California Law Review

VOL. 78

JANUARY 1990

No. 1

Copyright © 1990 by California Law Review, Inc.

An Ordinary Economic Rationale for Extraordinary Legal Sanctions

David D. Haddock†
Fred S. McChesney††
Menahem Spiegel†††

Legal scholars have typically viewed the set of extraordinary legal sanctions (that is, remedies that systematically over- or undercompensate plaintiffs) as logically independent and often inefficient. In this Article, the authors offer a single, generally applicable model that predicts and explains the role extraordinary sanctions play in an efficient legal system. Starting with punitive damages, the authors show that extraordinary sanctions are necessary in those situations in which the expected imposition of liability rules, which seek to make the plaintiff whole, would encourage a defendant wrongly to take a plaintiff's property rather than negotiate for it. The authors distinguish their model from two others, the court-error model and the illicit-benefits model. They then extend their model to account for many other controversial and seemingly unrelated extraordinary legal remedies: injunctions, stipulated damages, collateral source recoveries, wrongful death awards, and criminal sanctions.

---

† Associate Professor of Law, Northwestern University.

†† Robert T. Thompson Professor of Law and Business and Professor of Economics, Emory University.

††† Associate Director, Center for Energy and Environment, City and Regional Planning Department, University of Pennsylvania.

The authors acknowledge support from the Program in Civil Liability at the Yale Law School and from the John M. Olin Program in Law and Economics at the University of Chicago Law School. Helpful comments on earlier drafts were received from Ian Ayres, Richard Craswell, Lloyd Cohen, Louis De Alessi, Robert Ellickson, Richard Epstein, David Friedman, D. Bruce Johnsen, Jason Johnston, Jonathan Macey, A. Mitchell Polinsky, Roberta Romano, Richard Speidel, and Gordon Tullock; and in presentations at the Law and Economics Workshop at the University of Chicago, the University of Delaware Department of Economics, the Emory University School of Law, the Georgetown University Law Center, and the Workshop in Civil Liability at the Yale Law School.

[P]lurality is not to be assumed without necessity. . . . [W]hat can be done with less is done in vain with more.

—William of Occam ("Occam's Razor")¹

This Article offers a property rights analysis of the way an efficient legal system would mold legal remedies. The Article focuses on the efficient protection of entitlements and the "extraordinary" remedies required to defend them. Extraordinary remedies are those that either exceed or fall short of ordinary, provable, compensatory damages.²

The Article seeks to make two contributions. First, in the spirit of Occam's Razor, it demonstrates that a single, concise economic model can predict a variety of seemingly unrelated extraordinary sanctions that have previously elicited equally various and detailed analyses. The model presented here (the extraordinary sanctions model) subsumes some of the competing analyses as special cases applying under limited circumstances, but it explicitly contradicts other analyses. Second, the Article furthers the analysis of extraordinary damages by explaining the principles that should, and often do, motivate courts to make subsidiary and superordinary awards.

Separate literatures now examine extraordinary remedies in tort,

---

1. Quoted in 3 THE NEW PALGRAVE: A DICTIONARY OF ECONOMICS 691 (J. Eatwell, M. Milgate & P. Newman eds. 1987).

2. This Article treats interchangeably the terms "ordinary, provable, compensatory damages" (henceforth "ordinary damages") and "damages intended to make the plaintiff whole." Modern tort law equates these two measurements of damages. See, e.g., *Beaulieu v. Elliott*, 434 P.2d 665, 670 (Alaska 1967) (footnote omitted) ("The general principle underlying the assessment of damages in torts cases is that an injured person is entitled to be replaced as nearly as possible in the position he would have occupied had it not been for the defendant's tort."). But see *infra* notes 59-63 and accompanying text (showing that the proper method for making a plaintiff whole will depend upon degree of market liquidity). Economists employ a similar concept with "compensated" demand curves, which leave the consumer's utility at the level it would have been had no price change occurred. See J. HIRSHLEIFER, PRICE THEORY AND APPLICATIONS 152-55 (4th ed. 1988); D. MCCLOSKEY, THE APPLIED THEORY OF PRICE 75-78 (2d ed. 1985).

including punitive damages,³ restitution,⁴ and the collateral source rule,⁵ often considering them in conjunction with the legal role of insurance.⁶ Other literatures consider criminal penalties,⁷ liquidated damage

3. Punitive damages are awards to the plaintiff over and above the plaintiff's ordinary provable injury, and they typically exceed ordinary damages. A diverse, contradictory, and often confusing literature discusses punitive damages. See Cooter, *Economic Analysis of Punitive Damages*, 56 S. CAL. L. REV. 79 (1982) (concluding that punitive damages are appropriate where the defendant acted intentionally and violated the legal standard by a wide margin); Ellis, *Fairness and Efficiency in the Law of Punitive Damages*, 56 S. CAL. L. REV. 1 (1982) (arguing that use of punitive damages for retribution and deterrence is misguided); Owen, *Civil Punishment and the Public Good*, 56 S. CAL. L. REV. 103, 117-21 (1982) (arguing for increasing fairness and efficiency in assessing punitive damages rather than for abolishing them); Owen, *Crashworthiness Litigation and Punitive Damages*, 4 J. PROD. LIAB. 221 (1981) (analyzing punitive damage awards in crashworthiness cases and cautioning against abuse of such awards); Owen, *Problems in Assessing Punitive Damages Against Manufacturers of Defective Products*, 49 U. CHI. L. REV. 1 (1982) (defending punitive damage awards in appropriate products liability cases but pointing out the potential for unfair punishment and inaccurate awards); Owen, *Punitive Damages in Products Liability Litigation*, 74 MICH. L. REV. 1257 (1976) (arguing that punitive damages can be used successfully to deter marketing of defective products); Priest, *Punitive Damages and Enterprise Liability*, 56 S. CAL. L. REV. 123 (1982) (suggesting that punitive damages and enterprise liability actually reduce safety by prompting reduced consumer investment in safety); Schwartz, *Deterrence and Punishment in the Common Law of Punitive Damages: A Comment*, 56 S. CAL. L. REV. 133 (1982) (arguing that deterrence and punishment do not adequately explain the actual practice of assessing punitive damages). A symposium in the *Alabama Law Review* on punitive damages appeared late in 1989, too late for us to take full advantage of all the papers. See generally *Symposium: Punitive Damages*, 40 ALA. L. REV. 687 (1989).

4. See, e.g., Wittman, *Liability for Harm or Restitution for Benefit?*, 13 J. LEGAL STUD. 57 (1984) (using economic theory as framework for analyzing choices between liability rules and restitution rules); Wittman, *Should Compensation Be Based on Costs or Benefits?*, 5 INT'L REV. L. & ECON. 173 (1985) (showing that asymmetries in information play a role in determining whether restitution should be based on costs or benefits). See generally W. KEETON, D. DOBBS, R. KEETON & D. OWEN, PROSSER AND KEETON ON THE LAW OF TORTS § 94, at 672-75 (W. Keeton 5th ed. 1984) [hereinafter PROSSER AND KEETON] (discussing restitution and its historical development).

5. The collateral source rule permits the winning plaintiff to collect ordinary, and possibly punitive, damages from the defendant while also receiving additional compensation from third parties such as insurers, employers, relatives, or friends. See, e.g., Fleming, *The Collateral Source Rule and Loss Allocation in Tort Law*, 54 CALIF. L. REV. 1478 (1966) (exploring means of making reallocation more equitable); Maxwell, *The Collateral Source Rule in the American Law of Damages*, 46 MINN. L. REV. 669 (1962) (examining the collateral source rule in terms of four types of additional compensation received: insurance, employee benefits, gratuities, and social legislation); D. Ellis, *The Law and Economics of the Collateral Source Rule* (draft July 7, 1983) (unpublished manuscript on file with the authors) (concluding that the collateral source rule may not allocate resources efficiently).

6. See, e.g., Ellis, *supra* note 3, at 71-76 (arguing that insurance is inconsistent with the punishment and deterrence goals of punitive damage liability); Shavell, *On Liability and Insurance*, 13 BELL J. ECON. 120 (1982) (technical analysis of effect of liability insurance on accident-reduction incentives and risk allocation).

7. The criminal law makes no explicit effort to equate the statistical expectations of victims' injuries and the sanctions imposed on convicted criminals. Moreover, these sanctions often impose deadweight losses (imprisonment or corporal or capital punishment); alternatively, they benefit, not the victim, but either the state (fines, confiscation, or forced labor) or, more rarely, otherwise uninvolved third parties (compulsory pro bono publico work). The victim can sometimes collect ordinary, and possibly punitive, damages in an independent civil action. For a sample of the more important works involving economic analyses of the criminal law, see Becker, *Crime and*

clauses⁸ and punitive damages in contract,⁹ wrongful death awards,¹⁰ damages for patent and copyright infringement,¹¹ antitrust damages,¹²

Punishment: An Economic Approach, 76 J. POL. ECON. 169 (1968) (developing an economic analysis for optimal allocation of resources to deter and punish crime); Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193 (1985) (giving an economic interpretation of criminal law doctrine and suggesting that the doctrine promotes efficiency by preventing people from bypassing the market); Shavell, *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, 85 COLUM. L. REV. 1232 (1985) (developing a theory of nonmonetary sanctions); Stigler, *The Optimum Enforcement of Laws*, 78 J. POL. ECON. 526 (1970) (suggesting a scheme for rational enforcement of laws); Tullock, *The Welfare Implications of Tariffs, Monopoly and Theft*, 5 W. ECON. J. 224, 228-31 (1967) (considering the indirect social costs of crime). Posner, *supra*, makes reference to a number of other works on this subject.

8. Courts may refuse to enforce voluntarily negotiated contract clauses that stipulate a definite amount of damages to be paid to the promisee upon breach by the other party. In those instances, courts award the plaintiff promisee only ordinary damages. In other instances, however, courts enforce the liquidated damages clause even though the damages are either greater than or less than ordinary. See, e.g., Clarkson, Miller & Muris, *Liquidated Damages v. Penalties: Sense or Nonsense?*, 1978 WIS. L. REV. 351 (using concept of economic efficiency to justify distinction between liquidated damages and penalties); Goetz & Scott, *Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach*, 77 COLUM. L. REV. 554 (1977) (arguing that liquidated damages are, in many situations, efficient means of insuring against otherwise uncompensable losses due to breach); Rea, *Efficiency Implications of Penalties and Liquidated Damages*, 13 J. LEGAL STUD. 147 (1984) (arguing that enforcement of penalty clauses is generally undesirable).

9. See, e.g., Harman, *An Insurer's Liability for the Tort of Bad Faith*, 42 MONT. L. REV. 67 (1981) (examining development of tort of "bad faith," which imposes a greater liability on the insurer than its contractual obligation); Sullivan, *Punitive Damages in the Law of Contract: The Reality and the Illusion of Legal Change*, 61 MINN. L. REV. 207, 244 (1977) (arguing that expanded availability of punitive damages in contract has blurred the distinction between contract and tort); Comment, *Punitive Damages on Ordinary Contracts*, 42 MONT. L. REV. 93, 108 (1981) (authored by Laura Lee) (concluding that courts should closely examine punitive damages in contracts cases and permit them only when tortious elements are present).

10. In civil wrongful death cases, juries typically are instructed that a defendant is not liable for the losses suffered directly by the deceased—that is, for the expected utility lost as a result of an early death. To receive an award from the defendant, the estate must prove indirect losses to a third party, such as the spouse or child of the deceased. In criminal cases, the sanctions imposed are neither collectible by the decedent's estate nor calibrated to approximate the lost utility. For a discussion of wrongful death awards, see W. LANDES & R. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 186-89 (1987) (finding no rational explanation for the "inadequate" nature of wrongful death awards); Cohen, *Toward an Economic Theory of the Measurement of Damages in a Wrongful Death Action*, 34 EMORY L.J. 295 (1985) (arguing that basing a wrongful death award on the derivative loss to family or estate is economically inefficient because it undervalues consequences of tortfeasor's error); Friedman, *What Is "Fair Compensation" for Death or Injury?*, 2 INT'L REV. L. & ECON. 81 (1982) (arguing that "full compensation" of the plaintiff is economically inefficient); Komesar, *Toward a General Theory of Personal Injury Loss*, 3 J. LEGAL STUD. 457 (1974) (developing a conceptual framework for examining personal injury damage rules and applying it to damages for wrongful death).

11. For a general discussion and citations to literature and case law, see D. DOBBS, *HANDBOOK OF THE LAW OF REMEDIES* 434-56 (1973).

12. Ordinary damages are trebled in private antitrust awards. Additional sanctions, not received by a specific injured party, may be imposed in a separate criminal action. See, e.g., Breit & Elzinga, *Antitrust Penalties and Attitudes Toward Risk: An Economic Analysis*, 86 HARV. L. REV. 693 (1973) (using economic theory to examine the deterrent value of private antitrust damage awards).

and insider trading sanctions.¹³ Still other, less common, extraordinary sanctions exist. In some states, for example, a defendant who willfully cuts down a plaintiff's trees is liable for treble the amount of ordinary damages.¹⁴ Similar statutes permit multiple damages for takings of other commodities on the land, such as minerals and crops.¹⁵

Previous analyses of each extraordinary sanction are needlessly isolated, with arguments specialized to rather narrow institutional details. Moreover, some of those arguments are empirically empty, others are unnecessarily complex, and still others require ad hoc subarguments to account for discrepancies between their predictions and existing law. As shown here, extraordinary sanctions can be explained both more generally and more simply, and in a way that yields a larger number of testable, hence empirically meaningful, implications.

The most studied extraordinary remedy, and the one treated with the most skepticism and hostility by most law and economics scholars, is punitive damages in tort. To a majority of those observers, the recent trend of punitive tort awards seems unprincipled and counterproductive, because punitive damages overcompensate plaintiffs and thus overdeter potential defendants engaged in useful activities.¹⁶ This Article shows that punitive damages can be useful precisely because some defendant activities have no social value at any level and, in fact, impose net social costs whenever initiated. This Article ultimately shows that net social cost at any positive activity level is the key to understanding the range of extraordinary sanctions, subsidiary as well as superordinary ones.

The second contribution of this Article is to extend the analysis of punitive damages. The extraordinary sanctions model that explains punitive damages can predict and explain other extraordinary sanctions

---

13. Section 2 of the Insider Trading Sanctions Act of 1984, 15 U.S.C. § 78u(d)(2)(A) (1988), allows the Securities and Exchange Commission to recover civil penalties from any person who trades securities while in possession of material, nonpublic information. These penalties are measured by the gains, potentially trebled, to the inside traders rather than the losses befalling anyone else. See, e.g., Haddock & Macey, *Regulation on Demand: A Private Interest Model, with an Application to Insider Trading Regulation*, 30 J.L. & ECON. 311 (1987) (developing a private interest model of government behavior to examine regulations promulgated and enforced by SEC).

14. See, e.g., *Miller v. Wykoff*, 346 Mich. 24, 26-27, 77 N.W.2d 264, 264-65 (1956); *Bailey v. Hayden*, 65 Wash. 57, 60, 117 P. 720, 721 (1911). The trebled figure may approximate the trees' value in their best use, as part of a yard's landscaping perhaps, but only coincidentally.

15. D. DOBBS, *supra* note 11, at 346-48.

16. See *infra* notes 24-32 and accompanying text. But see R. COOTER & T. ULEN, *LAW AND ECONOMICS* 338 (1988) (arguing that punitive damages are principled and productive because they deter and punish intentional wrongdoing and reckless behavior); W. LANDES & R. POSNER, *supra* note 10, at 302-07 (concluding that punitive damages are economically efficient in some circumstances); cf. S. SHAVELL, *ECONOMIC ANALYSIS OF ACCIDENT LAW* 146-51 (1987) (examining models of accidents for which it is "desirable" that liability exceed actual losses); Landes & Posner, *New Light on Punitive Damages*, *REGULATION*, Sept.-Oct. 1986, at 33 (concluding that punitive damages are rarely imposed except in cases of intentional wrongdoing).

as well. Like punitive damages, those other extraordinary sanctions play a valuable role in a legal system that protects property rights efficiently. Thus, the standard criticisms of those extraordinary sanctions are misguided.

Extraordinary sanctions are aberrant only if law is thought always to strive to make plaintiffs as a class whole ex post or (what is operationally the same thing) to achieve optimal deterrence by compensating plaintiffs for their exact losses (appropriately adjusted for underdetection or mismeasurement of injury). From that perspective, a punitive damage award—or any extraordinary sanction—is tautologically inefficient. Efficient law seeks optimal deterrence. Sometimes, that will require making *defendants* whole from an ex ante perspective. In those cases, the legal system will try to strip away all benefit that the defendant expected from his actions, an objective that often calls for an extraordinary, “punitive” sanction.¹⁷ When an interaction between a defendant and a plaintiff alters the aggregate wealth of society, making the plaintiff whole is not the same thing as making the defendant whole.¹⁸

Our argument differs from that of a minority of scholars who approve of punitive damages because without them defendants would obtain “illicit benefits.”¹⁹ Our argument also differs from “court-error” models.²⁰ Neither the illicit benefits model nor the court-error model is as general as the one developed here, nor do they predict existing law as well.

The Article works along both an intensive and an extensive margin,

---

17. In other words, the court will “make a litigant whole” if it restores the litigant’s welfare to its initial level (“endowment” in economic terminology or “entitlement” in legal parlance). In the classic tort situation, the defendant (the tortfeasor) expected to be better off after committing the tort, while the plaintiff (the victim) is worse off after suffering it. A system that makes plaintiffs whole tries to give back to plaintiffs what they have lost. But when the system makes defendants whole, it takes back from defendants what they expected to gain. Because a single measure of damages will rarely serve both objectives, the system one observes leads in some cases to extraordinary damages, that is, awards that are noncompensatory to plaintiffs.

18. One essential difference between the present paper and many of its predecessors is its constant attention to the possible discrepancy between the plaintiff’s welfare loss from an interaction with the defendant and the defendant’s welfare gains. Thus, making the plaintiff whole and making the defendant whole are often inconsistent goals. Our analysis is directed toward discovering how and why a court decides which goal to pursue in any particular category of legal action.

19. These benefits include the pleasure defendants obtain from their intentionally harmful behavior. Punitive damages are supposedly appropriate because intentional harm is both immoral and more deserving of punishment than unintentional harm. See R. COOTER & T. ULEN, *supra* note 16, at 338; S. SHAVELL, *supra* note 16, at 147; Cooter, *supra* note 3, at 89-91.

20. Some scholars argue that punitive damages correct for courts’ tendency to underestimate plaintiffs’ damages and thus provide optimal incentives to avoid harmful behavior. See, e.g., Johnston, *Punitive Liability: A New Paradigm of Efficiency in Tort Law*, 87 COLUM. L. REV. 1385, 1388-89 (1987); Png, *Optimal Subsidies and Damages in the Presence of Judicial Error*, 6 INT’L REV. L. & ECON. 101 (1986); cf. R. POSNER, *TORT LAW: CASES AND ECONOMIC ANALYSIS* 32 (1982) (noting the argument without endorsing it); S. SHAVELL, *supra* note 16, at 149-50 (examining this argument).

explicating the model through an intensive examination of punitive damages before using it to explain an extensive set of extraordinary damages. Part I of the Article develops the basic model, a "property rule model" intended to address conduct that disregards property rights. Analysis of the competing explanations of punitive damages in tort—and the objections to these damages—reveals a frequently inappropriate reliance on the standard "liability rule model"²¹ that aims to make plaintiffs whole.

The liability rule model was first developed to explore the marginal conditions necessary to optimize a useful but externality-producing activity, and it is appropriate for many kinds of interactions. The central argument of Part I, however, is that the liability rule model fails in the situations that commonly lead to punitive awards. In contrast, a property rule model, like the model presented here, will show that it is more efficient to make defendants whole in certain illiquid market settings. Because the optimum level of some activities is zero, the extraordinary sanctions model focuses on total costs and benefits to parties under alternative legal rules rather than on marginal conditions under the rules. The model can therefore account for facets of punitive tort law heretofore unexplained or mistakenly criticized by other theories.

The extraordinary sanctions model developed in Part I relies on a series of Edgeworth boxes, now a familiar graphical device in the economic analysis of law.²² The graphs permit greater rigor in analyzing any individual legal doctrine than does purely verbal reasoning. Equally important for our purposes, the graphs illustrate how several seemingly distinct doctrines can be treated as aspects of a single model.

Part II shifts to the extensive margin. It applies the extraordinary sanctions model to other super- and sub-compensatory awards—penalty clauses in contract, wrongful death awards, the collateral source rule, and criminal penalties (including "corporate crime" statutes). As with punitive tort damages, such extraordinary sanctions are often criticized by scholars who analyze any one sanction in isolation. Part II shows that all of these sanctions are predicted by the model developed in Part I. In addition, that model also illuminates an interplay between efficient legal rules and private insurance markets.

For brevity, discussion of still other extraordinary sanctions is omitted, but the predictive content of the model with respect to other sanctions should be apparent. It is exactly the ability to explain numerous legal rules with one simple construct that most clearly sets the

---

21. See Brown, *Toward an Economic Theory of Liability*, 2 J. LEGAL STUD. 323 (1973).

22. The seminal treatment is Birmingham, *Damage Measures and Economic Rationality: The Geometry of Contract Law*, 1969 DUKE L.J. 49 (using economic analysis and the Edgeworth box to examine measurement of damages for breach of contract). For a more recent use of Edgeworth boxes in legal analysis, see Goetz & Scott, *supra* note 8, at 563-68.

extraordinary sanctions model apart from its precursors and gives the model its appeal. What could have been "done with less" until now has been "done in vain with more."

I

THE EFFICIENCY RATIONALE FOR MAKING DEFENDANTS WHOLE

A. *The Standard Model of Tort Remedies*

Modern economic analysis of tort law began with Coase and developed into the "standard liability model" through the work of Calabresi, Brown, and many others.²³ In that model, the purpose of legal sanctions is to internalize the external costs (taken to be unambiguously ascertainable) of socially beneficial activities, such as driving. Efficiency, and hence "optimal deterrence," requires that an actor take into account all the costs that a given activity imposes on the actor and on others. Under the assumptions of prohibitive transaction costs *ex ante*, decreasing marginal benefits of the activity, and constant or increasing marginal costs, efficient legal sanctions will equal the external costs at the margin.

In the standard liability model, a compensatory remedy that makes the plaintiff whole after the fact—that is, a remedy that compensates the plaintiff for the exact cost (injury) suffered from the defendant's actions—induces a rational and efficiently-informed defendant²⁴ to take the appropriate degree of care. When the law aims to influence the marginal behavior of individuals, setting appropriate marginal remedies—that is, the costs to be borne by defendants—is of obvious importance. The standard model implies that sanctions that undercompensate plaintiffs underdeter defendants, inducing excessive amounts of the useful but hazardous activity; for example, people drive too fast, too often, and too far. Sanctions that overcompensate plaintiffs overdeter defendants, causing them to forego activities that have net social benefits; people drive too slowly, too infrequently, and not far enough.²⁵

The standard liability model treats extraordinary damages as corrective, imposed only to address two sources of underdeterrence. First, if legal remedies exclude some items of loss, systematic undercompensation, and thus underdeterrence, result. Second, if some tortfeasors go

23. See generally Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960); G. CALABRESI, *THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* (1970); Brown, *supra* note 21. For an extensive bibliography, see C. GOETZ, *LAW AND ECONOMICS: CASES AND MATERIALS* 516-23 (1984). For a briefer, but more recent, bibliography, see R. COOTER & T. ULEN, *supra* note 16, at 371, 475-76.

24. An "efficiently-informed defendant" is not (and should not be) a perfectly informed defendant. See Stigler, *The Economics of Information*, 69 J. POL. ECON. 213, 224 (1961).

25. Cooter, *supra* note 3, at 85.

undetected or escape liability, assessing ordinary damages against the remaining tortfeasors creates an incentive *ex ante* to undertake too much of the activity, since the probability of paying the full cost of one's behavior is less than one. Thus, the traditional rationale for punitive damages in this context is "that compensatory damages don't really compensate fully, and punitive damages help close the gap."²⁶ Although some injured individuals are undercompensated while others are overcompensated, the statistical expectation is to be made whole; more important for economic analysis, potential defendants are thereby optimally deterred at the margin. There is nothing "extraordinary" about corrective damages, because they are intended merely to make plaintiffs expectationally whole.²⁷

Some sanctions, however, systematically overcompensate or undercompensate plaintiffs. These extraordinary sanctions, as the term is used here, purposely impose losses on defendants that deviate from awards that would make plaintiffs whole. Some awards, such as those for defamation or trespass, are superordinary by design, exceeding those that make the plaintiff whole. Others, such as those for wrongful death, are intentionally subordinary, falling short of demonstrable injury.

B. Competing Models and Their Failings

The standard liability model predicts corrective awards, but not superordinary and subordinary sanctions. Faced with courts' persistence in imposing such extraordinary sanctions, the typical law and economics scholar has reacted in one of three ways.

1. Courts' Misguided or Nefariously Destructive Intent

The largest group of scholars characterize extraordinary sanctions as mistakes (or worse). These scholars believe that since extraordinary sanctions over- or underdeter, the existence of such sanctions reflects the willingness of modern judges or juries to ignore efficiency concerns. George Priest, for example, complains of "the absence of theoretical

26. R. POSNER, *supra* note 20, at 32.

27. Australian law explicitly recognizes the distinction between damages intended to correct for mismeasurement and those intended to serve punitive purposes. It permits the award of corrective, "aggravated" damages in cases where ordinary, compensatory damages do not make plaintiffs whole for certain intangible losses. Punitive, "exemplary" awards then may be added to both compensatory and aggravated damages. See H. LUNTZ, D. HAMBLBY & R. HAYES, *TORTS: CASES AND COMMENTARY* 274-75 (1985).

The conclusions of the standard liability model are unobjectionable when its assumptions are met. Accident litigation is a typical example. Indeed, it will be seen below that the standard liability model is merely a special case of the more general extraordinary sanctions model. See *infra* notes 59-63 and accompanying text. The assumptions of the liability model, however, are substantially stronger than has been generally recognized, and when they are not met, that model yields inappropriate conclusions.

justification for punitive damage judgments."²⁸ Dorsey Ellis claims that "[p]unitive damages can be shown to promote efficient levels of deterrence only in those cases where expected liability for compensatory damages is less than expected harm to society [that is, where they are corrective] or where harm is deliberately caused and the satisfaction obtained by the actor is 'illicit.'" He concludes that "[p]resent punitive damage law extends far beyond these categories, however, and imposes additional costs on society."²⁹

Although William Landes and Richard Posner recognize a role for superordinary punitive awards in tort, they excoriate subsidiary ones:

Although . . . the damage rules of the tort system display a characteristic economic intuition, in one very important respect they seem out of phase with economics. This is the matter of valuing a life in a tort case where the victim dies ("wrongful death").

. . . .

The limitation of damages to survivors' pecuniary loss is very peculiar. It implicitly assumes—if, as we generally believe to be the case, tort law seeks to internalize the costs of accidents—that the average person derives no utility from living.³⁰

Thus, the dominant group of scholars has been "harshly critical" of many extraordinary sanctions,³¹ often calling for the abolition or drastic limitation of superordinary or subsidiary awards.³² Unless it is corrective, a remedy that does not make a plaintiff whole seems inefficient to most commentators. Because such a remedy systematically over- or undercompensates, the commentators conclude that it does not produce optimal deterrence for beneficial but externality-causing activities.

2. *The Court-Error Model*

A smaller group of law and economics scholars account for punitive damages within the standard liability model by interpreting them as roughhewn compensation for inevitable court error. Posner, for exam-

28. Priest, *supra* note 3, at 132.

29. Ellis, *supra* note 3, at 77. The notion of "illicit" satisfaction is discussed at *infra* notes 34-37 and accompanying text.

30. W. LANDES & R. POSNER, *supra* note 10, at 186-87.

31. Johnston, *supra* note 20, at 1388 (citing many of the scholars cited at *supra* note 3 as harsh critics of existing extraordinary damage awards).

32. On superordinary awards, see Sales & Cole, *Punitive Damages: A Relic That Has Outlived Its Origins*, 37 VAND. L. REV. 1117, 1154-66 (1984) (advocating abolition or severe limitation of punitive damages). For arguments supporting the call for limitations, see Ellis, *supra* note 3, at 76-78 (finding only a limited rationale for punitive damages); Priest, *supra* note 3, at 123-25 (adopting same view). On subsidiary awards, see W. LANDES & R. POSNER, *supra* note 10, at 187-89 (suggesting that wrongful death awards "bear no necessary relationship to the economically correct measure" of damages); Cohen, *supra* note 10, at 339-40 (suggesting a method of correcting the undervaluation of wrongful death damages). But see Friedman, *supra* note 10, at 81 (arguing that "full compensation" for death or lethal injury is economically inefficient).

ple, notes that in the case of a taking, the court must evaluate due compensation, "and its judgment is bound to be imperfect," so it should impose extra damages to ensure that plaintiffs are not undercompensated.³³

But the court-error model fails on several counts. First, it relies on the mere existence of measurement error by judges to justify punitive awards. Because measurement is costly, however, courts should measure compensatory damages imperfectly. Moreover, a court's measurement errors need not be systematically biased. In other words, even if actual awards reflect variance around theoretically correct awards, with some awards too high or too low, the average compensatory award is not necessarily inappropriate. High variance in awards admittedly implies real costs to risk-averse individuals—plaintiffs and defendants alike—but little else. When incorporated into the standard tort model, costly judicial errors do not necessarily yield nonoptimal deterrence, since expectational errors affect both parties. Nor, if deterrence levels are affected, is the amount of deterrence systematically too little or too much. Thus, even if the system were to use extraordinary sanctions to adjust for court error, one would have no way of knowing whether to increase or to decrease the damage award—or indeed, whether the initial award was biased at all.

The error-based model, reflecting only measurement errors, fails on another count. As explained by the extraordinary sanctions model presented below, a second sort of "error" is central to understanding extraordinary sanctions. This second sort occurs because, under predictable circumstances, it will be impossible even to know in any operational sense *what needs to be measured*. It hardly makes sense to term a court's failure to know the unknowable an "error." By contrast, the extraordinary sanctions model developed below shows that the second type of court "error"—the inability to know the unknowable—implies still other costs, ones that the court-error model overlooks entirely. Those are the costs of opportunistic behavior, which unambiguously merits extraordinary awards rather than corrective ones. Thus, the extraordinary sanctions model yields implications that sometimes differ from those of the court-error model. Existing legal rules more closely match the predictions of the extraordinary sanctions model, as shown below.

---

33. R. POSNER, *supra* note 20, at 34. Posner also notes that superordinary damages may compensate tort victims for their litigation costs. *Id.* As David Friedman notes, however, economizing on litigation costs and the other costs of the judicial system can, under some conditions, lead to subordinary awards rather than punitive damages. Friedman, *An Economic Explanation of Punitive Damages*, 40 ALA. L. REV. 1125, 1132-33 (1989).

3. *The Illicit Benefits Model*

Still a third construct employed by some law and economics scholars utilizes the concept of "illicit benefits" to account for extraordinary sanctions. Robert Cooter and Thomas Ulen, for example, say such sanctions are necessary to combat behavior that is "far worse morally" than the conduct for which ordinary damages are awarded.³⁴ Similarly, Steven Shavell remarks that "along with the usual costs of care, injurers may experience disutility that is not credited in the calculation of social welfare."³⁵ Some injurers "will obtain utility from the losses [they have caused, and] this utility may be . . . socially illicit . . . . It follows that for injurers to be induced to behave optimally, the magnitude of liability should equal the sum of losses caused and of illicit utility gained."³⁶

The "illicit benefits" explanation of punitive damages is not so much inconsistent with ours as it is incomplete and empirically empty. It is incomplete because illicit behavior would explain only sanctions that overcompensate plaintiffs, not subsidiary awards that systematically undercompensate plaintiffs. The illicit benefits construct cannot explain undercompensation without conjuring up a notion of "superlicit" behavior. This explanation would be unconvincing in most instances in which subsidiary awards are common—wrongful death cases, for example. Like the other prevailing models of various extraordinary sanctions noted in the introduction, the illicit benefits model addresses only a subset of the rules explained by the present model and has no application to other rules that are widely recognized as inconsistent with the standard liability model.

Second, and more important, the illicit benefits model is empirically empty, because it does not explain why some benefits are illicit while others are not. The only way to know which sources of utility are illicit under that model is to observe whether the law metes out superordinary sanctions. Such reasoning could be used to "explain" any configuration of legal rules, and hence can predict none. Indeed, such an "explanation" would seem malleable enough to fit into almost any theory of extraordinary sanctions, as it fits into our own. If, alternatively, one uses a sort of "behind-the-veil" intuition to avoid the tautological character of the illicit benefits model, the predictions yielded provide a less satisfactory picture of the law than does the present model, as will occasionally be noted below.

Widespread attitudes that particular behavior is "immoral" or that certain utility is "socially illicit" are endogenous phenomena to be

---

34. R. COOTER & T. ULEN, *supra* note 16, at 338.

35. S. SHAVELL, *supra* note 16, at 146.

36. *Id.* at 147.

explained, not exogenous phenomena that do the explaining.³⁷ The remainder of this Article explains why such attitudes arise, why certain behavior may, in consequence, be deemed "illicit," and why either over- or undercompensatory sanctions appear in the law. In brief, the extraordinary sanctions model, which subsumes the illicit benefits model as a special case, provides more and better-defined predictions.

C. *Making Defendants Whole: Punitive Damages and Restitution*

1. *Property Rules Versus Liability Rules*

Critics of extraordinary sanctions wrongly claim that potential defendants will choose to behave efficiently provided that they must always make plaintiffs whole expectationally, a norm that excludes extraordinary sanctions. Yet no such norm does, or should, prevail in court. The standard liability model, from which the norm of making the plaintiff whole flows, optimizes externality-producing activities that have net, but not marginal, social value. Some activities, however, have *no* social value at any level: their optimal level is zero. To eradicate such activities, efficient law would instead try to "make the *defendant* whole," that is, it would reduce to zero the expected gain available to the defendant from the injurious activity, leaving no incentive for him to attempt the activity in the first place.³⁸

The dichotomy between making plaintiffs whole and making defendants whole rests on the fundamental difference between liability rules and property rules, to use the familiar Calabresi-Melamed distinction.³⁹ An entitlement is protected by a property rule if the law condones surrender only through voluntary exchange. An entitlement has the lesser protection of a liability rule if it can be lost lawfully to anyone willing to pay court-determined compensation. That the latter entitlements can be taken legally, subject only to payment of compensation after the fact, means that liability rules may be strategically abused.

The Edgeworth Box in Figure 1 illustrates the distinction between property rules and liability rules. The utility function of some person π is mapped over π 's alienable entitlements: a composite good (on the

37. For a sample of attempts to explain these phenomena, see Posner, *Utilitarianism, Economics, and Legal Theory*, 8 J. LEGAL STUD. 103 (1979) (arguing that moral intuitions are based on maximization of social wealth); Cohen, *A Justification of Social Wealth Maximization as a Rights-Based Ethical Theory*, 10 HARV. J.L. & PUB. POL'Y 411, 412 (1987) (endorsing Posner's proposed relationship between moral intuitions and wealth maximization); Johnsen, *Wealth Is Value*, 15 J. LEGAL STUD. 263, 264 (1986) (arguing that interpretation of wealth maximization as a moral truth requires broader definition of social wealth).

38. See *infra* note 54.

39. See Calabresi & Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1105-06 (1972) (explaining that different entitlements are protected by different rules, called "property," "liability," and "inalienability" rules).

and Δ on the contract curve between R and U, but at neither of those endpoints. Let the contract point be C. Neither party is "made whole" by that transaction, because each is on a higher indifference curve at C than at E. If that joint advantage is possible, gains from exchange exist.

Ex ante, bargaining outcomes in thin markets are indeterminate. Although π is assumed for the moment to know Δ 's utility map, and, hence, the contract curve, the location of point C will depend on the ability and willingness of Δ to bargain. That, in turn, will depend on Δ 's intelligence and negotiating experience and the opportunity cost of Δ 's time. Furthermore, π 's gains will depend on the characteristics of Δ 's utility map. If Δ 's indifference curve at E had been steeper or more strongly curved, the trading lens would have had a different shape. Indeed, it might have been impossible for π to attain point C, or even the indifference curve i_1 , through negotiation.

A typical strict liability award following a taking would be for ordinary damages.⁴¹ That award makes the injured party "whole," that is, it returns π to the indifference curve that passes through the original entitlement—here, the curve i_0 that passes through E. Assume that it is costless for the parties to litigate and for the court to enforce its judgment. Also assume that the court neither mistakenly imposes liability in the absence of a violation nor fails to impose liability when a violation has in fact occurred. Under those conditions, a taking by Δ of E_t units of π 's goods will lead to a monetary remedy of tR . E_t is Δ 's preferred taking if Δ expects to be forced to compensate π , because it is at R that Δ reaches the highest indifference curve attainable (u_2), subject to the legal constraint that π must be returned to i_0 . R is on the contract curve between u_0 and i_0 , so the movement from E to R is Pareto efficient.⁴²

Since, by assumption, the law seeks to make the plaintiff whole, the court will compensate π . Thus, only π 's utility function matters; Δ 's wealth, insurance coverage, and all other factors related to Δ 's utility function are irrelevant to the court. If Δ 's indifference curves were steeper, for example, Δ would have taken less than E_t , but an ordinary remedy would still return π to i_0 , though at a different point.

Under present assumptions—thin markets, universal knowledge of π 's and Δ 's indifference maps, zero transaction and litigation costs—scholars typically draw two inferences from the Edgeworth box model. First, R, C, or any other point along the contract curve within the trad-

41. Strict liability, including appropriate controls on the behavior of potential plaintiffs, is assumed to apply to liability rule takings, such as those at issue in an eminent domain or personal injury proceeding. The extraordinary sanctions model would be similar if negligence rules applied. Indeed, many of the model's implications would be stronger, because negligence awards do not even restore plaintiffs expectationally to their pretaking utility levels.

42. A particular change is Pareto efficient if it makes at least one party better off without making another party worse off.

ing lens (indeed, *any* point within the trading lens) is Pareto superior to E,⁴³ where gains from exchange remain unexploited. Second, there is no welfare preference between the Kaldor-Hicks equivalents R and C,⁴⁴ which reflect only different asset distributions, each of which exploit all gains from exchange. Thus, many scholars conclude that there is no welfare rationale for property protection rather than liability protection, or vice-versa.

Those two inferences are incorrect, however, because the static (given-endowment) welfare model ignores opportunism. Despite the apparent social welfare equivalence of R and C, π and Δ have personal preferences— π prefers C and Δ prefers R, because C provides a higher level of utility for π , R a higher level of utility for Δ . If Δ expects to pay only a compensatory award, then Δ has an incentive to expend resources to be a taker, who gets all gains from exchange. At the same time, π has an incentive to expend resources to prevent a taking, thus preserving some of those gains. Because a taker in an illiquid market extracts all gains from exchange, liability remedies that compensate π in a legal sense—monetary awards of tR —are undercompensatory in a larger sense. They do not restore π 's preexisting opportunities for obtaining exchange value—opportunities for a share of any gains from exchange. The now impossible trade would have made π more than whole. "Full" compensation for the E_t taken would put π at some point on i_1 , or even back to E, from which π still could negotiate to i_1 at point C.⁴⁵

Allowing Δ to take from π and pay ordinary compensation (liability rules) rather than negotiate with π and share the gains from exchange (property rules), diminishes π 's incentive to invest scarce or unique resources in tradeable assets. For example, real estate improvements are less likely, and those undertaken are less valuable, when governments are more prone to exercise the power of eminent domain. Instead, traders will substitute present consumption for some investment. Furthermore, substitute investments less threatened by opportunistic takings will be enlarged, even where such investments are socially less valuable at the margin than the foregone alternatives. Viewing the Edgeworth box more dynamically, its dimensions at any time depend on the rules of exchange

---

43. One distribution is Pareto superior to another if it makes at least one party better off without making any other party worse off.

44. One distribution of assets is said to be Kaldor-Hicks equivalent to another when neither A nor B would be willing to pay the amount demanded by the other to move to the alternative distribution. See W. LANDES & R. POSNER, *supra* note 10, at 16.

45. In the standard liability model, efficient caretaking by potential defendants is induced when tortfeasors expect to bear the full costs that their actions impose on potential plaintiffs. The model here makes plain, however, that in a thin market, the cost imposed on plaintiffs is ambiguous; the utility from owning and using an asset (i_0) is not generally the same as the utility obtainable from exchanging it (i_1). See also *infra* notes 59-63 and accompanying text.

previously established, rules that earlier investors had expected would prevail at present.⁴⁶

In a liability rule regime, potential traders will face a prisoner's dilemma.⁴⁷ Even if compensation will be required, everyone's dominant strategy in illiquid markets will be to take the entitlements of others in order to garner more of the potential gains from exchange for oneself. But traders as a group will suffer if individuals can strategically substitute court-ordered compensation for negotiation.⁴⁸ Liability rules make thin-market entitlements a sort of commons, so that gains that would otherwise be available from exchange instead will be dissipated by searches for and defense of takeable assets,⁴⁹ and resources available for investment will be diverted toward less takeable uses.⁵⁰ Societies willingly bear such costs only when circumstances make property rules untenable. Property protection for entitlements is properly the norm; liability protection, and hence the standard liability model, should be an exceptional response to exceptional circumstances.

2. *An Efficiency Rationale for Punitive Damages*

As shown above, a defendant (Δ) is better off with a liability remedy of ordinary damages than with a bargained outcome, given the present assumption of thin markets, universal knowledge of π 's and Δ 's indiffer-

46. See D. Haddock & F. McChesney, *Bargaining Costs, Bargaining Benefits and Compulsory Non-Bargaining Rules for Bilateral Monopoly* 8-13 (Mar. 1989) (unpublished manuscript on file with the authors). The model developed there derives from the original work (in a very different context) of Gordon Tullock. See Tullock, *supra* note 7, at 224. Substituting liability protection for property protection opens new avenues for Tullock-type rent-seeking. For use of a Tullock-like model to analyze the criminal law intensively, see Posner, *supra* note 7. But see Buchanan & Faith, *Entrepreneurship and the Internalization of Externalities*, 24 J.L. & ECON. 95 (1981) (claiming that liability rules are superior to property rules in certain situations).

47. The prisoner's dilemma arises when decisions made by interdependent but isolated utility-maximizing individuals produce a suboptimal outcome for the group. Thus, decisions that are rational from the viewpoint of each party produce a situation where all parties are worse off.

48. See generally Haddock & Hall, *The Impact of Making Rights Inalienable*; Merrión v. Jicarilla Apache Tribe, Texaco, Inc. v. Short, Fidelity Federal Savings & Loan Ass'n v. de la Cuesta, and Ridgway v. Ridgway, 2 SUP. CT. ECON. REV. 1 (1983) (discussing cases in which the value of assets is enhanced for both buyers and sellers when the legal system enforces negotiated agreements).

49. Anderson & Hill, *The Race for Property Rights*, 33 J.L. & ECON. 177 (1990); Barzel, *Optimal Timing of Innovations*, 50 REV. ECON. & STAT. 348 (1968); cf. Gordon, *The Economic Theory of a Common-Property Resource: The Fishery*, 62 J. POL. ECON. 124 (1954) (concluding that competition dissipates rents from exhaustible resources because the common-property character of resources encourages over-exploitation); Haddock, *First Possession Versus Optimal Timing: Limiting the Dissipation of Economic Value*, 64 WASH. U.L.Q. 775, 777 (1986) (anticipated capture of future economic rent "induces abandonment of alternatives with positive current productivity," and "at the margin, expenditures to capture title will equal value of asset sought"); Hardin, *The Tragedy of the Commons*, 162 SCIENCE 1243 (1968) (discussing how property rights to takeable resources curb natural inclination to overuse).

50. D. Haddock & F. McChesney, *supra* note 46, at 17-19; see also Haddock, Macey & McChesney, *Property Rights in Assets and Resistance to Tender Offers*, 73 VA. L. REV. 701, 715-17 (1987) (examining the problem of takeable assets in the context of tender offers).

ence maps, and zero transaction and litigation costs.⁵¹ Thus, if legal remedies always imposed ordinary damages, as in the standard liability model, incentives would exist for defendants to forego consensual exchange in favor of socially costly takings—"contractual bypass," so to speak.⁵² To discourage such takings, a legal remedy must leave a defendant who takes with no advantage vis-a-vis a defendant who bargains.

Punitive damages provide one way to strip defendants of their gains from abusing liability rules. If return of the taken entitlement is impossible, a taking of Δ must yield a remedy that leaves Δ with no more utility than a contract would have provided. In Figure 2, such a remedy would move allocations from t to m or even beyond.⁵³

Even when a court is assumed to know indifference maps—so the measurement errors addressed by the court-error model can be ignored—the court ordinarily cannot know which voluntary bargain the parties would have struck; that is, it cannot know C , the particular point along the contract curve to which the parties would have negotiated. Thus, the court cannot know u_1 , the level of utility that Δ would have attained through negotiation. A court can know for certain only that Δ would have been better off respecting π 's property protection (through negotiation) if the remedy for the taking puts Δ at or to the right of P . Moving Δ to P makes Δ whole (or "compensates" Δ , in the economist's sense) by restoring Δ 's pretaking utility. Compensation in this sense requires stripping all the gains Δ expected ex ante from the action.⁵⁴ Therefore, the

51. This possibility has been recognized previously in various contexts. See Haddock & Spiegel, *Property Rules, Liability Rules, and Inalienability: One View of the Edgeworth Box*, in PAPERS PRESENTED AT THE FIRST MEETING OF THE EUROPEAN ASSOCIATION FOR LAW AND ECONOMICS 47, 63 (G. Skogh ed. 1984) (available from the Nationalekonomiska Institutionen, Lunds Universitet, Sweden) (comparing liability rules, which indirectly award all of the gains of trade to defendants, with negotiation, which requires defendants to share the gains from trade with plaintiffs); Rose-Ackerman, *I'd Rather Be Liable Than You: A Note on Property Rules and Liability Rules*, 6 INT'L REV. L. & ECON. 255, 259 (1986) (because defendants obtain all gains from trade under a liability rule, they will actually prefer paying damages over bargaining); Veljanovski, *The Employment and Safety Effects of Employers' Liability*, 29 SCOT. J. POL. ECON. 256, 257-58 (1982) (liability rules and contract rules produce equal resource allocations only in the context of perfect information); Oi, *The Economics of Product Safety*, 4 BELL J. ECON. & MGMT. SCI. 3, 5 (1973) (different liability rules lead to different resource allocations).

52. Posner, *supra* note 7, at 1195, uses the similar concept of "market bypass." "Contract bypass" is used here to emphasize that bypassing thick (highly liquid) markets is innocuous, but bypassing thin markets that use bargaining to generate contractual exchanges is not. See *infra* notes 59-63 and accompanying text.

53. If the remedy only moves allocations to m , Δ will be indifferent between bargaining for and taking π 's entitlements. An efficiency-minded court would want to impose a remedy greater than tm , so that Δ will have a disincentive to pursue the noncontractual solution.

54. Note that the proper focus is on the ex ante expected utility of defendants. For example, a potential defendant may unilaterally decide to impose an increased risk on a potential plaintiff, even though the plaintiff could have reasonably expected that the parties would negotiate to allocate the risk. The defendant may do this even though he realizes that, under certain states of the world, his actions will injure both parties. The defendant is willing to bear the risk of injury to himself. The

Figure 2

appropriate remedy for a violation of a property protection (property violation) is "punitive": it leaves Δ worse off than negotiation would have.⁵⁵ Merely compensatory damages are insufficient to discourage takings and encourage negotiation.

Consider a libel. Publisher Δ , believing that publication of certain information about π will increase circulation, could negotiate with π for the right to publish that information. But Δ has no incentive to negotiate if expected damages for π 's injury of E_t are only tR . On the other hand, anticipation of an additional punitive award of RP would discourage

---

defendant's motivation is that the expected cost of litigation (the probability of the plaintiff's injury times the ordinary damages if an injury does occur) is less than the cost of obtaining the plaintiff's agreement to bear the increased risk, given that the plaintiff will expect to negotiate into the trading lens, not along its edge.

Whether the defendant ultimately is also injured is largely immaterial to the discussion. The defendant has already taken the ex ante risk premium that would otherwise have gone to the plaintiff because the expected ordinary damages were less than the premium the plaintiff would have demanded. Making the defendant whole entails extracting the difference. If, ex post, the defendant has also been injured, that is too bad, but it is a different matter entirely.

55. Although W. LANDES & R. POSNER, *supra* note 10, at 160-63, reach a similar result, they do so by employing the court-error model, which yields fewer predictions about legal rules and predictions that are less useful in explaining the existing legal environment. See *infra* notes 64-66 and accompanying text.

unauthorized publication and maintain property protection for π 's entitlement. A similar expectation of ordinary remedies for voluntary, intentional trespass or nuisance would offer an analogous opportunity—use or pollute π 's land and pay mere compensation, rather than negotiate for use ex ante. Since such violations have no net social value, they too call for punitive sanctions.

Enforcing property rules requires stripping all gain (or more) from a taking. In thin markets, this requires consideration of evidence concerning the defendant's utility map.⁵⁶ In a thin market, a remedy that makes the defendant whole (strips all gain but nothing more) rarely makes the plaintiff exactly whole (restores all loss but nothing more). In Figure 2, it is only point e, where i_0 and u_0 intersect, that is equivalent for both π and Δ to the initial allocation at point E, where i_0 and u_0 also intersect. Because the remedy (tR) that compensates the plaintiff often will not be the remedy (tP) that would make the defendant whole, an appropriate award must be selected for each dispute. When a property violation has occurred, restoring the defendant's initial utility is desirable, since it eliminates the defendant's incentives to ignore the property protection.

Figure 2 helps to illustrate and dispose of an objection to punitive damages raised by the standard liability model. Making Δ whole following a taking of Et yields allocations at P. The combined reallocation resulting from Δ 's taking and the court's corrective measures misses the contract curve; thus, the final allocations are neither Kaldor-Hicks nor Pareto optimal if viewed statically. An ordinary remedy tR would lie on the contract curve, yet the law rejects that remedy for an alternative that seems inefficient. It would be easy to conclude that such an alternative is inappropriate, or that some goal besides efficiency lies behind this body of law. Many scholars have reached just such conclusions,⁵⁷ and in an unintended sense, they are correct: dynamically efficient law will sometimes forsake statically efficient outcomes.

In a more dynamic frame of reference, static "inefficiency" can be

---

56. Although the plaintiff's utility map now may seem irrelevant, this is not so, for reasons that will become clear in the discussion accompanying *infra* notes 108-15. This Article treats P as the exact remedy desired in a property violation case, not as a lower-bound remedy.

Although a punitive remedy at P is superior to the ordinary remedy at R from π 's perspective, it could be inferior to the thwarted contractual remedy at C. That depends on the degree of curvature of π 's indifference curve through C. Therefore, it is incorrect to presume that an extraordinarily bad outcome for Δ necessarily implies an extraordinarily good outcome for π .

That implication of the extraordinary sanctions model stands in obvious contrast with the court-error model. The latter views punitive damages as a means of ensuring that plaintiffs are at least fully compensated for losses in the face of measurement error; the extraordinary sanctions model views the level of compensation for plaintiffs as a by-product of remedies that make defendants whole. *But see infra* notes 116-22 and accompanying text (discussing reluctant victims and insurance).

57. See, e.g., Ellis, *supra* note 3, at 77; Priest, *supra* note 3, at 124.

understood as the cost of property protection for entitlements.⁵⁸ The standard liability model mimics missing market transactions. But imitating a market is appropriate only when circumstances make it unreasonable or unnecessary for the parties to rely on a market. In a property violation case, efficient law would not help mimic a missing exchange, but instead would encourage the principals facing other potential exchanges to bargain. It is of little importance whether violators of property rules can predict their liabilities—prediction is important only for takings properly governed by liability rules. Rather, the important expectation focuses on an investor's ability to reap the maximum future benefit from socially useful present investment. In thin markets, such expectation interests are left unprotected by liability rules.

D. Market Liquidity and the Choice Between Punitive and Ordinary Damages

If plaintiffs received as remedies the allocations that would have resulted from negotiation—"full" liability remedies, so to speak—defendants would not use courts opportunistically. But such awards are infeasible in thin markets, particularly once we abandon the implausible assumption that courts know how to make litigants whole, that is, once we recognize that courts cannot estimate the litigants' indifference maps with any precision. Although difficult, the task of making π whole pales in comparison with ascertaining a counterfactual bargaining outcome. Determining a bargaining outcome requires ascertaining all conceivable bargains that might have been reached, then *guessing* which particular bargain the parties would have struck.⁵⁹ From the court's perspective,

---

58. Narrow inefficiency (a cost) in pursuit of broad efficiency (a benefit) is common in the legal system. For example, towing an illegally parked car (at a cost) is narrowly inefficient because the owner prefers that it remain where it is, and nobody especially wants it to be where it is going. The "inefficiency" is a cost of no-parking zones, which, presumably, someone does value. Indeed, litigation itself is similarly "inefficient": the motivating damage often is done, and the law's sole function is a negative sum redistribution, counting litigation cost. Without such costs, law would not exist at all, but properly designed law affords benefits. A lamented "inefficiency" often is merely a readily apparent cost that accompanies a real but undiscerned benefit.

59. In other words, an effort to make a plaintiff whole requires an estimate of only one of π 's indifference curves. But estimating the counterfactual bargaining outcome would require an estimate of a substantial portion of both π 's and Δ 's utility functions. But even that would merely enable the court to recognize the infinity of points along the contract curve to which π and Δ could potentially have negotiated. The court would still have to guess which one of that infinity of points π and Δ ultimately would have agreed upon.

Although it is unrealistic to imagine that court awards precisely restore anyone's pretaking utility, it is credible to assume that jury awards are often unbiased, low-variance estimates of the awards necessary to make one of the litigants whole. Litigants are often drawn randomly from the population. Juries are also selected more-or-less randomly. Hence, a juror's contemplation of the award that would make the juror whole under similar circumstances will lead to an unbiased estimate of appropriate compensation. Because 12 independent estimates are averaged, unusually high or low ones will tend to offset each other in the jury consensus. By contrast, a juror's

Figure 3

π 's precise loss is already ambiguous in theory and costly to measure in practice. The court's task would be several orders of magnitude more difficult if, in addition, the court were required to estimate counterfactual market outcomes.

But relaxing the assumption of a thin market, and assuming instead that reasonable quantities of the good are available at a well-defined price, makes full liability protection feasible. There is no role for extraordinary damages in thick markets.⁶⁰ In a thick market, each party faces a budget constraint passing through the initial endowments and having a slope determined by the good's price, as shown in Figure 3. If Δ takes E of π 's goods, the proper remedy will restore π 's original budget

---

contemplation of what deal would have resulted in a thin market, following a hypothetical negotiation between traders whom no juror knew well, would lead to estimates with a high variance around any potential real market outcome.

On the other hand, if a litigant is not a natural person—not a “peer” of the jurors—it is less likely that a jury will arrive at an award with the same desirable attributes. That is, it is less likely that a randomly selected jury will contain a representative distribution of shareholders, managers, and other actors found in a typical corporation.

60. Corrective damages, as we have distinguished that concept above, may be appropriate, however. See *supra* notes 26-27 and accompanying text.

constraint at R' rather than π 's original utility at R . π regains the exchange value available at E ; π can again trade to C , or back to E if no exchange had originally been attractive.⁶¹ A strict liability rule and a property rule converge, but only because the meaning of compensation is altered when it is simple and cheap to estimate a counterfactual market transaction. In effect, strict liability is full liability in a thick market.

Hence, the proper way to "make the plaintiff whole" depends upon the degree of market liquidity. In a thick market, a remedy at C (or even R') is preferable to E , because transaction costs are lower. Therefore, efficiency-minded courts will award damages and will rarely grant equitable relief. Note, however, that an efficient remedy for a taking of E in a thick market (tR') will exceed ordinary damages for a similar taking in a thin market (tR), reflecting the court's lower information costs of ascertaining the full value of a taken entitlement. Conversely, the thick-market remedy (tR') will fall short of the punitive remedy a court will impose on an opportunistic taker in a thin market (tP).

In short, the role of punitive damages (and extraordinary sanctions in general) depends on market liquidity. A strict liability rule and a property rule converge as markets become more liquid and the incentives for Δ 's opportunism diminish. As this occurs, the rationale for extraordinary damages also diminishes. In a thick market, an allocation at R' makes both π and Δ whole relative to the contractual outcome at C (neglecting transaction costs), because each of them can exchange to C from R' . A liability rule in a thick-market setting requires little sacrifice of the benefits of property protections, because movement to the budget constraint restores π 's full potential for gains from exchange by giving π the resources to trade to the desired point along the budget constraint.⁶²

Even in thin markets, full liability protection can be had without recourse to punitive damages when the asset taken remains in existence. In this case, the sanction is equitable relief—extraordinary not in amount but in kind. If the asset is still available for court-ordered return to its owner, such relief strips away Δ 's gains. No monetary sanction is required to restore Δ to the pretaking utility level u_0 . A remedy at E is obviously preferable to one at R , since it safeguards property protections where Δ has chosen taking over negotiation. Ordinarily, an injunction restoring the parties to E will be superior to P , since the court need not then estimate the superordinary damages tP needed to put Δ back onto

61. If Δ had anticipated paying compensation, Δ would have taken only E' in order to move directly to C , because Δ must bear modest transaction costs to reach C from R' .

62. The thick-market/thin-market distinction is yet another point of departure between the present model and the illicit benefits construct. The illicit benefits model, described at *supra* notes 34-36, offers no prediction that the benefits derived from a taking will be deemed illicit only if the taken item is traded (if at all) on an illiquid market.

u_0 , a costly and error-prone process. Injunctive relief that restores the item taken returns Δ to u_0 with no possibility of error.⁶³

Thus, punitive damages are unnecessary to make defendants whole in two general cases: when litigants can be restored to the initial budget constraint in a thick market, and when litigants can be restored to E in a thin market. Indeed, under such circumstances, punitive damages would be undesirable, as will be shown in Part II.

E. Competing Error-Based Explanations

As mentioned above, other analyses suggest that courts should impose punitive damages to strip defendants of their gains. But those explanations do not rest on the desirability of making defendants whole. Based on the standard liability model, those analyses instead treat making plaintiffs whole as the first-best norm for optimal deterrence, while hypothesizing that court errors require superordinary damages as a second-best alternative. But this emphasis on court error is misplaced. Imperfect court judgment means merely that the variance of the expected remedy is positive. In thin markets, the variance of expected bargaining outcomes in the absence of takings is also positive. Why is variance in legal outcomes a more serious concern than the variance in market outcomes? Moreover, as already noted, court error does not necessarily produce bias in awards.

---

63. For an extended discussion of injunctions, see *infra* notes 88-91 and accompanying text.

Another type of extraordinary sanction, restitution, is sometimes elected by tort plaintiffs in those relatively rare cases where direct evidence of the magnitude of the benefit to the defendant is cheaply available:

[W]here the commission of a tort results in the unjust enrichment of the defendant at the plaintiff's expense, the plaintiff may . . . "waive" the tort action, and sue instead on a theoretical and fictitious contract for restitution of the benefits which the defendant has so received. . . . Restitution in quasi-contract, on the other hand, looks to what the defendant has received which in good conscience should belong to the plaintiff; and this may be either more or less than the amount of plaintiff's actual loss.

PROSSER AND KEETON, *supra* note 4, § 94, at 672-73 (footnotes omitted).

Analytically, restitution is a defendant-revealed measure for punitive damages. In practice, however, resort to restitution requires the plaintiff to produce evidence of the existence and extent of the tortious benefit, that is, the distance tP in Figure 3. Since it is difficult in an illiquid market to establish the extent of the defendant's gain, some courts do not permit suit for restitution unless the defendant has subsequently sold the taken asset and thereby revealed the precise amount of his benefit. See *id.* § 94, at 673-74.

By contrast, fact-finders awarding punitive damages can often make workable educated guesses of the size of the remedy tP in the absence of quantitative evidence. See *supra* note 59. Thus, the burden of proof weighs much more heavily on the plaintiff seeking restitution than on the one seeking punitive damages. Moreover, it is not legally sufficient for a plaintiff seeking restitution to show that the population of potential defendants benefits in aggregate from similar activity and that each individual defendant thus had an *ex ante* expectation of gain. The remedy of restitution requires a showing of an *ex post* benefit to the *particular* defendant before the court. In other words, restitution "is not permitted where the defendant has merely damaged the plaintiff, negligently or otherwise, without benefit to himself." PROSSER AND KEETON, *supra* note 4, § 94 at 673 (footnotes omitted). Punitive damages, in contrast, remain a possibility in such circumstances.

Other points of difference exist between the present extraordinary sanctions model and the court-error model. In explaining punitive damages in terms of court error, for example, Shavell remarks that "[w]ere courts to set liability equal to their best estimate of losses, . . . [a]n injurer might then engage in his activity when his benefit exceeds a court's estimate of losses rather than, as under bargaining, when his benefit exceeds true losses."⁶⁴ But as shown above, the plaintiff's "true losses" are not uniquely defined. Although the plaintiff could have attained some point along the contract curve, it is impossible to determine which point the plaintiff would have obtained. Shavell's statement is ambiguous, mingling errors in ascertaining compensation that would make plaintiffs whole with errors in estimating bargaining outcomes. Punitive damages cannot correct errors of the former type, except to guarantee that they will all be positive; punitive damages *can* eliminate errors of the latter type by eliminating the need to estimate bargaining outcomes at all. Thus, court error, though doubtless a feature of legal life, provides little aid in understanding punitive damages.

Both Posner and Shavell seem to believe that whenever a defendant could have chosen between taking and negotiation (and ordinary damages are ascertainable without error), making plaintiffs whole deters optimally, so punitive damages are unnecessary.⁶⁵ That conclusion is incorrect. Regardless of errors in making plaintiffs whole, courts ought to abandon the notion of optimal deterrence in many thin-market situations that provide the defendant with a reasonable choice between taking and negotiating. In those situations, courts should seek to eradicate, not optimize, takings in order to strengthen property protection. Punitive damages will then provide a first-best incentive. Thus, the presence or absence of error in reckoning ordinary damages is irrelevant—Occam's "plurality assumed without necessity."⁶⁶

---

64. S. SHAVELL, *supra* note 16, at 149.

65. *Id.*; R. POSNER, *supra* note 20, at 34.

66. Several other error-based analyses of punitive damages also rest on the assumed desirability of making plaintiffs whole in order to achieve optimal deterrence. It is suggested, for instance, that uncertainty created by judicial errors in assessing liability overdeters defendants, and that increasing the standards for (lowering the probability of) liability and imposing more punitive sanctions could both mitigate the overdeterrence problem and achieve optimal compensation for plaintiffs as a class. *See, e.g.,* Johnston, *supra* note 20, at 1390 (using the low-probability/high-sanctions model to describe the current legal system); Png, *supra* note 20, at 103-04 (assessing overdeterrence in the context of innocent parties paying damages). Admittedly, assessing punitive damages increases the likelihood that plaintiffs will be at least adequately compensated when injuries are difficult to value. In that sense, punitive damages resemble specific performance in contract, which is explained as avoiding erroneous undercompensation of plaintiffs when damages are difficult to value. *See, e.g.,* Kronman, *Specific Performance*, 45 U. CHI. L. REV. 351, 355 (1978); Schwartz, *The Case for Specific Performance*, 89 YALE L.J. 271, 272 (1979). Note, however, that this view of punitive damages assumes that court errors will be biased, that is, that courts will systematically undercompensate rather than overcompensate.

F. *Implications of the Model*

Many law and economics scholars believe that punitive civil damages have no efficiency rationale beyond correcting for underdeterrence or court error.⁶⁷ To the contrary, the extraordinary sanctions model here has described situations in which an expectation of punitive sanctions would advance efficiency. To what extent can the present model predict observed features of the law of punitive damages?⁶⁸

1. *Thin Markets and Punitive Damages*

The model does in fact explain several features of the law of punitive damages. Case law bears out the model's prediction that punitive tort sanctions would commonly be observed only in cases involving assets traded in illiquid markets. Charles McCormick summarizes the cases:

Violations of rights of bodily immunity are among the most frequent instances, and such [punitive] damages are constantly given in cases of assault, personal injury due to recklessness, and false imprisonment. . . . Recovery is allowed for infractions of interests of personality, such as slander and libel. Less common but well recognized is the liability [for] this type of relief in violations of pecuniary interests, whether in intangibles, as in cases of deceit, malicious interference with business relations, and infringement of trade-marks, or rights in tangible property. Of the last class are actions for the conversion of personal property and for intrusions upon land, destruction of shade trees, nuisance, pollution of streams, and interference with easements. . . . [T]he award of punitive damages in cases of injuries to family relations is frequent, as in cases of seduction, deceit in inducing marriage, and alienation of affection.⁶⁹

None of the assets discussed in the quotation—bodily integrity, reputation, land, specific capital, and so forth—comes in homogeneous units traded in liquid markets. They are precisely the sorts of property that, if taken rather than negotiated for, should persuade courts to award plaintiffs punitive damages.

2. *Making Defendants Whole*

The extraordinary sanctions model is based on the inapplicability of the standard liability model to thin-market takings. In those situations,

---

67. See, e.g., Ellis, *supra* note 3; Priest, *supra* note 3.

68. It is of methodological interest that the authors were unacquainted with many of the features of law discussed throughout this article until the implications of the model motivated the research reported here that delves into the rules that govern under specific circumstances. Thus, the predictive value of the model is not skewed by foreknowledge of the desired outcome; development of the model did not involve drawing bullseyes around arrows already shot into a board. For a criticism of that practice, see Balkin, *Too Good to Be True: The Positive Economic Theory of Law*, 87 COLUM. L. REV. 1447, 1461 (1987) (reviewing W. LANDES & R. POSNER, *supra* note 10).

69. C. MCCORMICK, HANDBOOK OF THE LAW OF DAMAGES 287 (1935).

defendants (not plaintiffs) must be made whole, unless transaction costs in the market are prohibitive. Punitive damage law does implicitly strive to "make defendants whole." No plaintiff "entitlement" to punitive damages exists; the award depends on particulars of a case that concern only the defendant, not the plaintiff.⁷⁰ The issue is not *what* was done (clearly, the plaintiff's entitlement was taken), but *why* (was it to avoid negotiation?): "It is usually the defendant's mental state that is said to justify a punitive award against him, rather than his outward conduct."⁷¹ Punitive liability is variously said to result because acts are "inalicious," "evil," "wicked," "morally culpable," or some similar description,⁷² although, as noted above in the discussion of the illicit benefits model, such terminology is conclusory rather than explanatory.

Such nebulous terminology in the case law could easily confuse commentators, but the explanation is fairly simple. For the many sorts of cases noted earlier where punitive damages are common, the "evil" act was the defendant's decision to take instead of acquire by voluntary agreement. For example, even utterly scurrilous lies would motivate no libel award to a plaintiff who had given clear advance approval. The legal evil, then, is not the content of the publication, but the contractual bypass.

Notably, courts will impose punitive damages only for intended or negligently mistaken takings.⁷³ For example, if a defendant intentionally trespasses the law will impose punitive damages.⁷⁴ Yet if a defendant exercising due care accidentally builds on another's land, the law will treat the trespass as a "legitimate" taking, and a court will order only ordinary damages.⁷⁵ Likewise, misappropriation of trade secrets or protected commercial ideas will result in punitive damages only if done intentionally.⁷⁶

In the extraordinary sanctions model, well-defined property rights are a necessary condition for the award of punitive damages. Since defining entitlements is costly, an efficient legal system will not define them for every potentially valuable resource; movement of assets into and out of

---

70. See, e.g., *Luke v. Mercantile Acceptance Corp.*, 111 Cal. App. 2d 431, 437-38, 244 P.2d 764, 769 (1952).

71. D. DOBBS, *supra* note 11, at 205; see also Grady, *Punitive Damages and Subjective States of Mind: A Positive Economic Theory*, 40 ALA. L. REV. 1197, 1219-24 (1989).

72. See, e.g., *Santiesteban v. Goodyear Tire & Rubber Co.*, 306 F.2d 9, 12 (5th Cir. 1962) ("malice"); *Cooperative Refinery Ass'n v. Young*, 393 P.2d 537, 540 (Okla. 1964) ("evil intent").

73. See W. LANDES & R. POSNER, *supra* note 10, at 149-153 (providing an extensive analysis of alternative plausible meanings of intent).

74. D. DOBBS, *supra* note 11, at 347 (noting cases).

75. Cf. *Dolske v. Gormley*, 58 Cal. 2d 513, 520-21, 375 P.2d 174, 179, 25 Cal. Rptr. 270, 275 (1962) (in ruling on request for mandatory injunction, court should consider good faith of party who constructed the encroachments).

76. See, e.g., *Williams v. Weissner*, 273 Cal. App. 2d 726, 743, 78 Cal. Rptr. 542, 551-52 (1969).

common ownership is a constant process.⁷⁷ As a resource becomes more valuable, the likelihood that claimants will try to capture previously ambiguous entitlements increases. As inconsistent claims to a resource arise, court adjudication to settle the claims often becomes necessary. In cases of ambiguous entitlements, the present model indicates that courts should not award extraordinary damages, since no one has unambiguously committed a property violation. This outcome is consistent with cases that address the subject.⁷⁸

The costs of property violations and the corresponding benefits of making defendants whole also explain punitive damages in "spite" cases where torts have been committed specifically to harm the plaintiff. Spite cases also illustrate the difference between the extraordinary sanctions model presented here and the illicit benefit model of punitive damages criticized above. Suppose Δ wants an apple; suppose also that he hates an orchard-owning neighbor π . Rather than buy an apple to eat, Δ takes π 's apple, expecting to be detected—perhaps even desiring detection so that π will be reminded of the hatred.

To account for the punitive damages that courts often award in such situations, commentators like Ellis and Cooter term the satisfaction that one person gets from inflicting an injury on another "illegitimate" or "illicit."⁷⁹ This illicit benefit supposedly necessitates the imposition of punitive damages in order to achieve optimal deterrence.⁸⁰ The commentators give no explanation for courts' refusal to count the defendant's utility in such cases; it is simply "a normative decision."⁸¹ In effect, Δ has no right to injure π because Δ has no right to injure π . This is no explanation at all.

In addition, extraordinary damages in such situations are inconsistent with the supposed optimal deterrence function of law in the

---

77. See Demsetz, *Toward a Theory of Property Rights*, 77 AM. ECON. REV. 347, 350-53 (Papers and Proceedings of the American Economic Association, May 1967) (describing the conditions necessary for the emergence of property rights); Anderson & Hill, *The Evolution of Property Rights: A Study of the American West*, 18 J.L. & ECON. 163 (1975) (examining "dynamic" evolution of property rights during the development of the American West).

78. One of the best-known examples of a court limiting damages to compensatory amounts in a "takings" situation where entitlement ownership was ambiguous is *Spur Industries, Inc. v. Del E. Webb Development Co.*, 108 Ariz. 178, 494 P.2d 700 (1972), where the court refused to impose the extraordinary (though normal) remedy of an injunction to abate a "nuisance."

79. See Cooter, *supra* note 3, at 90; Ellis, *supra* note 3, at 32-33.

80. Cooter, *supra* note 3, at 90; Ellis, *supra* note 3, at 32-33.

81. Ellis, *supra* note 3, at 32. For an illustration of the differing weights that scholars put on the defendant's utility in "illicit benefits" cases, compare Landes & Posner, *An Economic Theory of Intentional Torts*, 1 INT'L REV. L. & ECON. 127, 136 (1981) (arguing that wrongful conduct should be sanctioned without concern for overdeterrence) with Ellis, *An Economic Theory of Intentional Torts: A Comment*, 3 INT'L REV. L. & ECON. 45, 53-54 (1983) (arguing that gains to the injurer must be included in computing costs of avoidance). Landes and Posner have taken issue with Ellis' characterization of their argument. See W. LANDES & R. POSNER, *supra* note 10, at 157 n.9.

standard liability model, in which deterrence is achieved by making plaintiffs whole.⁸² If plaintiffs are supposed to be made whole, Δ 's pleasure (utility) from the taking is irrelevant—indeed, everything about Δ is irrelevant. According to the standard liability model, if damages are accurately assessed (with corrective adjustments when necessary), then no extraordinary damages are appropriate, because they would induce excessive care by potential defendants.⁸³

The model here shows that an efficient legal system would try to eradicate property violations, not seek to optimize them.⁸⁴ The marginal conditions applying to Δ 's activity level would be irrelevant. Only totals would matter— Δ 's expected cost, including punitive damages, would have to exceed Δ 's expected benefit. In spite cases, the defendant's utility is negatively interdependent with the plaintiff's, meaning that the defendant double-counts the taking; Δ values π 's apple first because Δ likes apples, then again because the taking harms π . To retain property protections, greater judgments against a spiteful Δ are required than against an unspiteful Δ . If π 's and Δ 's utilities are negatively interdependent, returning Δ to his original utility level requires stripping away both the value of the taken asset and the value to Δ of the insult inflicted on π .

The distinction between a spiteful and an unspiteful Δ turns on a difference in the magnitude of gains to Δ (to be reflected in the magnitude of appropriate sanctions), not on degrees of "illegitimacy." Δ 's taking of apples from a neighbor creates the same sort of loss whether or not Δ hates his neighbor; the activity is "illegitimate" in either case. Merely obtaining an apple does not convey illicit benefits, and, by hypothesis, the unspiteful defendant places no other importance on the taking. Consequently, the illicit benefits model implies that no extraordinary damages would follow a dispassionate taking. The present model indicates otherwise: extraordinary damages are still appropriate to force potential defendants to negotiate with potential plaintiffs. But the minimum efficacious punitive damages will be less when utilities are independent, because no double counting of π 's injury influences Δ 's behavior.

In the extraordinary sanctions model, punitive damages are tailored to the defendant's preferences (utility map) while ordinary damages are tailored to the plaintiff's preferences. This relationship clarifies a long vexing problem—the appropriate relationship between ordinary and

---

82. Extraordinary damages are, by contrast, wholly consistent with the preventive function achieved by making *defendants* whole.

83. See *supra* text accompanying notes 28-29.

84. The model assumes that court costs are zero. In practice, it is costly to discourage property violations, so efficient law would not seek total eradication. Nevertheless, any restraints imposed would be intended to optimize the behavior of courts, not that of potential defendants. Friedman, *supra* note 33, views positive court costs as a sufficient rationale for the imposition of extraordinary sanctions.

punitive damages. Courts have had difficulty in defining the relationship; many hold that there is none, that the size of punitive awards varies solely with the defendant's motive and purpose.⁸⁵ The model as developed so far indicates that no relationship should exist, because the magnitude of punitive damages in a given case depends only on the shape of the defendant's indifference curves. Other courts have held that punitive awards must bear some relation to ordinary damages.⁸⁶ As shown below, that too is consistent with the model, provided that the requirement is invoked only when plaintiffs can undertake efficient precautions to reduce the likelihood of a property violation.⁸⁷

3. Injunctions

The distinction between static and dynamic efficiency explains courts' willingness to order injunctions when a taking has occurred, even though injunctions, like punitive damages, place the parties off the contract curve. Punitive damages require an accurate estimate of indifference curves if Δ is to be made whole; an injunction makes Δ whole directly without any such estimate. If it costs less to replicate a bargaining environment than to estimate the appropriate compensation, then an injunction, which then forces the acquiring party to bargain for the good it took, is an efficient way for a court to maintain property protections. As noted above, if entitlements merit property protection in thin markets, then restoring initial entitlements is preferable to making plaintiffs whole. Point E in Figure 3 will be preferable to R, that is, providing the asset is essentially unaltered. From E, the parties can still negotiate to C. The bilateral monopoly tolerated before is essentially unchanged. The parties could not bargain to C if placed at R.

Moreover, there is no reason for additional extraordinary monetary damages when a court returns the parties to E by granting an injunction. Damages can return Δ to the indifference curve u_0 passing through E, but an injunction returns Δ to E and thus u_0 with certainty. An injunction is equivalent to adding punitive damages of RP to ordinary damages of tR without actually determining R and P, an equivalent that conserves a court's resources and avoids counterfactual conjecture.

---

85. See, e.g., *Bucher v. Krause*, 200 F.2d 576, 587 (7th Cir. 1952) (Under Illinois law, punitive damages "do not to any extent depend upon the amount of actual pecuniary damage sustained, but depend wholly upon the motive, purpose and condition of mind and heart of the wrongdoer and the circumstances and manner of his doing the wrong."), *cert. denied*, 345 U.S. 997 (1953).

86. See, e.g., *Hughes v. Babcock*, 349 Pa. 475, 481, 37 A.2d 551, 554 (1944). Courts often consider the defendant's financial worth in defining the relationship. See, e.g., *Aland v. Pyle*, 250 Pa. 1170, 106 A. 349 (1919); cf. *Thomas v. E.J. Korvette, Inc.*, 329 F. Supp. 1163 (E.D. Pa. 1971) (discussing the seeming inconsistency between these two propositions), *rev'd on other grounds*, 476 F.2d 471 (3d Cir. 1973).

87. See *infra* notes 108-15 and accompanying text. Efficient control of property violations then will require sanctions in addition to, or instead of, extraordinary awards to plaintiffs.

Thus, the extraordinary sanctions model predicts that extraordinary damages will be unavailable when equitable relief is granted, regardless of intent or negligence. Although that has long been the rule, it has perplexed commentators. Dan Dobbs writes that

[i]f the defendant's misconduct is sufficient in law courts to justify a punitive award, one might expect that equity courts would be equally willing to make such an award. Strangely enough, the traditional rule is otherwise, and a majority of courts that have examined the point probably still refuse to grant punitive damages . . . . The reasons for the rule against punitive awards in equity are sometimes difficult to understand . . . .⁸⁸

Not only were those reasons difficult for Dobbs to understand, they are difficult to understand under the illicit benefits model as well. If an offense produces illicit benefits, why should courts decline to strip them away through imposition of punitive damages merely because an injunctive remedy is available? Our model predicts and explains such a legal rule; the illicit benefits model does not.

Recall that the extraordinary sanctions model predicts that courts will grant injunctions when the taken asset is "essentially unaltered." In many cases, an alteration will reduce or destroy the asset's value; however, in some cases, an alteration may increase its value. In the latter cases, as much as in the former, court-ordered return of the altered asset cannot restore the initial entitlements at E. What does the extraordinary sanctions model predict in such a case?

First, when a defendant has unintentionally and nonnegligently taken an asset that is traded in thin markets and made an investment in it, such that it cannot be returned in its original state, a court should award only ordinary damages.⁸⁹ Suppose, for example, that working under a good-faith but mistaken belief concerning the title to certain land, Δ has built upon it. Since destroying the building would be wealth-reducing and also unnecessary for discouraging intentional takings, an attempt by the court to return the entitlement in its initial condition would be misguided. A court cannot now protect both parties' assets; if the investment is to be retained, either π must have the nonnegligent Δ 's building, or Δ must have the faultless π 's land. The court can infer that Δ is more likely to make efficient use of the improved property. Although the court might presume that π expected to use the land to good purpose, it could not know whether π would make good use of Δ 's building. The court would know by Δ 's actions, however, that Δ expected to use both the building and π 's land profitably. Consequently, the court would sanction the taking, awarding only ordinary, compensatory damages to π . In fact, courts apply that rule when a defendant's

---

88. D. DOBBS, *supra* note 11, at 211.

89. See *supra* note 75 and accompanying text.

taking is innocent and site-specific investments have ensued.⁹⁰

Second, when a defendant has intentionally or negligently taken an asset traded in thin markets and made an investment in it, a court should award extraordinary damages. In contrast to the previous example, suppose that Δ had taken real estate not by innocent means and had built upon it. Injunctive relief would require destroying the building or putting it in the hands of π , who would presumably use it less beneficially. Neither outcome is desirable; neither is necessary. The model predicts that Δ should keep the land, but pay extraordinary damages to π that put the parties at point P. Once again, that seems to be the legal rule.⁹¹

4. *Court-Error Models Distinguished: Nominal Damages, Vicarious Liability, and Insurance*

A number of features of punitive damage law predicted by the extraordinary sanctions model are inconsistent with the court-error model, in which punitive damages are imposed because of difficulty in valuing injury to plaintiffs. For example, if a plaintiff has clearly suffered no compensable injury, only nominal damages are at issue and valuation is error-free. Suppose that a plaintiff's land has been trespassed upon, but without damage. Or suppose that a plaintiff is defamed in a locale where he is unknown, so that no loss of reputation results. With no compensable injury alleged, an error in ascertaining how to make the plaintiff whole is unlikely, so the error model would deny punitive damages. The extraordinary sanctions model predicts just the opposite: to prevent defendants from "gaming" the system by taking instead of negotiating, courts should award punitive damages even though no compensatory damages would be awarded. In fact, courts often award punitive dam-

---

90. See D. DOBBS, *supra* note 11, at 205 ("[A]nything that negates a bad state of mind will usually preclude the punitive award, as for example, where the defendant is found to have acted in good faith, or on advice of counsel.").

91. See *Boomer v. Atlantic Cement Co.*, 26 N.Y.2d 219, 257 N.E.2d 870, 309 N.Y.S.2d 312 (1970). *Boomer* is one of the most studied cases of an intentional taking. Though it is usually discussed as a case of compensatory damages, recent research indicates that this characterization is inaccurate. The record indicates that on remand from the cited decision, "[t]he trial judge agreed with the plaintiffs that damages would not be limited to the decrease in fair market value" and that "the damages ultimately awarded exceeded the decline in the fair market value of the land." D. Farber, *Reassessing Boomer: Justice, Efficiency, and Nuisance Law* (Working Paper No. 37, John M. Olin Program in Law and Economics, Stanford Law School, Oct. 1987). *Boomer* involved only a few foreseeably injured parties, but the defendant chose to be a nuisance rather than negotiate with them. Farber implies that it was only "the size of [the defendant's] investment, its use of the best available pollution control, and the size of its work force" that induced the court to award plaintiffs extraordinary damages rather than an injunction. *Id.* at 7-8.

See also *Orchard View Farms, Inc. v. Martin Marietta Aluminum, Inc.*, 500 F. Supp. 984, 1025 (D. Or. 1980) ("Because the [defendant] company did not cooperate . . . in arranging to prevent this damage or to neutralize it through voluntary compensation arrangements, the company is liable to the plaintiff for an award of punitive damages.").

ages in such cases.⁹²

In other cases where the present model shows punitive damages to be undesirable, the error model generates no obvious predictions. Punitive damages for vicarious liability are an example. The extraordinary sanctions model predicts that if an agent takes rather than negotiates, thus justifying punitive sanctions against the agent, the agent's principal will escape liability for extraordinary damages unless the taking was authorized, or unless the principal failed to control the agent's foreseeable behavior prudently. In fact, the usual rule bars vicarious liability for punitive damages: "In the federal courts and in the majority of state courts, a principal cannot be held for exemplary damages for acts done by an agent unless it be proven that the principal has participated in or ratified the agent's wrongdoing."⁹³ The model developed here likewise would predict that, even if a public entity (for example, a municipal corporation) has no statutory or constitutional immunity against suit, it would be immune from punitive damages for its agents' acts, subject to the exceptions discussed above. This, too, seems to be the rule, although fewer cases exist.⁹⁴ The court-error model is not so much inconsistent with these doctrines as it is silent with respect to them.

Nor does the error model correctly predict the legal treatment of insurance. If the goal of both compensatory and punitive damages is to make the plaintiff whole, the law should enforce insurance contracts that pay for punitive damages awards against defendants. By contrast, the extraordinary sanctions model would deny enforcement of such insurance contracts. Insurance coverage of punitive sanctions thwarts the sanctions' intended effect of encouraging negotiation and discouraging takings, since the defendant himself is not "punished."⁹⁵ Thus, the model predicts that courts would refuse to enforce insurance clauses that provided reimbursement for punitive awards; indeed, that has become

---

92. See C. MCCORMICK, *supra* note 69, at 293 ("While a few courts have held that to sustain an award of exemplary damages there must be . . . substantial actual damage, the majority view is that a finding of a malicious wrong giving rise to a cause of action, even though only for nominal damages, is sufficient.").

93. *Id.* at 282. Some cases do impose punitive damages on principals under the doctrine of respondeat superior. See *id.*; D. DOBBS, *supra* note 11, at 214. The extraordinary sanctions model implies that punitive damages will be imposed on only those principals who have intentionally failed either to instruct their agents or to control their agents' predictable behavior.

94. D. DOBBS, *supra* note 11, at 217. A general rule is difficult to formulate because courts are uncertain whether they can impose punitive damages on public entities without statutory authorization. The Federal Tort Claims Act specifically bars such awards against the federal government. 28 U.S.C. § 2674 (1988).

95. Given the possibility of moral hazard, a preliminary issue is whether insurance companies would even intend to write such coverage. In fact, there has been considerable litigation on this point. "Before 1962, the cases that took up the insurer's liability for punitive damages based on the insured's malicious acts considered mainly the contract issue: did the policy provide for such coverage?" D. DOBBS, *supra* note 11, at 216.

the rule.⁹⁶ Notably, however, in the minority of jurisdictions that award vicarious punitive damages without inquiring whether the principal has participated in or ratified an agent's acts, the model predicts that principals could insure against punitive damages, since punitive damages in such jurisdictions do not "punish" the principal for bypassing a contract. That is also the rule.⁹⁷

In short, punitive damages are appropriate to make defendants whole if their strategic behavior would otherwise reduce aggregate wealth. Illiquid markets and heterogeneous assets are conditions necessary to make such awards useful in an efficient legal system. Measurement error in compensating plaintiffs is unnecessary and inadequate to explain the existing legal rules. Indeed, the court-error model can hardly be predictive, since it does not even attempt to establish when error is substantial enough to require punitive damages. The extraordinary sanctions model, which focuses on eradicating certain behavior by making defendants whole, is simpler and has greater predictive power.

5. *Privately-Stipulated Extraordinary Sanctions*

Finally, the model of punitive damages in tort presented here has implications for a recurring question in contract law: whether to enforce damages stipulated by the contracting parties ("liquidated damages" provisions). Many scholars believe that courts are right to refuse to enforce such clauses (calling them "penalties") when the stipulated damages exceed ex ante expected losses.⁹⁸ Such damages supposedly could only be either a mutual mistake or an attempt by an unscrupulous promisee to take advantage of an unsuspecting promisor.⁹⁹ By this rationale, stipulated damages should be unenforceable whenever they exceed a promisee's expected losses.

The present model indicates that this argument is wrong. Liquidated damages that exceed expected ordinary damages merely award some or all expected gains from breach to the promisee rather than the

96. The seminal case is *Northwestern Nat'l Casualty Co. v. McNulty*, 307 F.2d 432, 433-34 (5th Cir. 1962).

97. Dobbs makes this point in noneconomic terms:

[W]hile public policy might well preclude the malicious individual's shifting his punishment to an insurer, it very probably would not preclude that same shifting by one who is only vicariously liable for punitive damages. Thus, in the case of an employer who is held in punitive damages for an employee's malicious conduct, insurance seems to be an entirely reasonable solution, and . . . has received approval as consistent with the *McNulty* rule.

D. DOBBS, *supra* note 11, at 216 (footnote omitted).

98. See, e.g., Rea, *supra* note 8, at 167 ("There are no strong economic arguments for enforcing damages that are unreasonably large ex ante, and the doctrine can be justified as a method of identifying cases of mistake or unconscionability.").

99. See *id.*

promisor.¹⁰⁰ That solution leaves parties on the contract curve but results in a different, though agreed upon, allocation of the gains from breach. Thus, it must be as efficient in the short-run as one that awards all the gains from trade to the other party. It is also more efficient in the longer run, for some promisees are more likely to discover the opportunities to efficiently redirect the contracted-for goods. Liquidated damages provisions give plaintiffs as well as defendants an incentive to find situations in which breach would be efficient, rewarding them in accordance with their expected contributions to creating gains.

The efficiency impact of the ability to allocate gains from breach is particularly clear when a higher-valuing purchaser ϕ has been found. When courts allow Δ to breach and then sell to ϕ , requiring only compensation for π 's provable losses, they give all gains from the breach to the breacher Δ . This is appropriate, however, only if Δ alone can find ϕ . But sometimes π will be better positioned to find ϕ . π and Δ each benefit if π can expect (through the liquidated damages) to reap part of the benefits from allowing Δ to deliver the contracted-for goods to ϕ . Liquidated damages encourage π to search for ϕ . Because ϕ then is more likely to be found, the initial contract price Δ receives will be higher. Since π incurs costs when searching for a higher valuing purchaser, only property protection of the resale entitlement purchased from Δ will induce an efficient amount of search for ϕ . Properly designed liquidated damages clauses establish just that protection for π 's entitlement.

In other words, π and Δ will divide the gains from prospective reallocation of the contracted-for asset according to their relative abilities to locate opportunities for gain—if the law permits it. If Δ has sold π some or all of those prospective gains, an efficient legal system would enforce the division for which the parties had bargained, or (what is functionally equivalent) it would impose punitive damages in the event of a violation of the agreement.

A contract-law rule against awarding liquidated damages in excess of actual losses is thus completely inconsistent with courts' willingness to award punitive damages in tort. Both court-ordered punitive damages and privately stipulated damages are devices for altering the short-run division of the gains from trade; both deviate from an award that would make the plaintiff whole. But it is the long-run property-rights implications of both punitive awards in tort and liquidated damage clauses in contract that make each desirable in appropriate situations.¹⁰¹ An

100. This result of our general model is consistent with the analysis of liquidated damages in Goetz & Scott, *supra* note 8, at 593-94.

101. As has been recognized by other writers, the availability of extraordinary awards can lead to opportunistic behavior by potential plaintiffs. See Clarkson, Miller & Muris, *supra* note 8, at 368-71 (liquidated damages); S. SHAVELL, *supra* note 16, at 194-97 (punitive damages); Cooter, *supra*

efficient and consistent legal system would not only award punitive damages in tort, but also would give effect to liquidated damage clauses in contract—even when the parties had specified amounts that exceeded expected actual damages.¹⁰²

Thus, the model of extraordinary sanctions, initially developed in the context of punitive damages in tort, can readily be applied to the realm of liquidated damages clauses in contract. Part II will discuss other efficient legal rules implied by the model.

II

APPLICATIONS OF THE EXTRAORDINARY LEGAL SANCTIONS MODEL

The diagrams presented above showed that if potential taker Δ expects to be returned to the initial utility level (u_0) by a perfectly enforced property rule, Δ has no incentive to take π 's property. Where π 's property is protected only by a liability rule, Δ will limit takings to E_t in a thin market or to $E_{t'}$ in a thick one. Consequently, Δ will take more than E_t only if underdetection, underconviction, or inadequate sanctions are expected. Such expectations are characteristic of those parties committing crimes such as theft and murder. If a taking exceeds E_t , no comparable voluntary trade could have taken place, since reaching the contract curve at a height of less than t in Figure 3 would leave π worse off than at the initial entitlement at E . The problem for an efficient legal system is to select, wherever possible, a response that will alter Δ 's expectations of opportunistic gains at π 's expense. Corrective multiplicative awards are appropriate to adjust for systematic underdetection or underconviction. But quite apart from corrective awards, extraordinary sanctions will be as appropriate in the criminal context as in the civil.¹⁰³

Clearly, the dynamic losses that made R unattractive in Figures 1 through 3¹⁰⁴ also make unattractive any remedy that leaves π worse off

---

note 3, at 96-97 (punitive damages). This problem is addressed in *infra* notes 108-15 and accompanying text.

102. *Accord* C. MCCORMICK, *supra* note 69, at 290. The extraordinary sanctions model predicts that courts will assess punitive damages in contract less frequently than they will in tort, because parties to a contract have an opportunity to, and often do, negotiate *ex ante* for appropriate liquidated damages. Indeed, punitive damages for breach of contract, though legally permissible, are rare. D. DOBBS, *supra* note 11, at 818. Our model also predicts, however, that when courts refuse to enforce liquidated damage clauses under particular circumstances, they will be more willing to impose appropriate punitive damages instead.

103. Becker, *supra* note 7, at 177-79, argues that intentional underdetection coupled with corrective multiplicative awards constitutes efficient law enforcement. *But see* Stigler, *supra* note 7, at 526-28 (discussing circumstances in which Becker's argument fails because it ignores appropriate marginal deterrence).

104. *See supra* notes 45-50 and accompanying text.

than he would be at point C.¹⁰⁵ For example, if Δ steals the amount ES in Figure 4, but an award of SR'' returns π only to i_0 , then π loses the exchange value associated with ES. Δ is, of course, better off at S than at t, so potential defendants have increased incentives to invest time and other resources locating takeable assets.¹⁰⁶ Even if caught, a convicted thief would still find R'' preferable to C, because R'' lies on a higher indifference curve (u_2 rather than u_1). Consequently, any incentive to forsake thievery would be too small. Asset owners will counter by expending resources to avoid theft. Theft imposes costs without counter-vailing benefits; there is nothing to optimize, only something to eradicate. To avoid the costs arising from Δ 's opportunism, larcenous defendants, not victimized plaintiffs, should be made whole.¹⁰⁷

For thefts less than Ee' , a remedy that makes Δ whole exceeds one that would make π whole. This seems to imply a superordinary award for π , as discussed above with respect to punitive tort damages. The appropriate remedy for some range of thefts is thus analogous both to the extraordinary sanctions desirable when Δ takes rather than negotiates (fully expecting to compensate π with ordinary damages) and to privately-negotiated liquidated damage clauses in contract. In such instances, however, a superordinary award may create a complication—the moral hazard problem of "eager victims."

A. Eager Victims

The more the expected award following a taking exceeds π 's market expectation, the greater is π 's incentive to behave in a way that increases the probability that π 's property will be taken. Whenever there are potential "eager victims," as such plaintiffs will be called, courts must limit π 's award as well as reckon the optimal sanction against Δ .¹⁰⁸

105. Point C in Figure 3 is determinable only in liquid markets. See *supra* notes 59-63 and accompanying text.

106. Tullock, *supra* note 7, at 229-30.

107. Both Posner, *supra* note 7, at 1195, and Shavell, *supra* note 7, at 1242, implicitly recognize that making the defendant whole is the desirable response to crime because it promotes deterrence. Neither, however, notes or analyzes the contrast between awards of this type and awards that attempt to make plaintiffs whole.

108. Cooter, *supra* note 3, at 96-97, and S. SHAVELL, *supra* note 16, at 194-97, discuss eager victims, but only in the context of insurance and the problem of moral hazard. Yet the eager victim problem arises whenever expected extraordinary awards would leave plaintiffs better off than bargaining would. Awards that exceed contract outcomes subsidize behavior that actually increases the probability of crime.

Eager victims may explicitly use resources to increase the probability of losing property—for example, by strolling through a crime-prone neighborhood even though that route is not as enjoyable as safer routes. Alternatively, they may implicitly misuse resources by foregoing activities that would efficiently deter a taking—for example, by failing to park their cars in a garage.

Both the explicit and implicit misuse of resources by eager victims are forms of contributory negligence, of course. If that doctrine predictably barred or reduced recovery, the problem could be

Courts sometimes require punitive and ordinary damages to be related, as discussed above. That requirement may be a crude way of accounting for π 's moral hazard incentives: it reduces eager victim problems. In fact, courts that require the two awards to be reasonably related nearly always reduce extraordinary jury awards they find too large, rather than increase awards they consider too small.¹⁰⁹ When eager victim potential exists, π 's characteristics are relevant, even though the court's primary focus will remain on taking away Δ 's gains.¹¹⁰

In a system that makes Δ whole, however, π 's advantage from being an eager victim exists only within a bounded range of potential losses. Beyond E_v in Figure 4, π loses any incentive to "game the system" (encourage theft), because the remedy implied by the model—returning

Figure 4

controlled directly. See Brown, *supra* note 21, at 340 (arguing that optimum social benefit is achieved by barring recovery in event of contributory negligence). The illustrations in this note, however, indicate the difficulty in conclusively identifying many eager victim situations. And although his contributorily negligent "eager victimitis" makes π an undeserving plaintiff, it would not reduce the optimal sanction against Δ . This raises a question, to be explored in *infra* notes 137-39 and accompanying text, of the portion of any sanction which should be awarded to π versus the portion to be imposed as a fine or destructive sanction.

109. D. DOBBS, *supra* note 11, at 210-11 & n.42 (noting that the relationship rule helps courts to control "potentially excessive punitive awards").

110. See *supra* notes 70-87 and accompanying text.

Δ to u_0 —actually reduces π 's utility below the level (i_1) that he could have achieved through bargaining.¹¹¹ For takings that exceed Ee' , an award that makes Δ whole yields a subsidiary award that does not even restore the utility of π 's initial endowment (i_0). The extraordinary sanctions model predicts, therefore, that courts would rarely apply a "reasonable relationship" test to jury awards for very large plaintiff losses. An expectation that the award will make Δ whole is unlikely to induce potential plaintiffs to behave strategically where the loss is likely to be great.¹¹²

Eager victims also appear in liquidated damage cases. As discussed above, a legal system that employs punitive damages in tort to safeguard property should likewise enforce liquidated damage clauses in contracts, even when the damages stipulated exceed expected losses *ex ante*. One problem not considered at that point was the possibility that the stipulated damages will create an eager victim (promisee), who will seek to induce breach opportunistically if it becomes possible to realize an exceptionally attractive damage award as a result.

In itself, this version of the eager victim problem is no reason to bar enforcement of the liquidated damage clause. The risk of creating an eager promisee is foreseeable to the promisor; the parties account for that risk in setting the damage figure, that is, in dividing the gains from breach. The moral hazard (risk of creating an eager promisee), together with the payment for it, have been voluntarily accepted by the promisor in agreeing to the terms of the contract. The transaction creating the contract is therefore efficient.¹¹³

After a contract has been concluded, however, a different sort of eager victim problem materializes if, by chance, the actual losses anticipated in the event of breach become less than the damages stipulated in

111. This assumes that π and Δ would have bargained to C . In an illiquid market, the precise contractual outcome may be impossible to determine. Nonetheless, from an expositional perspective, it makes no difference where on the contract curve C is located.

The eager victim range may exceed that implied by the text. Stigler, *supra* note 7, at 527-28, argues that appropriate marginal deterrence—discouraging a defendant from increasing the severity of a plaintiff's injury, for example, by murdering a kidnap victim—requires a lower bound on the rate at which sanctions increase as violation severity increases. If for some violation(s) defendants are made whole—the smallest sanction that will be efficacious for a property violation—more serious offenses may require that defendants be made less than whole in order to maintain marginal deterrence. But if adequate marginal deterrence implies a sanction exceeding one that makes a defendant whole, awarding it to the plaintiff could create an eager victim for takings that exceed Ev .

112. For example, it seems unlikely that anyone will be enticed into a building protected by spring-guns merely because the building's owner will have to pay an award equal to the owner's benefit from placing the weapons there. The benefit to the owner is modest compared to the expected injury to one who is shot by one of the guns. See *Katko v. Briney*, 183 N.W.2d 657 (Iowa 1971) (letting stand on procedural grounds a punitive damage award in such a case).

113. See Shavell, *supra* note 6, at 120 (discussing the moral hazard problem as it applies to insurance contracts).

the contract. The unanticipated (and, hence, uncompensated) reduction in the potential loss makes the promisee prefer breach to performance, creating a danger that the promisee will seek to induce nonperformance by the promisor.¹¹⁴ There is less reason to respect the liquidated damage clause in such situations of ex post opportunism than there is when the clause merely reflects the parties' ex ante agreed division of the gains they may realize if a higher valuing purchaser is identified through luck or effort.¹¹⁵

B. Reluctant Victims

Figure 4 showed that the basic extraordinary sanctions model that was applied initially to punitive tort damages is easily generalized to cover theft and liquidated damages. In both situations, extraordinary awards will be appropriate in order to make defendants whole. All extraordinary sanctions considered so far have been superordinary, but the model is general enough to explain subsidiary awards, that is, awards that systematically undercompensate plaintiffs. As before, the key is the desirability of making defendants whole in certain situations.

Assume again that Δ has chosen to take rather than negotiate and that Δ does not intend or expect to compensate π . For takings greater than E_v in Figure 5, π cannot be given the utility formerly available through exchange (i_1) by an award that also makes Δ whole. Indeed, below point e' π cannot even be restored to the original level of utility (i_0) by such an award; below point g π cannot be restored to i_0 by any award Δ can make, because that would require allocations outside the box, that is, in excess of Δ 's wealth.¹¹⁶

Thus, for larger takings—that is, those exceeding E_v in Figure 5—remedies that restore defendants to their prior level of utility would not give full property protection to plaintiffs. Although such a remedy

114. Clarkson, Miller & Muris, *supra* note 8, at 368-72.

115. Penalty clauses in contract stipulations have a mirror-image in hostile takeovers. Through the stipulations in a contract, the promisee has purchased property protection from the promisor, who then may try to convert it into liability protection by selling to a higher-valuing third party (that is, by breaching). In the takeover context, present shareholders have not relinquished property protection, at least no more than is implied by the corporation's charter and by-laws. But in at least some instances an acquiror may try to convert some shareholders' protections into liability protections (appraisal rights following a "freeze-out") in order to sell to a higher-valuing third party. See generally Haddock, Macey & McChesney, *supra* note 50.

116. Below the contract curve, a taking moves into the range where the marginal value to Δ of the taken goods, although positive, is less than its value to π . Becker, *supra* note 7, at 173, recognizes the contrast between an entitlement's value to a defendant and its value to a plaintiff, but his model does not draw any legal distinction between takings in situations where liability rules are inevitable (such as common accident situations), and takings in situations where bilateral bargaining may reasonably take place. Stigler, *supra* note 7, at 527, argues that the law is generally unconcerned about the value to a criminal from a taking. Stigler implicitly raised, but did not investigate, the property rights distinctions which are central here.

Figure 5

would optimally deter potential defendants from such takings, it would force potential plaintiffs to go outside the legal system to obtain full protection for their entitlements.

It was noted above, however, that many plaintiff actions designed to avoid losses from takings require wasteful use of real resources, and so are to be discouraged. Hence awarding punitive rather than merely compensatory damages is a desirable response to many property violations under the assumptions of the extraordinary sanctions model.¹¹⁷ This observation suggests that a legal system predicated on the desirability of full property protection should switch from compensating defendants to compensating plaintiffs in the event of very serious property violations (where Δ 's taking exceeds E_v). Otherwise, π would be forced to rely on inefficient means of self-help.

In other words, in the face of a property violation, why would the legal system not simply opt for the larger of two alternative awards: the one that makes the defendant whole (with appropriate adjustments for eager victim possibilities), or the one that makes the plaintiff whole? The former award would strip away the defendant's gains from a smaller tak-

117. See *supra* notes 38-50 and accompanying text.

ing (one less than E_v in Figure 5), while affording the plaintiff at least some degree of property protection. The latter award would at least return the plaintiff to his pretaking level of utility in the event of a larger taking (one greater than E_v in Figure 5), while stripping away more from the defendant than he had expected to realize from the taking. As will be shown, however, this is not the most efficient solution.

1. *Insurance: A Peculiar Form of Self-Protection*

The supposition that efficient law would discourage all plaintiff self-protective investments ignores one peculiar form of self protection—insurance. In a modern economy, insurance can be purchased against many takings of physical property (for example, theft or vandalism) and of human capital (for example, wrongful death, murder, or personal injury). Unlike the forms of plaintiff self-protection discussed in Section I.C.1, insurance does not use up real resources, aside from the typically modest costs of administering the plan.¹¹⁸ For the most part, insurance merely effectuates nominal monetary transfers among those who consent *ex ante* to pool the risk of losses. That is, insurance transfers real resources *ex post* to the few participants who have suffered losses from the many who have not, but it does not consume resources in an effort to avoid the losses in the first place.¹¹⁹

Admittedly, although a large percentage of insurance premiums is used to cover mere transfers among the insured, administration of collections and transfers does consume real resources. But so does use of the legal system. In addition, since an insurance firm specializes in assessing and properly pooling risks, it can often perform those functions at a lower resource cost than could the non-specialized legal system.

In the extraordinary sanctions model, therefore, the rule of protecting property by making defendants whole would often persist, even when that rule did not fully compensate plaintiffs. The rule would be appropri-

---

118. Costs of administration ordinarily comprise a relatively modest proportion of the insurance premium for those forms of insurance that survive in the market. INSURANCE INFORMATION INST., INSURANCE FACTS: 1988-89 PROPERTY/CASUALTY FACT BOOK 21 (1988).

119. Insurance companies *do* expend real resources to avoid some sorts of losses, and they may require the insured to expend resources for the same purpose. These expenditures, however, are typically intended to prevent the insured from negligently causing or contributing to compensable injuries that would properly be judged under a liability rule. Accordingly, they are efficient expenditures under the standard liability model as elucidated by Brown, *supra* note 21. Alternatively, the expenditures may be intended to prevent the insured from committing a property violation. All of these expenditures will be appropriate whether or not the legal system functions ideally.

Real expenditures directed toward avoiding becoming the victim of a property violation, however, arise only because the legal system fails to deter costlessly all property violations. In other words, they materialize because an optimal but costly legal system will (unfortunately) not be a perfect legal system.

ate if plaintiffs were apt to purchase insurance to cover the additional amounts needed to afford themselves full property protection.¹²⁰ When available marketed insurance offers the lowest cost protection against property losses, the legal system can appropriately dispense with careful but costly investigations of precise plaintiff losses. Instead, the legal system could focus on a different set of indifference curves, those defining the gains to the defendant realized from the taking. Since neither insurance markets nor their potential substitute, the judicial system, function costlessly, plaintiffs would have to shoulder some of the cost attributable to potential takings by defendants. Overall, however, a rule making defendants whole, coupled with resort to insurance markets, would often give full property protection at the lowest total resource cost.¹²¹

By pointing to the combination of the legal remedies that give proper incentives to potential defendants plus the insurance markets that give proper incentives to potential plaintiffs, the extraordinary sanctions model can now explain several previously controversial legal rules. As

---

120. Note, however, that even if an insurance market for a particular sort of loss was feasible, plaintiffs might be unwilling to participate in it. Plaintiffs would utilize that particular insurance only if it were less costly to them than alternative resource-consuming forms of self-protection were.

121. This discussion should *not* be taken to imply that the entire body of rules permitting civil awards for takings could appropriately be abolished and replaced with private insurance. Optimal deterrence of property violations still requires that defendants be stripped of all gains from such violations. Hence, courts would necessarily be required to evaluate, as best they could, the defendant's indifference curves. From society's standpoint, paying over to the plaintiff the amount collected from the defendant would appear to be a zero-cost proposition, setting aside eager victim situations. Yet in those instances in which private insurance can pool risks over potential plaintiffs more cheaply than the legal system can (and in which insurance is less costly than resource-using forms of self-protection), the legal system should not accept the additional task of evaluating the indifference curves of plaintiffs. An implication of the model, then, is that courts would be more likely to take on the role of quasi-insurer in thick markets than in thin ones, because their costs of ascertaining compensation that would make the plaintiff whole will be reduced when courts can focus on market prices rather than on indifference curves.

Even so, the desirability of making a defendant whole would not seem to require that the sanction extracted from him be given to the plaintiff. Why not award it to the state instead? As will be discussed at *infra* notes 137-39 and accompanying text, that is sometimes an appropriate policy. In many instances, however, it is efficient to award substantial sums to plaintiffs, even if the awards fall grossly short of making them whole, because of the incentives the plaintiffs are given to prosecute cases against culpable defendants. In other words, with the possibility of at least *some* award, plaintiffs will assist courts in making defendants whole.

As the victim, a plaintiff often learns, automatically and unintentionally, a great deal about a taking, whereas it would be costly, if not impossible, for a public prosecutor to acquire similar information. For example, the plaintiff knows that he has suffered a broken jaw as a result of a punch thrown by the defendant. A state prosecutor, on the other hand, will know the same thing only if someone, such as the plaintiff, tells him, or if the prosecutor incurs substantial costs ferreting out information that has not been volunteered. Since subtle injuries will be difficult to prove, actions charging some types of property violation would rarely be brought unless a victim could expect sufficient compensation, as through an award from a defendant, to make the effort worthwhile for him as a plaintiff. Because the interests of society at large dictate that such cases be energetically prosecuted, it follows that under many circumstances, plaintiffs must expect to receive awards, even ones that are insufficient to make them whole.

shown below, the same rationale that has explained overcompensatory awards also governs awards for wrongful death, awards that systematically undercompensate plaintiffs.¹²² First, however, the model will be used to formulate a motivation for the collateral source rule. This controversial rule can now be shown to be very closely related to rules that result in undercompensatory awards to plaintiffs.

2. *The Collateral Source Rule*

Below v in Figure 5 the legal system fails to restore π 's entire economic losses; thus, as the level of taking increases, π would be expected to resort increasingly to various forms of self-help to supplement undercompensatory court-awarded damages. Like any legal remedy, all forms of self-help are costly. An efficient legal system would attempt to minimize those costs.

One common form of self-help is insurance. For a theft of Ee' , for example, a rule making Δ whole will undercompensate π , who would get only $e'e$. But if π can fully insure against losses, it will be possible for him to arrange for the additional ee'' necessary to restore the utility that would have resulted from bargaining (i_1).¹²³

Recovery by injured plaintiffs from two different sources has troubled courts and scholars. The collateral source rule allows plaintiffs to keep both a full recovery from defendants and also collateral benefits—insurance, gifts, salary, etc. The rule has been controversial.¹²⁴ As Dobbs notes, "legal arguments in favor of the collateral source rule are often unconvincing." Of these, "[p]erhaps the weakest . . . is the one that has been most mentioned in the courts—the wrongdoing defendant should not get the benefit of any reduction in the plaintiff's damages by a collateral source, since this would be a 'windfall.'"¹²⁵

As Figure 5 shows, however, that "weakest" argument actually has considerable merit. Permitting π 's collateral recovery to offset Δ 's obligation would vitiate the benefits of returning Δ to his pretaking position. Moreover, if collateral sources benefited defendants rather than plaintiffs, π would have to seek other ways to cover the potential loss ee'' .¹²⁶ π

122. See *infra* notes 128-36 and accompanying text.

123. It is likely that in many situations plaintiffs will only partially insure against potential losses, even in the absence of moral hazard problems for the insurer. See Friedman, *supra* note 10, at 83-85. That complication, however, does not affect the present analysis and is ignored here for simplicity.

124. See D. Ellis, *supra* note 5, at 2 (footnotes omitted) ("Most legal academics have been critical of the [collateral source rule] . . . . The courts have been more hospitable to the rule and have continued to expand its scope."); Fleming, *supra* note 5, at 1484 (criticizing the rule); cf. Maxwell, *supra* note 5, at 695 (concluding that the rule, while of doubtful benefit in an ideal legal system, is nonetheless necessary).

125. D. DOBBS, *supra* note 11, at 584, 586.

126. A few courts treat gifts from friends and relatives differently from contracted-for benefits,

seeks least-cost entitlement protection, and prohibiting the use of an efficient alternative can only raise costs. Although an eager victim risk may exist, insurance companies specialize in coping with that risk.¹²⁷ Resort to the judicial process is undesirable when more specialized institutions perform adequately. The collateral source rule incorporates the gains from such specialization.

The proper objective of a remedy for property violations is to strip all gains from defendants. Whether the plaintiff has insurance is irrelevant to this objective. Since the law seeks to make defendants whole in appropriate circumstances, it does not and should not offset awards against defendants with payments from collateral sources: otherwise, plaintiffs' property protections would be incomplete in some instances. This same reasoning applies equally well to other subsidiary awards.

3. *Wrongful Death Remedies*

Many commentators have noted that wrongful death awards are systematically subsidiary.¹²⁸ They typically fail to meet even a lower-bound estimate of the decedent's loss, a failure that has not been adequately explained.¹²⁹ The extraordinary sanctions model predicts that an efficient legal system will often fail to compensate the decedent in wrongful death cases, just as it does in other situations where it strives to enforce property rules. Instead, if the death was truly wrongful in ways that violated property protections, the law should strive to "compensate" the defendant in the economist's sense of stripping away all gains realized by the defendant.¹³⁰

As before, however, the general availability of insurance means that judgments that merely make defendants whole for large takings need not leave plaintiffs undercompensated overall. As discussed above, private

---

possibly because they are uncertain and therefore do not affect a plaintiff's search for substitutes. See *id.* at 582-83. Yet even if gifts are unexpected windfalls (a troublesome assumption in its own right), that only suggests that a plaintiff should return a gift to its source to the extent that it duplicates a recovery from the defendant. A rule requiring that gifts reduce a defendant's obligation diminishes property protection, because the amount a defendant must pay, not the amount a plaintiff will receive, affects a defendant's incentive to take plaintiff's property.

Under "subrogation" clauses in insurance contracts, some collateral benefits are indeed returned to the source. These clauses typically entitle the insurer rather than the injured party to sue the defendant, with the injured party collecting the contracted-for benefits from the insurance company regardless of the outcome of the suit. S. SHAVELL, *supra* note 16, at 143. Such clauses both reduce insurance premiums and increase plaintiff's certainty while maintaining the disincentive for potential defendants to neglect property rules.

127. Shavell, *supra* note 6, at 121-22.

128. See, e.g., W. LANDES & R. POSNER, *supra* note 10, at 186-89; Cohen, *supra* note 10, at 339-40; Komesar, *supra* note 10, at 475-76.

129. See W. LANDES & R. POSNER, *supra* note 10, at 314-15.

130. Note again that because the goal is deterrence, it is the ex ante expected gains of the defendant that are of interest, not the ex post realized gains.

arrangements (collateral sources) can compensate plaintiffs fully in cases not involving wrongful death. Likewise, an efficient legal system could leave potential plaintiffs or decedents to arrange their own compensation via insurance for wrongful death. This does not mean, of course, that remedies which make defendants rather than plaintiffs whole properly sanction all wrongful deaths. For some deaths (such as in traffic accidents), liability rule protections will still be appropriate because transaction costs make negotiation *ex ante* prohibitively costly. In such a case, where efficient negotiations can never take place, there is no reason to try to encourage defendants to negotiate by applying extraordinary sanctions. But many deaths occur in contexts where negotiation *ex ante* is feasible or even routine.

Wages for hazardous occupations, for example, include a risk premium negotiated in advance.¹³¹ Even so, an employer may subsequently attempt to "take" the employees' negotiation rights by increasing the risks facing the employees without bargaining with them over the riskier work conditions, and perhaps without their knowledge. In the most egregious situations, the employer may actually interfere with the information flow that otherwise would warn employees of the increased risk. In those circumstances, the wrongful death would appropriately be treated by courts as a property-rule violation, since bargaining was clearly feasible. The intent of the court would then be to impose a remedy on the defendant-employer sufficient to nullify *ex ante* any expected benefits from a repetition of such behavior by that employer or others similarly placed. Since the court focuses on the gain to the defendant, however, the remedy may leave plaintiffs undercompensated. But in situations in which insurance firms can provide coverage more cheaply than a court can, that is the efficient approach. Employees who are apt to be undercompensated in such instances are properly left to purchase their own level of additional compensation via insurance.

There are two exceptions to this general rule. First, making defendants whole by stripping them of gains may create an "eager estate" problem analogous to the eager victim problem discussed above,¹³² in which potential heirs have an incentive to invest resources to increase the probability of an apparently wrongful death.¹³³ The efficient response to

---

131. Deaths occurring in this context should properly be called "unfortunate deaths" rather than "wrongful deaths." "Unfortunate death" recoveries might well appear to be "subordinary" as compared to those for decedents who die in similar, but noncontractual situations. In fact, however, they are not subordinary, because compensation had been paid "up front" as a premium in the contract price for the decedent's complete or partial assumption of the risk.

132. See *supra* notes 108-15 and accompanying text.

133. An eager estate problem is more likely when a potential decedent has insurance coverage or assets in addition to human capital, because insurance may motivate the named beneficiaries to hasten insured's death (by murdering the insured or withholding care or attention from him, for

such a problem is to make defendants whole, but to give part of the award to the state, as criminal sanctions, rather than to the plaintiff.¹³⁴ Second, certain potential plaintiffs (children, for example) may not be able to contract for additional compensation via insurance. In that event, entitlements will not be protected by making defendants whole. To afford full property protection, the law must award the estate compensation that makes whole those third parties who could not contract for their own protection. Those third parties who can so contract should arrange for their own compensation (if any) through private collateral channels such as life insurance.

Admittedly, judicial opinions do little to elucidate the rationale for subsidiary recoveries. In most sorts of cases awards are compensatory (strict liability cases) or superordinary (more modest property violations). But for substantial property violations, awards will be systematically subsidiary. Maintaining property rules requires only that courts make defendants whole following violations. Plaintiffs' private access to collateral contracts frees courts from investigating all the peculiarities of every catastrophic loss. Since it need not concern itself with restoring plaintiffs' losses, the legal system can adequately punish property violations after considering only one indifference curve.¹³⁵

Through proper treatment of collateral sources, a legal system provides both deterrence against property violations through litigation and least-cost avoidance of plaintiffs' losses through private markets. If courts were less costly, or collateral contracts were more costly, the system could dispense with private contracts.¹³⁶ But modern insurance

---

example). Importantly, however, insurance coverage and asset ownership are under the control of those at risk, who presumably are better able to gauge the potential for an eager estate problem than is a court, and who can take whatever precautions are appropriate to control it. For example, the coverage of most insurance policies is limited by the insured to amounts sufficient to "look after" the dependents (or less), so an eager estate problem does not materialize with those contracts. In other words, potential decedents voluntarily purchase life insurance sufficient to cover not their own losses, but only the losses of their dependents. The wrongful death rules therefore mimic the insurance market.

134. See *infra* notes 137-39 and accompanying text.

135. For a well-reasoned argument that the law could come much closer to awards that would make the plaintiff population whole in such cases, if that were its goal, see Cohen, *supra* note 10, at 313-24. Cohen notes that many observers erroneously insist that the law cannot approach full compensation for the population of plaintiffs in wrongful death cases because the observers refuse to focus on *ex ante* rather than *ex post* concepts and because they also fail to appreciate the full implications of the probabilistic nature of most events that result in death. *Id.* at 295-96.

136. When the "load" of private insurance is higher, the legal system may take a more active role in compensating plaintiffs for losses. For example, W. LANDES & R. POSNER, *supra* note 10, at 186, note that "[p]rimitive and early societies . . . realized that a fatal injury inflicts a real loss for which compensation ought to be paid. But at some point in the evolution of the English common law the rule was changed . . . ." In contrast to the extraordinary sanctions model, Landes and Posner are critical of that change. *Id.* at 187-89.

Markets and courts are not the only alternatives to protect human capital. For example, the

agreements, employment contracts, and the like provide more efficient coverage than courts can. Courts are specialized to administer the law, not insurance. Moreover, since the insured pays, directly or indirectly, for private coverage, there is no *ex ante* incentive to misrepresent one's utility map, since contracts must be concluded and paid for. By contrast, misrepresentation is a real problem after an injury, when the damage is done and only compensation is at issue. Full court-ordered recovery for their losses would enable plaintiffs to bypass insurance contracts even though those contracts are preferable to litigation.

C. *The Role of Criminal Punishment*

The discussion of property violations for which defendants expect to escape even ordinary compensatory damages has focused on private plaintiffs. Yet many takings are punishable with both criminal and civil liability; wrongful death (murder or manslaughter) is an example. Similarly, common theft can result in actions for both criminal larceny and civil conversion; payment of civil damages by a defendant may leave unaltered the criminal sanctions the defendant must bear. Why the double system?

Figure 5 indicates two roles for criminal law. First, efficiency would sometimes require nonmonetary punishment for Δ because adequate monetary penalties are infeasible. For very large takings, after which Δ has destroyed or consumed the asset taken, the law cannot make Δ whole through a mere transfer, since that would require a sanction that exceeds Δ 's alienable wealth. In Figure 5, for a taking greater than Ej , a sanction greater than jj' is required to make Δ whole. But anything more than jj' exceeds Δ 's wealth. In that instance, Δ is "judgment-proof." To reduce Δ 's utility to u_0 then requires imposing destructive sanctions—imprisonment or corporal or capital punishment—through which the system expends its own resources to reduce those available to Δ .¹³⁷

"blood feud" enables survivors to obtain restitution directly from a "defendant" or his "insurers"—usually his relatives—when neither a formal insurance market nor the law can reasonably be expected to play any appreciable role. See Miller, *Choosing the Avenger*, 1 L. & HIST. REV. 159 (1983) (describing the Icelandic blood feud of the tenth to thirteenth centuries); Friedman, *Private Creation and Enforcement of Law: A Historical Case*, 8 J. LEGAL STUD. 399 (1979) (general discussion of Icelandic legal system during period of blood feud). Of course, as the value of human capital rises, the emergence of more specialized institutions to protect it is predictable and desirable.

137. This "destructive" element of criminal law becomes increasingly important as the probability of detection decreases, meaning that the sanction actually imposed must increase proportionally in order to keep constant the expected value of the sanction. See S. SHAVELL, *supra* note 16, at 284; Posner, *supra* note 7, at 1203; Shavell, *supra* note 7, at 1245.

After a similar explanation of nonmonetary criminal sanctions, Shavell, *id.* at 1242, notes that the utility gained by committing a particular criminal act may sometimes exceed the maximum possible sanction for that act. Hence, that defendant cannot be made whole, and so the crime cannot be deterred. Here, Shavell is correct; however, he then remarks that assuming perfect information by the court, "[u]nder the optimal system, . . . no sanction will be imposed on a party who

Second, using criminal law to punish defendants controls eager victim and eager estate incentives in the range immediately above v . If Δ is not judgment-proof, transfers ordinarily are preferable to economically inefficient criminal sanctions. But determinations of π 's actual losses and strategic behavior are costly, so some offenses would permit a civil suit only for provable compensatory damages, with the remainder of the appropriate sanction diverted elsewhere—perhaps as state-imposed monetary or destructive sanctions, which are criminal rather than civil.¹³⁸

Hence, "corporate crime" statutes, which place corporations themselves at risk of criminal prosecution, are also explained by the extraordinary sanctions model. If an agent commits a property violation that is foreseeable but not controlled by the corporation, the corporation (that is, the shareholders) should be returned to the level of profit it

---

committed an undesirable act if he could not possibly have been deterred . . . [I]t would serve no purpose and yet be socially costly to impose a sanction." *Id.*

On the latter point, Shavell is mistaken for three reasons. First, as Shavell himself now acknowledges, destructive sanctions (unlike monetary sanctions) serve legal functions in addition to deterrence, particularly incapacitation. See Shavell, *A Model of Optimal Incapacitation*, 77 AM. ECON. REV. 107, 107 (Papers and Proceedings of the American Economic Association, May 1987). Even assuming that a defendant could not have been deterred from an accomplished act, if that defendant is likely to be similarly undeterrable in the future, imprisonment or execution will prevent the defendant from committing other criminal acts for awhile or forever.

Second, Shavell ignores what the criminal law calls "general" deterrence. Unless other potential culprits also are assumed to have perfect information, they could not distinguish those previous culprits who were undeterrable—and so had gone unpunished—from those who could (and should) have been deterred—and so had consequently received punishment. Accordingly, the sorts of discriminatory penalties that Shavell envisions would make criminal sanctions seem to be random, unsystematic retribution. Potential criminals—even those who were deterrable—would thus underestimate the probability of being punished for legal violations.

Finally, refusing to punish undeterrable individuals (apparently because they are in some sense irrational) creates a perverse incentive for future defendants to act as if they are irrational. It thus becomes "rational" in Shavell's world to act irrationally and thereby escape criminal punishment. As more defendants escape punishment in this manner, the system of criminal justice is overwhelmed and the Shavellian legal system breaks down altogether.

The Shavellian strategy could work only as long as undeterrable individuals comprise a small proportion of the total population. But resort to the strategy increases the proportion of users. Hence, in the vocabulary of sociobiology, it is not an evolutionarily stable strategy. For a discussion of evolutionarily stable strategies, see R. DAWKINS, *THE SELFISH GENE* 69-86 (2d ed. 1989).

138. For some infractions, awards that seem merely compensatory would transform a few plaintiffs into eager victims. In those instances, private suits should be barred altogether. Some award to the plaintiff will often be desirable, however, in order to prevent the plaintiff from overexpending resources to reduce the probability of the infraction. On the latter point, see W. LANDES & R. POSNER, *supra* note 10, at 157.

Because a diversion of sanctions to the state controls eager victim problems, there is no reason for the law to strive to limit criminal penalties to those that make the defendant whole, given the assumptions here. When the assumption that courts never erroneously impose liability is relaxed, however, an upper bound on the magnitude of criminal sanctions will be appropriate. Such errors deter some socially desirable acts that may be mistakenly branded criminal, and the greater the sanction, the greater the number of desirable acts discouraged. Similarly, if the cost to the legal system of imposing sanctions increases with the magnitude of the sanction, the optimal sanction will be limited. See Shavell, *supra* note 7, at 1243.

would have earned in the absence of the agent's actions.¹³⁹ Expectation of such a penalty promotes internal corporate controls that encourage agents to respect plaintiffs' property rights. If the required punitive sanctions would lead to eager victim behavior when converted into plaintiffs' awards, part or all of the sanction should be imposed as a criminal fine.

Declaring an inanimate object—the corporation—a “criminal” is semantically inelegant; it is, of course, shareholders, not the legal fiction we call “the corporation,” who actually bear the sanction. Terminology aside, however, the logic of corporate (that is, shareholder) crime statutes is sound. Property rules require sanctions that leave defendants with no advantage from a violation; however, avoiding an eager victim problem requires an expectation that plaintiffs will be left with no advantage over a market outcome. Those simultaneous demands on the legal system are sometimes mutually inconsistent unless some third party, such as the state, claims part or all of the award.

CONCLUSION

A remedy that “makes the plaintiff whole” can create incentives for a defendant to bypass negotiation in favor of an outright taking. Takings discourage investment in takeable assets and create deadweight losses, since defendants spend resources to find such takeable assets and plaintiffs spend resources to protect them. Consequently, an efficient legal system often will opt for a remedy that makes a defendant rather than a plaintiff whole when the defendant intentionally takes, rather than negotiates for, a property-protected entitlement. Once that is properly understood, many seemingly inexplicable and disjointed legal rules become sensible and unified, issuing from a single dynamic model of extraordinary sanctions.

In the extraordinary sanctions model, many seemingly different aspects of the law of remedies are seen to be logically related. For example, injunctions are merely one form of extraordinary sanction, and the law's refusal to award additional punitive damages when injunctive relief is granted becomes predictable. The model has been used here to explain still other aspects of tort and criminal law, the boundary between them, and aspects of contract law. Several puzzling matters that concern extreme plaintiff losses also have been clarified. For substantial takings, awards that would make plaintiffs whole exceed those that would make defendants whole, and the divergence grows with the size of the taking. Thus, if the law makes defendants whole following property violations, it must accord a role to private insurance for protecting plaintiffs from very

---

139. Indeed, there is no reason to distinguish between corporations and individuals in any of the taking situations discussed above. Whenever the litigant is a corporation, isoprofit curves should be substituted for isoutility (indifference) curves in Figures 1 through 5.

large losses. This insight explains controversial and seemingly unrelated legal phenomena, such as the collateral source rule and systematic undercompensation of plaintiffs in serious personal injury or wrongful death cases. The model also indicates the circumstances under which third parties like the state should prosecute and punish defendants, or divert monetary sanctions from victims.

Liability rules are sometimes appropriate, particularly when a market price exists that can readily set compensation levels (as when markets are highly liquid), or when the cost of enforcing property rules would be prohibitive (as with ordinary traffic accidents). Circumspect use of extraordinary sanctions instills respect for property protections, so most litigated cases are, in fact, governed by a liability rule. But this fact means only that the observed sample of litigation is biased toward liability cases, not that an undeviating goal of good law is to make plaintiffs whole.

The argument here is neither that judges are omniscient nor that the law is unswervingly efficient. The model merely refutes familiar arguments that noncorrective damages that systematically and intentionally over- or undercompensate plaintiffs must be inefficient. Such damages need not be inefficient; in the situations discussed here, they are not. Nor does mastery of the various aspects of law considered here demand mastery of a plethora of models. To paraphrase William of Occam, plurality need not be assumed in this instance. With greater intellectual efficiency, what had been done in vain with more now can be done with less.