of punitive damages in filling the deterrence gap.

3. Evidence of Underdeterrence from Economic Models—or of the Likelihood of Escape from Liability.—In theory, one might speculate that the

277. This computation would be accomplished by comparison of the statewide total of actual damages recovered, with estimates of the total amount of actual harm caused by tortious behavior during the same period, including that which was not compensated.

278. A typical example is that of an insurer that strategically denies small claims in bad faith because it believes that most are too small to litigate, so that the rare loss due to a compensatory award (perhaps even with small-cap punitives) is overwhelmed by unlawful gains. *Cf.* Dobbs, *supra* note 220, at 866 (using the example in a slightly different context to show how extra compensatory damages would continue to be awarded under a deterrence-measure system of punitive-damage awards).

279. That is, it will overcompensate if the ratio or cap is the principal limit and if other, misleading guidance (such as net-worth evidence) is all that directs the jury.

280. *Cf.* BMW of N. Am., Inc. v. Gore, 116 S. Ct. 1589, 1603 (1996) (noting that "[w]hen the ratio is a breathtaking 500 to 1, however, the award must surely 'raise a suspicious judicial eyebrow'" (quoting TXO Prod. Corp. v. Alliances Resources Corp., 509 U.S. 443, 482 (1993) (O'Connor, J., dissenting))).

281. For example, the famous McDonald's coffee case arguably featured a large deterrence gap, and yet the size of the award engendered disrespect for the law. *See supra* note 2. It is plausible to maintain that this disrespect should not have resulted in reduction of the award, because it may not have been based on a full public appreciation of the evidence. The trial judge's response was to find that McDonald's was "willful" and "reckless," to uphold the punitive finding, but to remit most of the amount. *See supra* note 2.

282. *See supra* Part II.B.

best kind of evidence to determine a punitive award would come from economic measurement of the deterrence gap. Expert witnesses would attempt to convey to the jury the information contained in Figure 2, above, as applied to the specific case. This would require an economist to testify concerning his estimates of the typical firm's marginal cost curve, as well as the marginal cost curve that results when actual damages are added. Then, this expert would estimate—and perhaps sketch for the jury—the social marginal cost curve. The dollar value of the deterrence gap would emerge from this process. If the estimates were accurate, the punitive award would be economically efficient,[283] and it also would serve deontological purposes,[284] provided that its effects could be adequately communicated to the jury and the public.

The theoretical accuracy of this methodology, however, is matched only by its empirical indeterminacy. As is indicated above, marginal cost curves are difficult and expensive to derive; so too are projections of actual damages.[285] And for the social marginal cost curve, dollar estimates must be placed upon aesthetics, environmental preservation, tolerance of risks, and for that matter, human life itself.[286] And all of this indeterminacy is aside from the issue of the admissibility, and the partisanship, of expert testimony.[287] One would expect plaintiffs' and defendants' estimates of each of the relevant factors to differ wildly. A wide variety of damage models might be arguable for each of the ingredients. The trial judge would face significant difficulties in determining, pursuant to *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,[288] the "reliability" and "fit" of the testimony thus proffered.[289] In summary, this kind of evidence is the most accurate and desirable in theory, but it often would be difficult to present in practice.

As an alternative to the kind of diffuse evidence that would measure underdeterrence economically, Polinsky and Shavell argue that the punitive award should depend upon the likelihood that the de-

---

283. *See supra* Part I.B.

284. The deontological purpose would be served because the punitive award would be related to the wrongfulness of the conduct. *See supra* Part II.B.

285. *See supra* Part II.A.2.

286. *See supra* Part II.A.1.

287. *See* Stan V. Smith, *Pseudo-Economists: The New Junk Scientists*, 47 FED'N INS. & CORP. COUNS. Q. 95, 95-103 (1996) (arguing that economists as witnesses may be prone to lack of qualification and to biased economic assessments).

288. 509 U.S. 579 (1993).

289. *Id.* at 597. *See generally* G. Michael Fenner, *The* Daubert *Handbook: The Case, Its Essential Dilemma, and Its Progeny*, 29 CREIGHTON L. REV. 939 (1996) (thoroughly reviewing the issues raised by *Daubert*).

fendant might have escaped having to pay for the harm for which she should be responsible.[290] Indeed, they argue that the estimation of this factor should be the principal task of the jury in fixing the amount, with a resulting multiplier supplied to the jury in a table, to be applied to the actual damages to compute a base amount that then would be adjusted by other specified factors. This factor, the likelihood of escape, suggests that several kinds of evidence might be relevant. For example, other instances of similar conduct that escaped detection are circumstantial indicators of the probability of nondetection in this case. The presence or absence of persons or agencies that would cause detection is another indicator. Furthermore, efforts by the defendant to conceal the conduct would be circumstantially relevant, as would openness by the defendant, to the opposite effect. Polinsky and Shavell's system is comprehensive enough, and yet precisely targeted enough, so that it may provide a sound means of guiding the jury toward a rough estimate of the deterrence gap.

4. "Profitability" Evidence.—As an alternative to evidence about the deterrence gap, perhaps the profitability of the conduct at issue can be examined in some kinds of cases.[291] Perhaps it could be determined more readily than could the deterrence gap,[292] although it still would involve evidence from experts based upon models, with all of the indeterminacy that these two factors entail.[293]

On the other hand, some cases may not be fit for profitability evidence.[294] An accident, after all, is accidental; it may involve aberrational conduct. If a particular employee, for example, embarks upon a course of embezzlement that is not detected by the firm's account-

---

290. POLINSKY & SHAVELL, *supra* note 52, at 8-9, 24-34. The authors propose a set of jury instructions telling jurors that their "principal" task is to estimate this probability of escape. *Id.* at 104-09. The multiplier is defined as $(1-P)/P$, where P is the probability of liability (and $1-P$ therefore is the probability of escape). Thus, the multiplier ranges from 0 when the probability is 0, and 0.11 when the probability is 0.1, up to 4.0 when it is 0.8, and 9.0 when it is 0.9. *Id.* at 110.

291. *See* Dobbs, *supra* note 220, at 868-88 (suggesting that the defendant's profit may be the "best possible measure of extracompensatory damages . . . [although] it is not the only one").

292. For example, it would not entail nearly so much indeterminacy caused by incommensurate social values. *See supra* Part II.A.2.

293. On the one hand, economic modeling is admissible as relevant to many kinds of issues, ranging from liability questions for claims such as antitrust violations to computation of compensatory damages. *See supra* note 81 and accompanying text. On the other hand, determining the profitability of a tortious act requires judgments about allocation to legitimate or illegitimate factors. *See supra* Part III.B.3. These judgments may be particularly sensitive to partisan manipulation. *See supra* note 287.

294. *See generally* Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1 (1991) (involving a misappropriation of insurance premiums and a finding of fraud).

ing controls, it may be difficult to estimate the "profitability" of the course of conduct to the firm.[295] In this event, it is unlikely that the firm itself has profited; its accounting controls were set up precisely to prevent the event. Estimating the profitability of using this particular system of accounting controls, as versus that of another, would be better than pure profit evidence, but this factor may be difficult to measure and seems far removed from the deterrence gap for which it is supposed to serve as a surrogate.

Also, "profitability" is not easy to define, and it readily can be misunderstood. It has more obvious meaning in the case of an economic actor, as opposed to one who acts on non-market motives, such as a sexual harasser or a murderer.[296] Furthermore, this kind of reasoning should not be taken as a reason to penalize a firm simply because it has shown a profit overall. Profit, or payment for entrepreneurship, management, and use of capital, is an essential part of the market system.[297] It is not a sign that the firm is doing something wrong; instead, it is more often a sign that the firm is doing something right. Assessing punitive damages in whopping amounts merely because the firm earned a twenty-percent profit last year, or in small amounts because it earned two percent, would be as irrational both economically and deontologically as basing the amount on its net worth.[298]

Furthermore, it is equally irrational to base punitive damages on the profitability of the activity or product that was associated with the damage. Usually, this factor will involve a socially desirable good or service: the providing of medicines, the selling of banking services, or the serving of coffee. Thus, for example, it was fallacious (or at least imprecise) for plaintiff's counsel in the McDonald's coffee case to urge a determination of punitive damages based on a multiple of several days' worth of coffee sales, or even coffee profit for that matter.[299]

---

295. *See id.* The profit approach in such a case may fairly measure punitive damages for an employee, who acted intentionally, but not for the company whose liability is based either upon inadvertence or upon a non-fault theory such as respondeat superior.

296. Thus, the punitive damages in such a case might better be assessed on the basis of the more nebulous but arguably better targeted factors in the *BMW* case. *See supra* Part II.B.2.

297. *See* SAMUELSON & NORDHAUS, *supra* note 52, at 660-61, 745-47 (discussing how profits influence factors of production and total product).

298. *See supra* Part III.B.1.

299. *See Damage Award Cut in McDonald's Case, supra* note 2. Given the crude condition of current evidence law governing this issue, however, it is arguable that counsel may have been operating at the state of the art in so arguing. *See generally* S. Reed Morgan, *McDonald's Burned Itself*, TEX. LAW., Sept. 12, 1994, at 18, *available in* LEXIS, News Library, Txlawr File (offering the perspective of the plaintiff's lawyer who tried the case).

Selling coffee was not what McDonald's did that arguably was wrong, and this measure had only a distant relationship to the deterrence gap.

Instead, profitability inquiries would need to be more carefully focused, if their probative value is to avoid being eclipsed by their tendency to prejudice or mislead. The issue, at minimum, should be narrowed to the gain in productivity achieved by the arguably wrongful act of McDonald's of overheating its coffee. This act, not the selling of coffee, was the legitimate focus of the deterrence function of the tort system.[300] Determining the answer to this more narrow inquiry may require more sophistication, and therefore more complexity, than sales or profitability figures alone, but in Rule 403 terms, the gain in probative value and the avoidance of prejudice are worth the greater complexity.[301]

If these precautions are observed, profitability may be a valid factor in some cases. It particularly would be useful in cases involving repetitive conduct with small but significant damages. The fast-food restaurant that serves dangerously heated coffee and externalizes the cost of injuries, for example, can estimate the cost savings related to this conduct and the arguably decreased sales that would result from underheated coffee. The plaintiff, through an expert witness who considers data provided by the fast-food firm, can make an alternate estimate. The result often would be correlated with underdeterrence, particularly if combined with evidence of the number of resulting injuries. Although the correlation would be inexact, it would be practical to produce this evidence before a jury.

5. Evidence of Other Conduct of the Defendant—or of Other Recoveries, Actual or Punitive, for the Same Conduct.—One particularly striking case involving other-conduct evidence is *John Deere Co. v. May*,[302] in which the plaintiffs obtained judgment in a product liability action against the John Deere Company, whose bulldozer ran over and killed the plaintiffs' decedent.[303] The plaintiffs offered evidence of thirty-four other incidents in which John Deere bulldozers allegedly had shifted into gear while standing in neutral with their engines running.[304] The appellate court upheld the admissibility of this evidence, in part as

300. *See supra* Part I.A.1.

301. FED. R. EVID. 403 (excluding evidence if its relevance is "substantially outweighed" by counterweights).

302. 773 S.W.2d 369 (Tex. App. 1989, writ denied).

303. *Id.* at 371.

304. *Id.*

relevant to punitive damages: "Furthermore, they [plaintiffs] could recover exemplary damages if they proved that John Deere knew of the greater danger but acted with conscious indifference."[305]

The issue in *May* was primarily whether the evidence was "substantially similar"[306] to the event on trial, and whether it passed the Rule 403 test.[307] The evidence was admitted primarily for the purpose of determining liability.[308] However, if admitted to determine recoverability of punitive damages, it presumably also would be relevant to determination of the amount. In some cases, particularly where a regular pattern of injuries from the course of conduct can be derived, this evidence, together with profitability evidence, can be useful.[309] For example, the fast-food firm that serves super-heated coffee arguably is taking advantage of a deterrence gap if its product has caused numerous injuries, similar to the one on trial, without having to pay damages and without changing its course of conduct.[310]

But repetitive-accident evidence is prejudicial. As the *May* court put it, "Extraneous incidents . . . can be extremely harmful to the defense."[311] When such evidence is submitted on punitive damages, either for liability or for amount, the court should take care to ensure that the incidents are substantially similar or relate to the same course of conduct.[312] The court also should undertake a Rule 403 balancing before admitting the evidence.[313]

A separate issue relates to the availability of credit against punitive damages for other compensatory or punitive awards paid.[314] Such evidence is relevant because it reduces the size of the deterrence gap in economic terms; it also is prejudicial, however, because it injects

---

305. *Id.* at 373.

306. *Id.* at 372.

307. *Id.* at 373-74 (citing TEX. R. CIV. EVID. 403, which states that relevant evidence may be excluded on special grounds); *see also supra* notes 237-238 and accompanying text (discussing FED. R. EVID. 403).

308. *May*, 773 S.W.2d at 377-79.

309. *See id.* at 374 (reasoning that proof of one hundred other occurrences, in addition to the documented thirty-four instances, would have increased the probability that the death of plaintiff's decedent occurred as plaintiff alleged).

310. *See supra* note 2 and accompanying text.

311. *May*, 773 S.W.2d at 374.

312. *See id.* at 372-73 (articulating "[w]hat constitutes reasonably similar circumstances under the facts presented").

313. *Cf. id.* at 373-74 (applying TEX. R. CIV. EVID. 403 to determine whether the relevant evidence should be excluded because its probative value is "substantially outweighed" by the danger of unfair prejudice that would result from its admission); *see also supra* notes 237-238 and accompanying text (discussing FED. R. EVID. 403).

314. *See* MODEL PUNITIVE DAMAGES ACT § 10 (1996) (allowing a court to credit any judgment sought to be enforced if it determines that the judgments are unfairly duplicative).

the fact of other verdicts or settlements against the defendant. There are several ways in which this problem could be addressed. First, the trial before the jury could be bifurcated, with the issue of credit to be determined in the latter hearing.[315] A second and probably better approach, however, would be to have the credit determined by the court alone. In some instances, the amount of the credit may be determinable as a matter of law.[316] When it is not so determinable, the credit involves the kinds of calculations a court of equity would make, and therefore, it should be done by a judge without the intervention of a jury.[317] Finally, the credit issue is closely related to the judge's review of the sufficiency of evidence to uphold the verdict, both for judgment as a matter of law and remittitur purposes.[318]

C. *Evidence for the Jury, Rather Than Mere Law for the Court*

Many other kinds of evidence might be arguably relevant to the economic and moral purposes of punitive damages. For example, evidence of efforts to conceal the wrongful course of conduct, or destruction of evidence, has become admissible in some cases.[319] Broad objectives such as the need to "punish" or "deter," or global characterizations such as the "reprehensibility" of the defendant's conduct, are used in some of the cases.[320] Presumably, these issues are related to such factors as the magnitude of expected losses, the profitability of the wrongful conduct, and the deterrence gap.

What is striking, however, is that sometimes jurors are given virtually no information that is closely associated with the deterrence gap when asked to determine punitive awards.[321] They may be provided little in the way of instructions about the appropriate range of

---

315. *See supra* note 9 and accompanying text.

316. This would be the case, for example, if the number and size of other instances to be credited is mathematically determinable, and the overlap of the punitive awards is not genuinely at issue.

317. If, for example, the different situations involve similar but not identical conduct, the punitive-credit considerations resemble those in *Tull v. United States*, 481 U.S. 412 (1987), which held that although the water pollution penalty case was subject to jury-trial right, the right did not override Congress's determination that the judge alone should compute the penalty. *Id.* at 425-27.

318. *Cf.* MODEL PUNITIVE DAMAGES ACT § 10 cmt. (1996) (explaining the requirements of the trial court in attempting to decide whether a reduction or credit should be granted).

319. *See* BMW of N. Am., Inc. v. Gore, 116 S. Ct. 1589, 1596 (1996) (reviewing evidence of BMW's nondisclosure policy regarding refinished cars in the assessment of the punitive award).

320. *See, e.g., id.* at 1599 ("Perhaps the most important indicium of the reasonableness of a punitive damages award is the degree of reprehensibility of the defendant's conduct.").

321. *See infra* notes 329-333 and accompanying text.

amounts.[322] Even when they are given instructions, they frequently are not given particularized evidence focused on the amount of punitive damages (except the irrational factor of net worth).[323] As the preceding sections suggest, the use of some sort of evidence that serves to determine the relationship of the award to economic and deontological purposes—such as ratios, profitability of the wrongful conduct, and repetitiveness and seriousness of injuries in other cases—probably would improve the jury's function.[324] Many of these types of evidence can and should be presented in the latter stage of a bifurcated trial, only after the jury has found liability for punitive damages under correct legal standards.

In such cases as *BMW of North America, Inc. v. Gore*[325] and *Pacific Mutual Life Insurance Co. v. Haslip*,[326] the Supreme Court has considered the constitutional limits on punitive awards.[327] As is developed in the second Part of this Article, *BMW* creates substantive limits upon the absolute amount of punitive awards by a three-factor test, involving the "degree of reprehensibility" of the defendant's conduct, the ratio between plaintiff's actual loss and the punitive award, and other civil or criminal sanctions available for comparable misconduct.[328] The *Haslip* case, on the other hand, had to do with the instructions used to control the jury's discretion; there, the jury had been told nothing beyond the fact that it had "discretion" to award punitive damages to "punish" the defendants and to "protect[ ] the public."[329] In dissenting, Justice O'Connor pointed out that the State's punitive-damages scheme already required the trial court to consider "a list of seven factors . . . relevant to the size of a punitive damages award":

> "'(1) Punitive damages should bear a reasonable relationship to the harm that is likely to occur from the defendant's conduct as well as to the harm that actually has occurred. If the actual or likely harm is slight, the damages should be relatively small. If grievous, the damages should be much greater.

---

322. *See infra* notes 331-332 and accompanying text.

323. *See supra* Part III.B.1.

324. *See supra* Part III.B.2-5.

325. 116 S. Ct. 1589 (1996).

326. 499 U.S. 1 (1991).

327. *See BMW*, 116 S. Ct. at 1604 (holding that a grossly excessive punitive award transcends the constitutional limit); *Haslip*, 499 U.S. at 23-24 (holding that the common-law method for assessing punitive damages is not per se unconstitutional).

328. *See supra* notes 207-217 and accompanying text.

329. *Haslip*, 499 U.S. at 19 (internal quotation marks omitted).

> "'(2) The degree of reprehensibility of the defendant's conduct should be considered. The duration of this conduct, the degree of the defendant's awareness of any hazard which his conduct has caused or is likely to cause, and any concealment or "cover-up" of that hazard, and the existence and frequency of similar past conduct should all be relevant in determining this degree of reprehensibility.
>
> "'(3) If the wrongful conduct was profitable to the defendant, the punitive damages should remove the profit and should be in excess of the profit, so that the defendant recognizes a loss.
>
> "'(4) The financial position of the defendant would be relevant.
>
> "'(5) All the costs of litigation should be included, so as to encourage plaintiffs to bring wrongdoers to trial.
>
> "'(6) If criminal sanctions have been imposed on the defendant for his conduct, this should be taken into account in mitigation of the punitive damages award.
>
> "'(7) If there have been other civil actions against the same defendant, based on the same conduct, this should be taken into account in mitigation of the punitive damages award.'"[330]

The problem that Justice O'Connor saw, however, was that "[u]nfortunately, . . . [the state] courts do not give . . . [these] factors to the jury."[331] She argued that later application of the factors on appeal "does not cure the vagueness of the jury instructions. . . . After-the-fact review of the amount in no way diminishes the fact that the State entrusts its juries with standardless discretion."[332]

Justice O'Connor has a point. Even if one does not agree with her as a matter of constitutional law, it would make sense to provide the jury with more guidance than to instruct them, as Justice O'Connor paraphrased it, with nothing more than, "[t]hink about how much you hate what the defendants did and teach them a lesson."[333] As a matter of state law, it would make sense to use many of

---

330. *Id.* at 51-52 (O'Connor, J., dissenting) (quoting Green Oil Co. v. Hornsby, 539 So. 2d 218, 223-24 (Ala. 1989) (quoting Aetna Life Ins. Co. v. Lavoie, 505 So. 2d 1050, 1062 (Ala. 1987) (Houston, J., concurring specially))). In *Hornsby*, these factors were promulgated for court review, but not as jury instructions. *See Hornsby*, 539 So. 2d at 223.

331. *Haslip*, 499 U.S. at 52 (O'Connor, J., dissenting).

332. *Id.*

333. *Id.* at 49; *see also* POLINSKY & SHAVELL, *supra* note 52, at 104-10 (setting out proposed jury instructions implementing the authors' theory that likelihood of escape from liability should be used to derive a multiplier that can be applied to actual damages to compute a

these factors in jury instructions and also to allow evidence about them.

Of the seven *Haslip* factors listed above, the fourth should be excluded. The "financial position of the defendant" is not sufficiently relevant to either economic or deontological bases for punitive damages to compensate for its tendency to prejudice and mislead.[334] The sixth and seventh factors—composed of criminal sanctions and civil damages actually paid by the defendant for similar conduct—should not be given to the jury, for the reasons stated above. Instead, they should be considered by the court alone, as adjustments to the jury's award.[335] But, consistently with the critiques set out above,[336] the other *Haslip* factors—the first, second, third, and fifth—should be provided to the jury in appropriate instructions from the court. In addition, evidence relevant to those factors should be admissible, provided it passes the test of Rule 403.[337]

CONCLUSION

Determination of the amount of a punitive award should be controlled primarily by economic analysis. The ultimate measure should be the size of the deterrence gap created by failures of injured persons fully to recover all losses. In theory, if this amount could be computed, it would supplement the compensatory award in a way that would optimize the allocation of resources between production and safety. The alternative approach of concentrating principally on blameworthiness confounds the accuracy of the process because of the greater difficulty of fixing and evaluating this philosophical abstraction. Furthermore, except in the instance of deliberate violation of a clear rule or norm, blameworthiness is associated with unjustified risk and with probabilities of nonliability, and these factors are correlated with the deterrence gap.

But economic analysis cannot be the exclusive ingredient in all such determinations. There must be allowance for informational deficiencies. This factor may result in situations in which a less precise

---

base punitive award, which then is to be modified in other ways specified by the instructions).

334. *See supra* Part III.B.1.

335. *See supra* notes 314-318 and accompanying text.

336. *See supra* Part III.B.2-5. *See generally* Thomas M. Melsheimer & Steven H. Stodghill, *Due Process and Punitive Damages: Providing Meaningful Guidance to the Jury*, 47 SMU L. REV. 329 (1994) (advocating that juries be given careful instructions that not only detail the purposes of punitive damages but also provide some benchmarks by which they should be calculated).

337. *See supra* note 313.

measure may be superior if it is more workable. In addition, moral considerations should provide a reality check against disproportionately low or high awards. As Professor Hurd explains the matter, consequentialist theories such as economic analysis provide the more general solution, and "the principal payoff of deontological maxims is their ability to define and patrol the borders of consequential justification."[338] The constitutional decisions emphasize moral blameworthiness and proportionality,[339] and this deontological focus is appropriate for those kinds of questions, although it would not be a good guide to calculation in the typical case.

In a few kinds of cases, it might be appropriate for expert witnesses to estimate the size of the deterrence gap by economic models and to provide this information to jurors by their testimony. Partisan manipulation of the results, however, is a severe impediment to the use of this evidence, given the problems of incommensurability and uncertainty.[340] Perhaps, nevertheless, there are some kinds of cases in which it will be realistic to use this evidence, such as those involving commercially measurable losses, a defined universe of claims, and ready analysis of defendants' production costs.

As an alternative method of approaching the problem, it might be better to use secondary but more readily determinable factors that are related to the deterrence gap. Profitability evidence, for example, generally should be admissible. It normally should be confined to the profitability of the injuring practice, and it should not be measured by the gains on legitimate, nontortious activity. Evidence of other events similar to the injury at trial should be admissible, subject to Rule 403,[341] in the second part of a bifurcated proceeding. Careful jury instructions should tell the jurors both the purposes of punitive damages and the way in which the evidence should lead to the determination, which in turn should be explained simply in terms of the deterrence gap.

Two types of evidence now in widespread use are inconsistent with this approach, and they should be avoided. First, evidence of net worth has little value in either economic or moral terms. The theoretical relationship that it has to risk tolerance is weak and uncertain, and it is dwarfed by the prejudice and tendency to mislead that such evidence inevitably carries. Wealth evidence tends to inflate punitive

338. Hurd, *supra* note 177, at 254.

339. *See supra* notes 207-217 and accompanying text.

340. *See supra* note 287.

341. *See* FED. R. EVID. 403; *see also supra* notes 237-238 and accompanying text (discussing Rule 403).

damages arbitrarily, in ways that are unrelated to either the moral blameworthiness of the defendant's conduct or to the deterrence gap. Such evidence would be vehemently criticized if offered on any other issue, and it is anomalous that it is admitted as a measure of punitive damages.

Second, fixed-dollar caps exhibit the opposite kind of arbitrariness in that they undervalue the deterrent in some cases. Caps based upon ratios of punitives to compensatories also may be arbitrary, but they have the virtue of tying the award to actual losses, which are related at least roughly to both the deterrence gap and to moral blameworthiness. The elimination of these two misleading devices, net-worth evidence and fixed caps, would go far toward improving our regimes for determining punitive damages, and a focus on the deterrence gap would bring us closer to a system that would avoid assessing amounts that are either too large or too small.