



1985

An Economic Theory of the Criminal Law

Richard A. Posner

Follow this and additional works at: http://chicagounbound.uchicago.edu/journal_articles

 Part of the [Law Commons](#)

Recommended Citation

Richard A. Posner, "An Economic Theory of the Criminal Law," 85 Columbia Law Review 1193 (1985).

This Article is brought to you for free and open access by the Faculty Scholarship at Chicago Unbound. It has been accepted for inclusion in Journal Articles by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

AN ECONOMIC THEORY OF THE CRIMINAL LAW

*Richard A. Posner**

INTRODUCTION

The economic analysis of criminal law began on a very high plane in the eighteenth and early nineteenth centuries with the work of Beccaria and Bentham,¹ but its revival in modern times dates only from 1968, when Gary Becker's article on the economics of crime and punishment appeared.² Since then there has been an outpouring of economic work on criminal law, concentrated in the following areas: the optimal tradeoff between certainty and severity of punishment, the comparative economic properties of fines and imprisonment, the economics of law enforcement and criminal procedure, and above all the deterrent and preventive effects of criminal punishment (including capital punishment).³ Notice, however, what is missing from this list: the substantive doctrines and concepts of criminal law, about which there has been little economic writing.⁴ This is in striking contrast to the

* Judge, U.S. Court of Appeals for the Seventh Circuit; Senior Lecturer, University of Chicago Law School. I am grateful to Gary S. Becker, Isaac Ehrlich, George P. Fletcher, William M. Landes, Geoffrey P. Miller, Norval Morris, A. Mitchell Polinsky, Steven Shavell, Steven J. Shulhofer, A.W.B. Simpson, George J. Stigler, and participants in the Law and Economics Workshop and the Seminar on Rational Choice Methods in Social Sciences (both at the University of Chicago), for helpful comments on a previous draft of this Article, and to John H. Langbein for discussion of the subject matter of the Article. The Article was stimulated in part by critical remarks on the economic approach to criminal law made by George Fletcher at a conference held on November 16, 1984, on New Directions in Law and Economics, sponsored by Columbia Law School's Center for Law and Economic Studies.

1. See C. Beccaria, *On Crimes and Punishments* (H. Paolucci trans. 1963); J. Bentham, *An Introduction to the Principles of Morals and Legislation*, in 1 *Works of Jeremy Bentham* 1, 86-91 (J. Bowring ed. 1843); J. Bentham, *Principles of Penal Law*, in 1 *Works of Jeremy Bentham*, *supra*, at 365.

2. See Becker, *Crime and Punishment: An Economic Approach*, 76 *J. Pol. Econ.* 169 (1968).

3. For a comprehensive bibliography of research on the economics of crime and punishment to 1980, see *The Economics of Crime* 411-26 (R. Andreano & J. Siegfried eds. 1980); and for an excellent review of almost the entire literature, see D. Pyle, *The Economics of Crime and Law Enforcement* (1983).

There is also some interesting literature on the economics of organized crime. See, e.g., P. Reuter, *Disorganized Crime: The Economics of the Visible Hand* (1983); T. Schelling, *Choice and Consequence: Perspectives of an Errant Economist* chs. 7-8 (1984).

4. There is a short treatment of substantive issues of criminal law—the germ of this Article—in my economic analysis text. See R. Posner, *Economic Analysis of Law* ch. 7 (2d ed. 1977). Both Becker, *supra* note 2, and Stigler, *The Optimum Enforcement of*

situation with respect to tort law, though tort and criminal law are closely related. Tort notions seem to lend themselves to economic translation and elaboration—the Learned Hand formula for negligence being the most dramatic example⁵—while the concepts that dominate the substantive criminal law, such as attempt, conspiracy, entrapment, insanity, and premeditation, seem alien to the economist's way of thinking about problems. In particular, the pervasive emphasis placed in the criminal law on punishing harmless preparatory activity, on the mental state of the accused, and, related to both points, on the moral character rather than the consequences of behavior, suggests a decidedly noneconomic perspective.

But I think this is wrong, and that the substantive doctrines of the criminal law, as of the common law in general, can be given an economic meaning and can indeed be shown to promote efficiency.⁶ That, at any rate, is the burden of this Article. I certainly do not want to be understood, however, as arguing that every rule of the criminal law is efficient, or that efficiency is or ought to be the only social value consid-

Laws, 78 J. Pol. Econ. 526 (1970), written in response to Becker, also touch on some substantive issues, and in so doing anticipate (by many years!) some of the analysis of substantive criminal law doctrine in this Article. A paper by Steven Shavell, originally written in 1980, contains many parallels to my own analysis. I had not read his paper before completing the first draft of mine, and although I have read it since and revised my own, I have made no changes based on his analysis. I have thus refrained from discussing the principles of causation in criminal law, a topic omitted from my original draft and one on which Shavell's paper contains a fascinating discussion. See Shavell, *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, 85 Colum. L. Rev. 1232 (1985).

Skepticism about the prospects for a positive economic theory of substantive criminal law is the leitmotif of the papers—all but mine, of course—in *Nomos XXVII: Criminal Justice* pt. IV (J. Pennock & J. Chapman eds. 1985). In the same vein, see Seidman, *Soldiers, Martyrs, and Criminals: Utilitarian Theory and the Problem of Crime Control*, 94 Yale L.J. 315 (1984). As I said in my paper in *Nomos*, “[t]he problem with the economic analysis of criminal law is not that it is incomplete or lacks rigorous philosophical foundations, though it is and does, but that the economic analysts have yet to tackle the principal concepts that trouble legal analysts of the field—such concepts as attempt, conspiracy, diminished responsibility, provocation, insanity, strict criminal liability, recklessness, compulsion or necessity, and premeditation.” Posner, *Comment on On the Economic Theory of Crime*, in *Nomos XXVII: Criminal Justice*, at 310, 311 (J. Pennock & J. Chapman eds. 1984). This Article, like Shavell's, *supra*, is an effort to repair the omission.

5. See *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947); see also R. Posner, *Tort Law: Cases and Economic Analysis* ch. 1 (1982) (discussing formula); Landes & Posner, *The Positive Economic Theory of Tort Law*, 15 Ga. L. Rev. 851, 884–85 (1981) (same).

The Hand formula of negligence is $B < PL$, where B is the hurden of precaution, P is the probability of an accident if B is omitted, and L is the magnitude of the loss if the accident occurs. If $B < PL$, then B is cost-justified, and omitting B is negligent.

6. On the “efficiency theory” of the common law, see, e.g., R. Posner, *supra* note 4, pt. II; R. Posner, *supra* note 5; Landes & Posner, *supra* note 5. Although criminal law is no longer a pure common law field, most of its doctrines are of common law rather than statutory origin.

ered by legislatures and courts in creating and interpreting the rules of the criminal law.

My analysis can be summarized in the following propositions:

1. The major function of criminal law in a capitalist society is to prevent people from bypassing the system of voluntary, compensated exchange—the “market,” explicit or implicit—in situations where, because transaction costs are low, the market is a more efficient method of allocating resources than forced exchange. Market bypassing in such situations is inefficient—in the sense in which economists equate efficiency with wealth maximization⁷—no matter how much utility it may confer on the offender.

2. Much of this market bypassing cannot be deterred by tort law—that is, by privately enforced damage suits. The optimal damages that would be required for deterrence would so frequently exceed the offender’s ability to pay that public enforcement and nonmonetary sanctions such as imprisonment⁸ are required.

3. Such sanctions are extremely costly for a variety of reasons, and this, together with the socially worthless character of most of the sanctioned conduct, has a number of implications for efficient criminal law doctrine, such as that unsuccessful attempts should be punished in order to economize on costlier punishments for completed crimes. The threat of punishing attempts, as we shall see, makes the completed crime more costly in an expected sense and therefore less likely to be committed. I contend that the main differences between substantive criminal law and substantive tort law can be derived from the differences in (1) the social costs of criminal and tort sanctions and (2) the social benefits of the underlying conduct regulated by these two bodies of law. I contend, in short, that most of the distinctive doctrines of the criminal law can be explained as if the objective of that law were to promote economic efficiency.

I. THE FUNCTION OF CRIMINAL LAW

A. *An Economic Typology of Crimes*

In this section I try to derive the basic criminal prohibitions from the concept of efficiency; I argue that what is forbidden is a class of inefficient acts. As this is a controversial endeavor I think it important to note that the rest of the Article does not depend on it—that it would be little affected if I took as given that society wants to prevent the acts that it calls murder, theft, rape, etc., and did not inquire why.

When transaction costs are low, the market is, virtually by definition, the most efficient method of allocating resources. Attempts to by-

7. See R. Posner, *The Economics of Justice* 66–107 (1981).

8. Even the fine, we shall see, is not a purely monetary sanction equivalent to tort damages.

pass the market will therefore be discouraged by a legal system bent on promoting efficiency. If I covet my neighbor's car, it is more efficient to force me to negotiate with my neighbor—to pay him his price—than it is to allow me to take his car subject to being required by a court to pay the neighbor whatever the court decides the car is worth. If I happen to have no money but want a car, it would be inefficient to let me just take a car. Indeed, unlike the first case, this transfer cannot possibly improve the allocation of resources—that is, it cannot move resources from a less to a more valuable employment—because value is a function of willingness to pay. Since I am unwilling (because unable—but it does not matter why) to pay my neighbor's price for the car, it follows that the car would be less valuable in an economic sense in my hands than in his.⁹ Moreover, if I am allowed to take the car I will have an incentive to expend resources on taking it and my neighbor will have an incentive to expend resources on preventing it from being taken, and these expenditures considered as a whole, yield no social product.

In short, it is inefficient to allow pure coercive transfers of wealth—"pure" implying that the transfer is not an incident of a productive act. But this is an important qualification. The invention of a new product or process can also cause all sorts of wealth transfers that are involuntary from the standpoint of the losers, but invention increases, as well as transfers, wealth in a way that merely taking someone's wealth from him does not. Invention is not just a coercive or involuntary transfer, and it would be infeasible to force the inventor to identify and negotiate terms of compensation with all the losers.

The role of the criminal law in discouraging market bypassing is obscured by the fact that the market transaction that the criminal bypasses is usually not a transaction with his victim. If someone steals my car, normally it is not because he wants *that* car and would have bought it from me if the criminal law had deterred him from stealing it. He steals to get money to use in buying goods and services from other people. The market transaction that he bypasses is the exchange of his labor for money in a lawful occupation. But it is still market bypassing.

Although the market-bypassing approach provides a straightforward economic rationale for forbidding theft and other acquisitive crimes—such as burglary, robbery, fraud (false pretenses), embezzlement, extortion (by threat of violence), most kidnapping, some murder, some assault and battery, some rape—we must also consider "crimes of passion," which loom large in thinking about the criminal law and which may seem to have nothing to do with bypassing markets. Such crimes can be defined in economic terms as crimes motivated by inter-

9. The car might, of course, confer more utility (pleasure, satisfaction) on me than on my neighbor, but there is a difference between utility in a broad utilitarian sense and value in a (perhaps narrow) economic sense, where value is measured by willingness to pay for what is not yours already, or willingness to accept payment for what is yours. See R. Posner, *supra* note 7, at 66–67; *infra* text following note 10.

dependent negative utilities. An example is murdering someone because you hate him rather than because you want his money. These are not wealth transfers in any obvious sense and may seem to have nothing to do with bypassing the market. It might seem, therefore, that before we could pronounce such conduct inefficient we would have to compare the offender's utility with the victim's disutility.¹⁰ We could not do this without exceeding the conventional limits of economics, which do not allow interpersonal comparisons of utilities, just as we could not describe a theft as efficient because the impecunious thief would derive greater pleasure from his act than the pain suffered by his wealthy victim.

Now as a matter of fact it is a pretty safe empirical guess that most such conduct does create net disutility. The whole idea is to inflict as much disutility on the victim as possible, and it is unlikely that every disutility experienced by the wretched victim confers an equal and opposite utility on the offender. Indeed, there would seem to be a fundamental asymmetry between the pleasure that one would obtain from killing another person who has sullied one's honor, and the victim's pain, broadly defined to include the disutility to him of losing his life. But I want to emphasize four other economic, rather than utilitarian, points:

1. Coercion arising from interdependent negative utilities cannot increase the wealth of society and therefore cannot be an efficient act. If *A* kills *B* because the resulting disutility to *B* confers utility on *A*, the wealth of the society is not increased even in the unlikely event that the total amount of human happiness is increased.

2. The dichotomy between acquisitive crimes and crimes of passion is overstated. Acquisitive crimes bypass explicit markets; crimes of passion often bypass implicit markets—for example, in friendship, love, respect—that are the subject of a growing economic literature illustrated by Becker's work on the family.¹¹ Less obviously, crimes of passion often bypass explicit markets too. The essential characteristic of a market, and the source of the ethical appeal of market systems, is that in a market people have to be compensated for parting with the things that have value to them, unless transaction costs are prohibitive. Someone who gets his satisfactions in life from beating up other people, without compensating them, rather than from engaging in trade with them is thus bypassing explicit markets. This point is obscured by the fact, noted earlier in the context of acquisitive crimes, that the victims of the crimes and the people that the aggressor would be trading with if he were not committing crimes are different people.

To sum up, one who spends his time brawling rather than working

10. This is a common criticism of attempts to apply economics to deliberate wrongdoing. See, e.g., Ellis, *An Economic Theory of Intentional Torts: A Comment*, 3 *Int'l Rev. L. & Econ.* 45 (1983).

11. See G. Becker, *A Treatise on the Family* (1981).

is bypassing an explicit market;¹² if he spent his time raping rather than dating women he would be bypassing an implicit market. The essential point in both cases is that he would not be deriving his satisfactions in life from acts that confer benefits on other people.

3. Allowing coercion would create incentives for potential victims to spend heavily on self-protection and for potential aggressors to spend heavily on overcoming the victims' self-protective efforts. All this spending would yield little if any net social product.

4. Some crimes of passion are costly and inefficient efforts at self-help. *A* slanders *B*, and *B*, instead of suing *A*, kills him. The suit would have given *B* almost the satisfaction that killing *A* did, and at far lower social cost.

It may clarify the analysis to consider two seemingly more problematic examples of the concept of crime as pure coercive transfer:

(1) *Counterfeiting*. This can be viewed as a form of theft by false pretenses, the false pretense being that the "payor" is actually paying. The victim is whoever has the money when it is discovered to be counterfeit. Even if the counterfeiting is never discovered so that no individual or firm suffers a loss, counterfeiting imposes a social cost measured by the resources consumed in counterfeiting and in trying to prevent counterfeiting and also by the social costs of the inflation caused by counterfeiting. An increase in the stock of currency will have an inflationary effect—a very considerable one if counterfeiting is not punished at all.

Actually, the counterfeiter whose counterfeiting is never discovered differs from the thief only in the number of his victims. Both take slices of the social pie without putting in anything in return, but the victims of the undiscovered counterfeiter are all those who pay higher prices as a result of the increase in the amount of currency in circulation.

All of this assumes that even if counterfeiting is not a crime, counterfeit money is not legal tender. If it is, there can be no individual victims of counterfeiting, but incentives to work and save will be totally undermined. Anyone who wants anything will simply print up some money and "buy" the thing he wants.

(2) *Rape*. Suppose a rapist derives extra pleasure from the coercive character of his act. Then there would be (it might seem) no market substitute for rape, suggesting that rape is not a pure coercive transfer and should not, on economic grounds, anyway, be punished criminally. But the argument would be weak:

(a) Because there are heavy penalties for rape, the rapes that take place—that have not been deterred—may indeed be weighted toward a form of rape for which there are no consensual substitutes; it does not

12. Professional boxing would be an example of a lawful market alternative to battery.

follow that the rape that is deterred is generally of this character.¹³

(b) Put differently, the prohibition against rape is to the marriage and sex "market" as the prohibition against theft is to explicit markets in goods and services.¹⁴

(c) Given the economist's definition of "value," even if the rapist cannot find a consensual substitute (and one such substitute, prostitution, is itself illegal), it does not follow that he values the rape more than the victim disvalues it. There is a difference between a coerced transaction that has no consensual substitute and one necessary to overcome the costs of consensual transactions; only the second can create wealth, and therefore be efficient. Indeed, what the argument boils down to is that some rape is motivated in part or whole by the negative interdependence of the parties' utilities, and this, as I have argued in connection with crimes of passion, is no reason for considering the act efficient.

(d) As with my earlier discussion of crimes of passion, it is important not to take too narrow a view of market alternatives. Supposing it to be true that some rapists would not get as much pleasure from consensual sex, it does not follow that there are no other avenues of satisfaction open to them. It may be that instead of furtively stalking women they can obtain satisfactions from productive activities, that is, activities in which other people are compensated and thus derive benefits. This is an additional reason to think that the total wealth of society would be increased if rape could be completely repressed at a reasonable cost.

All this may seem to be a hopelessly labored elucidation of the obvious, that rape is a bad thing; but I think it useful to point out that economic analysis need not break down in the face of such apparently noneconomic phenomena as rape.

All of the pure coercive transfers that I have discussed are intentional torts at common law; and the subset of intentional torts that consists of pure coercive transfers (not all intentional torts are such) represents the largest category of criminal acts. Let us call them category (1). There are, however, several other categories:

(2) Tax evasion, price-fixing, and other examples of nonproduc-

13. This is a general problem with inferring the character of criminal conduct from observations of actual criminals. Not only are the observations limited to the criminals who are caught, but, more importantly, they are by definition limited to criminals who are not deterred by what are rather heavy penalties even when discounted by the small probability of actually punishing most criminal acts. Observed criminal conduct must have more of a mad dog character than the deterred conduct would have if it were not deterred.

14. This is suggested by the high fraction of rapes—approaching 50% in some surveys—in which the rapist and the victim have a prior acquaintance. See McDermott, *Rape Victimization in 26 Cities* 51 (Law Enforcement Assistance Admin., U.S. Dep't of Justice (1979)) (Analytic Rep. SD-VAD-6, App. A).

tive wealth-shifting conduct made criminal by statute. In contrast, category (1) crimes were punishable at common law.

(3) Voluntary exchanges incidental to activities that the state has outlawed. Some examples are pimping and prostitution, engaging in deviant (but voluntary) sexual relations, selling pornography, selling babies for adoption, selling regulated transportation services at prices not listed in the carrier's published tariffs, and trafficking in narcotics.

(4) Certain menacing but nontortious preparatory acts such as unsuccessfully attempting or conspiring to murder someone where the victim is not injured and the elements of a tortious attempt are not present. They would not be present if, for example, the victim did not know of the attempt at the time it was made.

(5) Conduct that if allowed would thwart other forms of common law or statutory regulation. Examples are leaving the scene of an accident, bribing judges and other public officials, and fraudulently concealing assets from a judgment creditor.

(6) Blackmail, and certain other forms of private law enforcement when these are made criminal.

Categories (3) and (6) create obvious difficulties for a positive economic analysis of law. It is hard for an economist to understand why the voluntary exchange of valuable goods should be criminal. Such exchange, *prima facie* at least, promotes rather than reduces efficiency—whether it concerns hard-core or soft-core pornography, cocaine or cigarettes, common carriage or contract carriage. The qualification is important, however. Voluntary transactions may have such serious effects on third parties that when those effects are taken into account the transaction is not value-maximizing after all.¹⁵

Category (6) is mysterious because it might seem that blackmailers, vigilantes, and others who prey on criminals would be valued auxiliaries in the war on crime rather than criminals themselves. Informers *are* valued auxiliaries; why should not blackmailers be? Like informers they are private enforcers of the community's ethical norms, including those embodied in the criminal law. Although the question is too difficult to be done complete justice to here, I shall venture to suggest an answer, though not a complete one, as it is inapplicable to the punishment of one who blackmails with discreditable but not incriminating information.

A person who learns that someone else is a criminal could in a regime that allowed blackmail either sell the information to the police—police pay informers, sometimes handsomely—or sell secrecy to the criminal. By outlawing sale to the criminal, society reduces the price of information to the police (by removing a competitor from the buying side of the market) and at the same time raises expected punish-

15. Whether any or all of the examples I have given should be condemned on this ground is not a question examined in this Article.

ment costs to the criminal. For it is not true that the money the criminal pays the blackmailer is equal to the punishment costs he would undergo if he were convicted and sentenced for his crime. As we shall see, most criminals cannot pay optimal fines—and their ability to pay puts a ceiling on what the blackmailer can extract from them. It might seem, however, that if the criminal cannot bid on the information, informers will have lower incomes, so there will be less informing and expected punishment costs will decline. But the police can (in principle, at least) control the level of informing by the prices they pay informers; it is not necessary to admit the criminal into the market.

B. *Why Isn't Tort Law Enough of a Social Control?*

Although the major criminal prohibitions seem explicable as measures for discouraging inefficient behavior rather than for achieving moral objectives that economics may not be able to explain—the major exception being the prohibition of victimless crimes—this does not explain why there is a criminal law, given that there is a law of torts and that it predates criminal law.¹⁶ An explanation of why the six categories of criminal activity have not and cannot be left to tort law leaps to mind for categories (3) and (4): no one is hurt, at least in any very direct sense. But the answer is superficial because society could allow whom-ever the law was intended to protect to sue for punitive damages. A better answer is that detection is difficult where there is no victim to report and testify against the wrongdoer. The answer is incomplete because, as we shall see, punitive damages can be adjusted upward to take account of the difficulty of detection. In principle, this device could take care of category (5) crimes as well. But as we shall also see, the higher the optimal level of punitive damages, the less likely they are to be collectable.

Another question about categories (3) and (4) is, why punish acts that do not hurt anybody? For category (3) the answer lies, as I have said, outside of economics, or at least outside the scope of this Article. For category (4) the answer is bound up with the question of why tort law is not adequate to deal with categories (1) and (2) (coerced transfers in violation of common law or statutory principles). The proper sanction for a pure coercive transfer is something greater than the law's estimate of the victim's loss—the extra something being designed to confine transfers to the market whenever market-transaction costs are not prohibitive.¹⁷ We can be a little more precise: the extra something should be the difference between the victim's loss and the offender's gain, and then some. To understand this, assume first that the gain is greater than the loss: *B* has a jewel worth \$1000 to him, but worth

16. See R. Posner, *supra* note 7, at 192, 203–04.

17. See Calabresi & Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 *Harv. L. Rev.* 1089 (1972).

\$10,000 to *A*, who steals it ("converts" it, in tort parlance). We want to channel transactions in jewelry into the market, and this requires that the coerced transfer be a losing proposition to *A*.¹⁸ If *A* is risk neutral, if the probability of *B*'s getting and collecting a judgment against *A* is one (an important assumption, to be relaxed shortly), and if legal proceedings are costless, then making *A* liable for damages of only \$1000 will not do the trick, and even making him pay \$10,000 (restitution) will not quite do it, but will just make him indifferent between stealing and buying. We shall have to add something on, and make the damages, say, \$11,000.

Of course, the jewel might be worth less to *A* (or to its ultimate purchaser from *A*) than to *B*, since *A* is not planning to pay for it. In that event a smaller fine would deter *A* given our assumption that the probability of punishment is one. If the jewel is worth only \$500 to *A*, damages of \$501 will be enough. But as we cannot determine subjective values, we shall want to base damages on the market value of the jewel (especially since if the subjective value is lower than market value, he can sell, and thus realize that value as his gain from the theft, less any expenses of sale), and then add on a hefty bonus to take account of the possibility that the thief may place a higher subjective value on the jewel than does the victim.

With regard to crimes of violence, such as murder, battery, and rape, which inflict nonpecuniary as well as or instead of pecuniary loss, it is not so easy to set a money value on the victim's loss, although tort law does of course make such estimates. Quite properly, they often are very high. For a crime that creates a substantial probability of death, the optimal damages may in fact be astronomical. This is clearest in the case where one person deliberately kills another. If the average person (someone not extraordinarily altruistic toward his heirs) were asked how much money he would demand to surrender his life on the spot, his answer would be that no finite offer would be high enough, since he would get no utility out of the money. For similar reasons, the average person would demand a very high price to incur a substantial risk of death even though he might demand only a small premium to take a small risk of death.¹⁹ This nonlinearity suggests why tort law may be adequate for many small risks of death (for example, the risk of being killed in an automobile accident caused by negligence), but not for the large risks that are created by crimes of violence.

18. Of course, if we really knew the respective values to *A* and *B*, the superiority of market to coercive transactions would be less marked. It is precisely because subjective values are hard to determine except as revealed in markets that a market system is economically and ethically attractive. The example is therefore unrealistic—but its purpose is didactic rather than descriptive.

19. For good summaries of the literature on risk premia and the value of life, see Dardis, *The Value of a Life: New Evidence from the Marketplace*, 70 *Am. Econ. Rev.* 1077 (1980); Nichols, *The Regulation of Airborne Benzene, in Incentives for Environmental Protection* 145, 173–75 (T. Schelling ed. 1983).

Optimal tort deterrence of the pure coercive transfer would require even heavier punitive damages than suggested thus far, for we have ignored the problem of concealment. Being a byproduct of lawful public activities, accidents usually are difficult to conceal. But when the tortfeasor's whole object is to take something of value, he will naturally try to conceal what he is doing and will often succeed. Subject to qualifications not necessary to address here, the formula for deciding how large an award of damages (D) must be if the probability (p) that the tortfeasor will actually be caught and forced to pay the damages is less than one is

$$D = L/p \quad (1)$$

where L is the harm caused by the tortfeasor in the case in which he is caught, and includes any adjustment to discourage bypassing the market by a coerced transfer. If $p = 1$, L and D are the same amount. But if, for example, $L = \$10,000$ and $p = .1$, meaning that nine times out of ten the tortfeasor escapes the clutches of the law, then D , the optimal penalty, is $\$100,000$. Only then is the expected penalty cost to the prospective tortfeasor (pD) equal to the harm of his act (L).

Once damages for the pure coercive transfer are adjusted upward to discourage efforts to bypass the market, to recognize the nonlinear relationship between risk of death and compensation for bearing that risk, and to offset concealment, it becomes apparent that the optimal damages will often be very great—greater, in many cases, than the tortfeasor's ability to pay. This is further true because bonding or compulsory insurance cannot be used to bring monetary incentives to bear on people who lack substantial liquid assets when those people are deliberate tortfeasors. It generally is impossible to buy insurance against intentional misconduct, because of the acute moral-hazard problem.²⁰ Also, coercive transfers are more attractive to the poor than to the rich, since the poor have only a limited ability to use the market as an alternative route to getting their wants satisfied. The problem, then, is not only that the optimal penalty for pure coercive or involuntary transfers is high relative to the average person's wealth, but also that it is extremely high relative to the wealth of the people most likely to consider attempting to bring about such transfers.

This has not always and everywhere been true. Primitive and ancient societies (including Anglo-Saxon England) have relied much more heavily than has our society on a form of tort damages (usually fixed in amount rather than assessed individually in each case)—“bloodwealth,” “wergeld,” “composition”—to control crime, apparently with some success. Among the things that make this approach feasible in such societies are the lack of personal privacy, which makes

20. The moral hazard is the danger that the insured will be induced by the fact that he has insurance to commit the act against which he has insured and thereby escape the costs of the act while reaping its benefits.

probabilities of apprehension and conviction high, and the principle of collective responsibility, which makes the offender's kinship group liable for his damages, thus enabling the society to set fines that exceed the individual's ability to pay.²¹

Three responses to the problem in our society of the infeasibility of primary reliance on fines to deter antisocial behavior are possible (enforcement responses—I exclude as beyond the scope of this Article such alternatives as reducing inequalities of wealth), and all are used. One is to impose disutility in nonmonetary forms, such as imprisonment or death. Another is to reduce the probability of concealment, and so lower D in equation (1), by maintaining a police force to investigate crimes.²² A third response, which involves both the maintenance of a police force and the punishment of preparatory acts (category (4)), is to prevent criminal activity before it occurs. If, as seems a good guess but no more than that, economies of scale, coupled with the danger to political stability of encouraging the growth of private armies, make public policing more efficient than private, the state is in the enforcement picture (whether or not prosecution itself, as distinct from policing, is public or private) and thus has a claim to any monetary penalties imposed. Hence these penalties are paid to the state as fines rather than to the victims of crime as damages. The victims can seek damages if the crime is also a tort, whether common law or statutory. The optimality of this feature is discussed briefly in Part III.

In cases where tort remedies, including punitive damages, are an adequate deterrent because they do not strain the potential defendant's ability to pay, there is no need to invoke criminal penalties—penalties which, as we shall see in the next part, are costlier than civil penalties even when just a fine is imposed. In such cases, the misconduct probably will be deterred. If in a particular case it is not, even though the tort remedy is set at the correct level and there is no solvency problem to interfere with it, so that the tort remedy must actually be applied to maintain the credibility of the tort deterrent, there still is no social gain from using a criminal sanction.²³ Although in some cases, notably anti-trust cases, affluent defendants are both prosecuted criminally and sued civilly, criminal sanctions generally are reserved, as theory predicts, for cases where the tort remedy bumps up against a solvency limitation.

This means that the criminal law is designed primarily for the

21. See R. Posner, *supra* note 7, at 197.

22. An alternative would be to pay bounties to private enforcers, but that involves technical difficulties discussed in Landes & Posner, *The Private Enforcement of Law*, 4 *J. Legal Stud.* 1 (1975).

23. Tort remedies do not operate with perfect efficacy. If they did, there would never be a litigated tort case. But the occasional, and inevitable, failures of the tort system do not in themselves provide a strong argument for criminal remedies. Those remedies are necessary for classes of cases where tort law is bound to fail, as where the defendants cannot pay tort damages and therefore are not deterred by the threat of being ordered to pay them.

nonaffluent; the affluent are kept in line, for the most part, by tort law. This may seem to be a left-wing kind of suggestion ("criminal law keeps the lid on the lower classes"),²⁴ but it is not. It is efficient to use different sanctions depending on an offender's wealth. The suggestion is not refuted by the fact that fines are a common criminal penalty. They are much lower than the corresponding tort damage judgments, and hence usable even against relatively nonaffluent offenders, for two reasons. The government invests resources in raising the probability of criminal punishment above that of a tort suit, which makes the optimal fine lower than the punitive damages that would be optimal in the absence of such an investment. Second, a fine is a more severe punishment than its dollar cost. Almost every criminal punishment imposes some non-pecuniary disutility in the form of a stigma, enhanced by such rules as forbidding a convicted criminal to vote. There is no corresponding stigma to a tort judgment.

II. OPTIMAL CRIMINAL PENALTIES

A. *Limitations on Severity, with Special Reference to Fines*

We have seen that the main thing the criminal law punishes is the pure coercive transfer, or, as it might better be described in a case of tax evasion or price-fixing, the pure involuntary transfer, of wealth or utility. In discussing what criminal penalties are optimal to deter such transfers, I shall assume that most potential criminals are sufficiently rational to be deterrable—an assumption that has the support of an extensive literature.²⁵

We saw earlier that the sanction for a pure coercive transfer should be designed so that the criminal is made worse off by his act, but now a series of qualifications must be introduced. First, some criminal acts actually are wealth-maximizing. Suppose I lose my way in the woods and, as an alternative to starving, enter an unoccupied cabin and "steal" some food. Should the punishment be death, on the theory that

24. See, e.g., Carlen, *Radical Criminology, Penal Politics and the Rule of Law*, in *Radical Issues in Criminology* 7, 20–22 (P. Carlen & M. Collison eds. 1980).

25. See D. Pyle, *supra* note 3, ch. 3 (reviewing the literature). It should be noted, however, that most estimates of the elasticity of the crime rate to changes in either the probability of apprehension or conviction or the severity of punishment (changes in the former usually are found to have a greater deterrent effect than equivalent changes in the latter) are less than one. See *id.* at 39–58 for a review of the studies. Thus, increasing the average length of prison sentences by 10%—a large increase—would reduce the crime rate by less than 10%. Very large increases in the length of imprisonment would run into a serious discounting problem. See *infra* note 39 and accompanying text. Large increases in the probability of apprehension and conviction, on the other hand, would require heavy additional investments in police forces, prosecutors' offices, and courts. The costs of bringing the crime level down may help explain why the crime rate is so high, and why a high crime rate need not signify that the criminal justice system is inefficient.

This assumption of deterability is relaxed in Part III of this Article.

the crime saved my life, and therefore no lesser penalty would deter? Of course not. The problem is that while the law of theft generally punishes takings in settings of low transaction costs, in this example the costs of transacting with the absent owner of the cabin are prohibitive. One approach is to define theft so as to exclude such examples; the criminal law has a defense of necessity that probably would succeed in this example. But defenses make the law more complicated, and an alternative that sometimes will be superior is to employ a somewhat overinclusive definition of the crime but set the expected punishment cost at a level that will not deter the occasional crime that is value-maximizing.

There is a related but more important reason for putting a ceiling on criminal punishments such that not all crimes are deterred. If there is a risk either of accidental violation of the criminal law or of legal error, an expected penalty will induce innocent people to forgo socially desirable activities at the borderline of criminal activity. The effect is magnified if people are risk averse and penalties are severe. If, for example, the penalty for carelessly injuring someone in an automobile accident were death, people would drive too slowly, or not at all, to avoid an accidental violation or an erroneous conviction. True, if through the concept of intentionality and defenses such as necessity the category of criminal acts is limited to cases where, in Hand formula terms, there is a very great disparity between B and PL ,²⁶ the risk of either accident or error will be slight and the legal system will be freer about setting heavy penalties.²⁷ But not totally free: if the consequences of error are enormous, even a very slight risk of error will generate costly avoidance measures. And, as there are costs of underinclusion if the requirements of proof of guilt are set very high, it may make sense to make proof easier but at the same time make the penalty less severe in order to reduce avoidance and error costs.

Once the expected punishment cost for the crime has been set, it becomes necessary to choose a combination of probability and severity of punishment that will bring that cost home to the would-be offender. Let us begin with fines. An expected punishment cost of \$1000 can be imposed by combining a fine of \$1000 with a probability of apprehension and conviction of one, a fine of \$10,000 with a probability of .1, a fine of one million dollars with a probability of .001, etc. If the costs of collecting fines are assumed to be zero regardless of the size of the fine, the most efficient combination is a probability arbitrarily close to zero and a fine arbitrarily close to infinity.²⁸ For while the costs of apprehending and convicting criminals rise with the probability of apprehen-

26. See R. Posner, *supra* note 5; Landes & Posner, *supra* note 5.

27. See Landes & Posner, *An Economic Theory of Intentional Torts*, 1 *Int'l Rev. L. & Econ.* 127 (1981).

28. This is a major theme of Becker, *supra* note 2. He also notes many of the practical limitations of fines.

sion—higher probabilities imply more police, prosecutors, judges, defense attorneys, and so forth because more criminals are being apprehended and tried, than when the probability of apprehension is very low—the costs of collecting fines are by assumption zero regardless of their size. Thus, every increase in the size of the fine is costless, and every corresponding decrease in the probability of apprehension and conviction, designed to offset the increase in the fine and so maintain a constant expected punishment cost, reduces the costs of enforcement.

There are, however, many objections to assuming that the cost of collecting a fine is unrelated to its size:

(1) For criminals who are risk averse, an increase in the fine will not be a costless transfer payment.²⁹ In Becker's model,³⁰ the only cost of a fine is the cost of collecting it, because either the fine is not paid—the crime is deterred—or, if paid, it simply transfers an equal dollar amount from the criminal to the taxpayer. But for a risk-averse criminal, every reduction in the probability of apprehension and conviction, and corresponding increase in the fines imposed on those criminals who are apprehended and convicted, imposes a disutility not translated into extra revenue of the state. Thus, the real social cost of fines increases for risk-averse criminals as the fine increases. Nor is this effect offset by the effect on risk-preferring criminals, even if there are as many of them as there are risk-averse criminals. To the extent that a higher fine with lower probability of apprehension and conviction increases the utility of the risk preferrer, the fine has to be put up another notch to make sure that it deters—which makes it even more painful for the risk averse.

(2) The stigma effect of a fine (as of any criminal penalty), noted earlier, is not transferred either.

(3) The model implies punishment of different crimes by the same, severe fine. This uniformity, however, eliminates marginal deterrence—the incentive to substitute less for more serious crimes.³¹ If robbery is punished as severely as murder, the robber might as well kill his victim to eliminate a witness. Thus, one cost of making the punishment of a crime more severe is that it reduces the criminal's incentive to substitute that crime for a more serious one. To put this differently, reducing the penalty for a lesser crime may reduce the incidence of a greater crime. If it were not for considerations of marginal deterrence, more serious crimes might not always be punishable by more severe penalties than less serious ones.

There is, however, a tradeoff between marginal deterrence and total deterrence, as shown in equation (2):

29. See Polinsky & Shavell, *The Optimal Tradeoff Between the Probability and Magnitude of Fines*, 69 *Am. Econ. Rev.* 880, 884–85 (1979).

30. See Becker, *supra* note 2.

31. See Stigler, *supra* note 4.

$$M_r = R(f_r) \times p_{m|r}(f_r) \quad (2)$$

M_r , the number of murders committed in the course of robberies, is a product of the number of robberies (R) and the probability that, given a robbery, a murder will occur ($p_{m|r}$). Both R and $p_{m|r}$ are functions of the penalty for robbery (f_r)—but R is a negative function, and $p_{m|r}$ a positive function. It is impossible to say a priori which dominates. If R were very sensitive to increases in the penalty, there might well be fewer murders—and of course many fewer robberies—if robbery were punishable by death.

In this example the greater and the lesser offense are complements rather than substitutes. Suppose we were speaking not of robberies and murders in the course of robberies, but of auto theft and bicycle theft. If the punishment for bicycle theft were raised to that of auto theft, there would be more auto theft. Moreover, even if all crimes were punished with the same severity, some marginal deterrence could be preserved by varying the probability of punishment with the severity of the crime: that is, by looking harder for the more serious criminal. Maybe, then, marginal deterrence should not be a very important factor in the design of a schedule of penalties.

(4) Limitations of solvency cause the cost of collecting fines to rise with the size of the fine—and for most criminal offenders to become prohibitive rather quickly. The solvency problem is so acute that the costs of collecting fines would often be prohibitive even if the probability of punishment were one and fines correspondingly much smaller than in the model. This explains the heavy reliance on non-pecuniary sanctions, of which imprisonment is the most common today. Imprisonment both reduces the criminal's future wealth, by impairing his lawful job prospects, and imposes disutility on people who cannot be made miserable enough by having their liquid wealth, or even their future wealth, confiscated.

(5) The solvency limitation is made all the more acute because a fine generally is considered uncollectable unless the criminal has liquid assets to pay it.³² The liquidity problem may seem superficial and easily solved by requiring payment on the installment plan or by making the fine proportional to future earnings. But these are more costly forms of punishment than they seem, because by reducing the offender's net income from lawful activity, they increase his incentive to return to a life of crime.

(6) Very low probabilities are difficult to estimate accurately. Criminals might underestimate them or overestimate them, resulting in too little or too much deterrence.

32. However, the Sentencing Reform Act of 1984, Pub. L. No. 98-473, Title II, § 211, 98 Stat. 1837 (to be codified at 18 U.S.C. § 3551), and Criminal Fine Enforcement Act of 1984, Pub. L. No. 98-596, 98 Stat. 3134 (to be codified in scattered sections of 18 U.S.C.), not only greatly increase the fines for federal crimes but also greatly improve the methods for collecting them.

B. *Nonmonetary Sanctions*

1. "*Afflictive*" *Punishment, with Special Reference to Death.* — The foregoing analysis shows that there is a place in the criminal justice system, and a big one, for imprisonment; and perhaps for other nonmonetary criminal sanctions as well. Since the cost of murder to the victim approaches infinity, even very heavy fines will not provide sufficient deterrence of murder, and even life imprisonment may not impose costs on the murderer equal to those of the victim. It might seem, however, that the important thing is not that the punishment for murder equal the cost to the victim but that it be high enough to make the murder not pay—and surely imprisoning the murderer for the rest of his life or, if he is wealthy, confiscating his wealth would cost him more than the murder could possibly have gained him. But this analysis implicitly treats the probability of apprehension and conviction as one. If it is less than one, as of course it is, then the murderer will not be comparing the gain from the crime with the loss if he is caught and sentenced; he will be comparing it with the disutility of the sentence discounted by the probability that it will actually be imposed. Suppose, for example, that the loss to the murder victim is one hundred million dollars, the probability of punishing the murderer is .5, and the murderer's total wealth is one million dollars and will be confiscated upon conviction. Then his expected punishment cost when he is deciding whether to commit the crime is only \$500,000—much less than his total wealth.

This analysis suggests incidentally that the much heavier punishment of crimes of violence than seemingly more serious white-collar crimes³³ is not, as so often thought, an example of class bias. Once it is

33. The following table will give some sense of the disparity:

Length of Sentence, in Months, by Type of (Federal) Felony, 1981	
Violent	
First-degree murder	180.3
Bank robbery	161.2
Assault	49.4
Rape	103.6
Kidnapping	250.9
Nonviolent	
Burglary	58.0
Bank larceny	68.6
Bank embezzlement	35.6
Tax fraud	38.0
SEC fraud	33.0
Counterfeiting	45.8
Drug offenses	55.5
Bribery	35.9
Antitrust	4.1

Source: Bureau of Justice Statistics, U.S. Dep't of Justice, *Sourcebook of Criminal Justice Statistics 1982*, at 464–67 (1982).

recognized that most people would demand astronomical sums to assume a substantial risk of death, it becomes apparent that even very large financial crimes are less serious than most crimes of violence. The same people who would accept quite modest sums to run very small risks of death would demand extremely large sums to run the substantial risks that many crimes of violence create, even when death does not ensue. This point holds even if the white-collar crime (say, violating a pollution regulation) creates a safety hazard, provided that the probability that the hazard will result in the death of any given person is low. Even if it were a virtual certainty that some people would die as a result of the crime, the aggregate disutility of many small risks of death may be much smaller than a single large risk of death to a particular person. This is the nonlinear relationship between utility and risk of death that I have stressed.³⁴

By the same token the argument sketched above for capital punishment is not conclusive. Because the penalty is so severe, and irreversible, the cost of mistaken imposition is very high; therefore greater resources are invested in the litigation of a capital case. Indeed, if I am right in suggesting that the cost of death inflicted with a high probability (a reasonable description of capital punishment) is not just a linear extrapolation from less severe injuries, it is not surprising that the resources invested in the litigation of a capital case may, as one observes, greatly exceed those invested in litigation in cases where the maximum punishment is life imprisonment, even if there is no possibility of parole. The additional resources expended on the litigation of capital cases may not be justified if the added deterrent effect of capital punishment over long prison terms is small. But there is scientific evidence to support the layman's intuition that it is great.³⁵

Capital punishment is also supported by considerations of marginal deterrence, which require as big a spread as possible between the punishments for the least and most serious crimes. If the maximum punishment for murder is life imprisonment, we may not want to make armed robbery also punishable by life imprisonment, for then armed

These figures include only persons sentenced to prison; in addition, the fraction of convicted defendants who are sentenced to prison is higher in the violent crime categories. See *id.*; see also Hagan & Nagel, *White-Collar Crime, White-Collar Time: The Sentencing of White-Collar Offenders in the Southern District of New York*, 20 *Am. Crim. L. Rev.* 259 (1982) (results of multiple regression analysis reveal generally more lenient sentences for white-collar offenders).

34. This point is overlooked in "radical" critiques of criminal law. See, e.g., S. Box, *Power, Crime, and Mystification* 9 (1983).

35. See, e.g., D. Pyle, *supra* note 3, ch. 4; Ehrlich, *The Deterrent Effect of Capital Punishment: A Question of Life and Death*, 65 *Am. Econ. Rev.* 397 (1975); Ehrlich & Gibbons, *On the Measurement of the Deterrent Effect of Capital Punishment and the Theory of Deterrence*, 6 *J. Legal Stud.* 35 (1977); Layson, *Homicide and Deterrence: A Reexamination of the U.S. Time-Series Evidence (August 1984)* (unpublished manuscript) (on file at the offices of the Columbia Law Review). The evidence has not gone unchallenged, of course. See D. Pyle, *supra* note 3, ch. 4, for discussion and references.

robbers would have no additional incentive not to murder their victims. But arguments based on marginal deterrence for a differentiated penalty structure are inconclusive, as we saw earlier, particularly when the greater offense is a complement of the lesser one, as is often the case with murder. Moreover, the argument does not lead inexorably to the conclusion that capital punishment should be the punishment for simple murder. For if it is, then we have the problem of marginally deterring the multiple murderer. Maybe capital punishment should be reserved for him, so that murderers have a disincentive to kill witnesses to the murder, though again the number of such complementary murders may be less if the initial murder is punished severely.

An important application of this principle is to prison murders. A prisoner who is serving a life sentence for murder and is not likely to be paroled has no disincentive not to kill in prison, unless prison murder is punishable by death. Considerations of complementarity might argue for making out-of-prison murders capital also, since reducing the number of murders and the fraction of murderers in prison would reduce the occasions for prison murder. What makes little sense is to have capital punishment for neither out-of-prison nor prison murders, so that the latter becomes close to a free good. This is the present situation in federal law. Notice that varying the probability of apprehension and conviction cannot preserve marginal deterrence in this situation. The probability of apprehension and conviction in the prison murder case is close to one; the problem is that for the murderer already fated to spend the rest of his life in prison, there is no incremental punishment from being convicted of murder again.

Of course there is no realistic method of preserving marginal deterrence for every crime, although medieval law tried. It is a reasonable conjecture (if no more than that) that because more medieval than modern people believed in an afterlife, because life was more brutal and painful, and because life expectancy was short, capital punishment was not so serious a punishment in those days as it is today. Furthermore, because society was poor, severe punishments were badly needed and law enforcement was inefficient, so that devoting much greater resources to catching criminals would not have been feasible or productive. In an effort to make capital punishment a more costly punishment to the criminal, especially gruesome methods of execution (for example, drawing and quartering)³⁶ were prescribed for especially heinous crimes, such as treason. Boiling in oil, considered more horrible than hanging or beheading, was used to punish murder by poisoning; since poisoners were especially difficult to apprehend in those times, a heavier punishment than that prescribed for ordinary murderers was (economically) indicated.

36. This punishment was still "on the books" in 18th century England. For the grisly details, see 4 W. Blackstone, *Commentaries* *92.

The hanging of horse thieves in the nineteenth century American West is another example of a penalty whose great severity reflects the low probability of punishment more than the high social cost of the crime. But the most famous example is the punishment of all serious (and some not so serious) crimes by death in pre-nineteenth century England,³⁷ when there was no organized police force and the probability of punishment was therefore very low for most crimes.³⁸

Death is not the only modern form of "afflictive" punishment. Flogging is still used by many parents and, in attenuated form, in some schools. The economic objection to punishing by inflicting physical pain is not that it is disgusting or that people have different thresholds of pain that make it difficult to calibrate the severity of the punishment—imprisonment and death are subject to the same problem. The objection is that it may be a poor method of inflicting severe but not lethal punishment. Just to inflict a momentary excruciating pain with no aftereffects might be a trivial deterrent, especially for people who had never experienced such pain; while to inflict a level of pain that would be the equivalent of five years in prison would require measures so drastic that they might endanger the life, or destroy the physical or mental health, of the offender. For slight punishments, fines will do. Incidentally, I do not mean, by omission, to disparage noneconomic objections to "afflictive punishment." But this is an Article about economics.

The infliction of physical pain is not the only way in which the severity of punishment can be varied other than by varying the length of imprisonment. Size of prison cell, temperature, and quality of food could also be used as "amenity variables." It may seem very attractive from a cost-effectiveness standpoint to reduce the length of imprisonment but compensate by reducing the quality of the food served the prisoners; the costs of imprisonment to the state, but not to the prisoners, would be reduced. The problem is that this would make information about sanctions very costly, because there would be so many dimensions to evaluate. Time has the attractive characteristic of being one-dimensional, and differs from pain in that it has more variability. But as a matter of fact, society does vary the amenities of prison life for different criminals. Minimum security prisons are more comfortable than intermediate security prisons, and the latter are more comfortable than maximum security prisons. Assignments to these different tiers are related to the gravity of the crime, and in the direction one would predict.

2. *Imprisonment.* — If society must continue to rely heavily on imprisonment as a criminal sanction, there is an argument—subject to ca-

37. See, e.g., Langbein, *Shaping the Eighteenth-Century Criminal Trial: A View from the Ryder Sources*, 50 U. Chi. L. Rev. 1, 36-49 (1983).

38. Many capital sentences, however, were commuted to banishment to the colonies.

veats that should be familiar to the reader by now, based on risk aversion, overinclusion, avoidance and error costs, and (less clearly) marginal deterrence—for combining heavy prison terms for convicted criminals with low probabilities of apprehension and conviction. Consider the choice between combining a .1 probability of apprehension and conviction with a ten-year prison term and a .2 probability of apprehension and conviction with a five-year term. Under the second approach twice as many individuals are imprisoned but for only half as long, so the total costs of imprisonment to the government will be the same under the two approaches. But the costs of police, court officials, and the like will probably be lower under the first approach. The probability of apprehension and conviction, and hence the number of prosecutions, is only half as great. Although more resources will be devoted to a trial where the possible punishment is greater, these resources will be incurred in fewer trials because fewer people will be punished, and even if the total litigation resources are no lower, police and prosecution costs will clearly be much lower. And notice that this variant of our earlier model of high fines and trivial probabilities of apprehension and conviction corrects the most serious problem with that model—that is, solvency.

But isn't a system under which probabilities of punishment are low "unfair," because it creates *ex post* inequality among offenders? Many go scot-free; others serve longer prison sentences than they would if more offenders were caught. However, to object to this result is like saying that all lotteries are unfair because, *ex post*, they create wealth differences among the players. In an equally significant sense both the criminal justice system that creates low probabilities of apprehension and conviction and the lottery are fair so long as the *ex ante* costs and benefits are equalized among the participants. Nor is it correct that while real lotteries are voluntary the criminal justice "lottery" is not. The criminal justice is voluntary: you keep out of it by not committing crimes. Maybe, though, such a system of punishment is not sustainable in practice, because judges and jurors underestimate the benefits of what would seem, viewed in isolation, savagely cruel sentences. The prisoner who is to receive the sentence will be there in the dock, in person; the victims of the crimes for which he has not been prosecuted (because the fraction of crimes prosecuted is very low) will not be present—they will be statistics. I hesitate, though, to call this an economic argument; it could be stated in economic terms by reference to costs of information, but more analysis would be needed before this could be regarded as anything better than relabeling.

There is, however, another and more clearly economic problem with combining very long prison sentences with very low probabilities of apprehension and conviction. A prison term is lengthened, of course, by adding time on to the end of it. If the criminal has a significant discount rate, the added years may not create a substantial added

disutility.³⁹ At a discount rate of ten percent, a ten-year prison term imposes a disutility only 6.1 times the disutility of a one-year sentence, and a twenty-year sentence increases this figure to only 8.5 times; the corresponding figures for a five percent discount rate are 7.7 and 12.5 times.

Discount rates may seem out of place in a discussion of nonmonetary utilities and disutilities, though imprisonment has a monetary dimension, because a prisoner will have a lower income in prison than on the outside. But the reason that interest (discount) rates are positive even when there is no risk of default and the expected rate of inflation is zero is that people prefer present to future consumption and so must be paid to defer consumption. A criminal, too, will value his future consumption, which imprisonment will reduce, less than his present consumption.

The discounting problem could be ameliorated by preventive detention, whereby the defendant in effect begins to serve his sentence before he is convicted, or sometimes before his appeal rights are exhausted.⁴⁰ The pros and cons of preventive detention involve issues of criminal procedure that would carry us beyond the scope of this Article, and here I merely note that the argument for preventive detention is stronger the graver the defendant's crime (and hence the longer the optimal length of imprisonment), regardless of whether the defendant is likely to commit a crime if he is released on bail pending trial.

The major lesson to be drawn from this part of the Article is that criminal sanctions are costly. A tort sanction is close to a costless transfer payment. A criminal sanction, even when it takes the form of a fine, and patently when it takes the form of imprisonment or death, is not. And yet it appears to be the optimal method of deterring most pure coercive transfers—which are therefore the central concern of the criminal law. These points will be seen to have important implications for substantive criminal doctrine.

III. SUBSTANTIVE PRINCIPLES OF CRIMINAL LAW

A. Preventing Crime: Herein of Multiple-Offender Laws, Attempt and Conspiracy, Aiding and Abetting, Entrapment

1. *Introduction.* — The theory of the criminal sanction presented in Parts I and II was purely one of deterrence. The state rations the demand for crime by setting a high price for it in the form of an expected cost of paying a fine or going to prison for committing a crime, but people are actually fined or imprisoned only to maintain the credibility of the deterrent system. This view leaves many important features of

39. See Block & Lind, *An Economic Analysis of Crimes Punishable by Imprisonment*, 4 J. Legal Stud. 479, 481 (1975).

40. Both devices are found in the Bail Reform Act of 1984, 18 U.S.C.A. §§ 3141-3150 (Supp. 1985).

the criminal justice system unexplained. One is that a repeat offender is usually punished more severely than a first offender even if the repeat offender served in full whatever sentences were imposed for the earlier crimes; another is that fines often are proportional to wealth. Consumers in competitive markets are not charged higher prices just because they are wealthier than other consumers or have bought the same product previously, and they certainly are not required to give back the thing they have bought if they have not consumed it yet, as a thief would be required to do. A similar puzzle is the punishment of the "inchoate" crimes, such as attempts and (unsuccessful) conspiracies. If the purpose of the criminal law is to make the criminal regard the full costs of his acts, why punish him when his conduct, because thwarted, imposes no costs? Another puzzle is that imprisonment is often thought to serve the additional value, besides deterrence, of preventing further criminal acts by the imprisoned criminal while he is in prison. Yet assuming that the criminal justice system maintains a proper schedule of prices for unlawful acts, why should anyone care that the criminal, if not imprisoned but punished with equivalent severity by some method that left him at large, might commit further criminal acts? Presumably he would do so only if the acts were socially (as well as privately) cost-justified.

A clue to the answers to these questions lies in the fact that the emphasis on preventing, rather than simply pricing, crime falls on the common law crimes—crimes whose essence is a coerced transfer in a setting of low transaction costs. Very little of the criminal activity in this category is socially cost-justified; examples such as the theft from the cabin under conditions of dire necessity are quite rare, and that example may be a noncrime by virtue of the defense of necessity. The high incidence of the common law crimes reflects not their social desirability, which is close to zero, but the costs of making punishment severe enough to achieve one hundred percent deterrence. If we therefore take as our admittedly exaggerated working hypothesis that but for the high cost of criminal sanctions the optimum level of criminal activity would be zero, we shall be driven to conclude that these sanctions are not really prices designed to ration the activity; the purpose so far as possible is to extirpate it.⁴¹ The smaller the proportion of socially cost-justified crimes, the smaller are the social costs and the larger the potential social benefits of preventing them if prevention is possible at a reasonable cost. This point explains the emphasis in the criminal law on prevention, which would make no sense in a market setting or even an unintentional-tort setting. It explains why fines should be proportional to the criminal's wealth, quite apart from any notions of a just distribution of wealth,⁴² and why a thief who is caught

41. This distinction is stressed in R. Posner, *Economic Analysis of Law* 357-59 (1972), and in Cooter, *Prices and Sanctions*, 84 *Colum. L. Rev.* 1523 (1984).

42. The principle of the diminishing marginal utility of income implies that a heav-

should be required to return what he has stolen in addition to whatever other punishment is meted out to him, even if the victim is not seeking restitution (maybe the victim is another thief!). It also shows that imprisonment is not so much more costly socially than fines as first appears. Imprisonment confers a social benefit that fines do not, by preventing the criminal from committing crimes (except in prison—an important exception today) during the term of his imprisonment.

2. *Multiple-Offender Laws*. — The practice, systematized in multiple-offender laws, of punishing repeat offenders more severely than first offenders illustrates the last point. The repeat offender has demonstrated by his behavior a propensity for committing crimes. Therefore, by imprisoning him for a longer time we can expect to prevent more crimes during his period of imprisonment than we would do if we imprisoned a first offender, whose propensities are harder to predict, for the same period. The same prison resources “buy” a greater reduction in crime. But if this were the only basis for the heavier punishment of repeat offenders, we would not observe the practice with crimes punishable mainly by fines, because fines have no preventive effect. There is more. The practice raises the price of crime to people who, judging by their past behavior, value crime more than other people do. If our object is to minimize the amount of crime, we must “charge” more to people who value that activity more. We could do this by uniformly increasing the punishment for the particular crime, but selective increases in the severity of punishment are less costly. Heavier punishment of repeat offenders may also be necessary because the stigmatizing effect of criminal punishment diminishes with successive punishments.⁴³

An important qualification must be entered here. The effect of prevention in actually reducing crime depends on the elasticity of supply of offenders.⁴⁴ If it is very high, then the principal effect of taking one criminal out of circulation is, by making room for another, to attract a person into crime from a lawful occupation or to cause a part-

ier fine is necessary to impose the same disutility on a rich as on a poor person, assuming (perhaps dubiously) that the average rich and average poor person have the same marginal utility schedules, and are merely situated at different points on them.

43. Another reason for the practice of punishing recidivists more heavily is suggested in Rubinstein, *Offenses that May Have Been Committed by Accident—An Optimal Correction Policy for Offenses that May Have Been Committed by Accident*, in *Applied Game Theory* 406 (S. Brams, A. Schotter & G. Schwödianer eds. 1979). The fact that the defendant has committed previous crimes makes us more confident that he really is guilty of the crime with which he currently is charged, and therefore the risk of error if a heavy sentence is imposed is less. There is merit to this point, but it overlooks the fact that a previous offender is easier to convict than a first offender, because if he takes the stand the prosecution can introduce his record of convictions to try to undermine his credibility.

44. This point is emphasized in Ehrlich, *On the Usefulness of Controlling Individuals: An Economic Analysis of Rehabilitation, Incapacitation, and Deterrence*, 71 *Am. Econ. Rev.* 307, 315–19 (1981).

time criminal to allocate more of his time to crime. With regard to "business-like" crimes such as trafficking in drugs, the elasticity of supply of offenders may be quite high. But with regard to crimes of violence, it is probably low, as most people in our society are quite averse to the personal risks involved in such crimes. This analysis implies that society will place more emphasis on preventing, relative to deterring, crimes of violence than acquisitive nonviolent crimes—an interesting question for further investigation.

Another point to notice about the greater punishment of repeat offenders may help to explain why the higher the percentage of young people in the population is the higher the crime rate is. Young people, as beginners in crime, are punished less severely, hence are less deterred and less incapacitated from committing crimes.

3. *Attempt and Conspiracy.* — Consider now the punishment of attempts. A man enters a bank, intending to rob it, but a guard spots him and seizes him before he can do any harm. The fact that he came so close to robbing it indicates that he is quite likely to try again unless restrained, so by putting him in prison we can probably prevent some robberies. This is one benefit of punishing attempts. Another is to increase the expected costs of bank robbery to the robber without making the punishment for bank robbery more severe, which would create all the problems discussed earlier. The robber cannot be certain that his attempt will succeed, and if it fails he will not merely forgo the gains from a successful robbery, but will incur additional (punishment) costs.

In equation (3), for example, the net expected income ($E(I)$) from some crime in a world without punishment for attempts is the difference between the expected gain from the crime, which is a product of the gain if the crime succeeds (G) and the probability of success (p_s), and the expected punishment cost ($p_p f$). If the attempt fails, there is no gain, but no punishment cost either.

$$E(I) = p_s G - p_p f \quad (3)$$

But now suppose an unsuccessful attempt is punished by penalty f_a , imposed with a probability p_{pa} (probability of punishment for an attempt). Now the prospective criminal's net expected gain falls.

$$E(I) = p_s G - p_p f - (1 - p_s) p_{pa} f_a \quad (4)$$

Punishing attempts is thus like maintaining a public police force. It is a way of increasing expected punishment cost without making the sanction for the completed crime more severe, and we have seen that the costs of criminal sanctions mount rapidly as the sanctions are made more severe.

The attempter in our example will not, however, be punished so severely as if he had actually robbed the bank. There are two economic reasons for this: to give offenders an incentive to change their minds at the last moment (a form of marginal deterrence), and to minimize the costs of error, since there is a higher probability that an attempter re-

ally is harmless than that a person punished for an actual robbery has really done nothing.

What if the defendant had simply said to a friend, who turned out to be a police informant, "I intend to rob that bank," but had taken no steps toward accomplishing his aim? This would not be a criminal attempt. The probability that such a person would actually rob a bank is much less than it would be if he were caught on the verge of doing so, so the social benefits from imprisoning him are much less; to put it differently, the expected costs of error are higher.

Sometimes attempts fail not because they are interrupted but because the attempter has made a mistake. He may have shot what he thought was a man sleeping in a bed but it turns out to be a pillow. Or he may have made a voodoo doll of his enemy and stabbed it repeatedly in the mistaken belief that this would kill the enemy. The question for the economist is whether the nature of the mistake makes it highly unlikely that the attempter will ever succeed. If no crime will be prevented by imprisoning him, there will be no social benefit of imprisonment but cost aplenty. The second hypothetical case is of this character. But while there is some authority for not punishing the first kind of attempt,⁴⁵ there is more for not punishing the second. The issue ought to be dangerousness, not whether the defendant mistook the end or the means.

The attempt that fails because of a mistake rather than because it is interrupted provides the strongest case for punishing an attempt less severely than the completed crime. If the punishment for attempted murder were the same as for murder, one who shot and missed (and was not caught immediately) might as well try again, for if he succeeds, he will be punished no more severely than for his unsuccessful attempt.

Conspiracies to commit criminal acts are punished whether or not they succeed. Where the conspiracy succeeds, punishing it as a separate crime makes the punishment for the underlying crime greater than if only one person had committed it and also confers certain procedural advantages on the prosecutor. The special treatment of conspiracies makes sense because they are (though only on average, of course) more dangerous than one-man crimes.⁴⁶ If they were more dangerous only

45. See R. Perkins & R. Boyce, *Criminal Law* 621 (3d ed. 1982), which incidentally is an excellent treatise on criminal law, as are G. Fletcher, *Rethinking Criminal Law* (1978); W. LaFare & A. Scott, *Handbook on Criminal Law* (1972); G. Williams, *Textbook of Criminal Law* (2d ed. 1983). On preparatory crimes, such as attempt and conspiracy, Schulhoffer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. Pa. L. Rev. 1497 (1974), is a particularly good treatment. On the basic principles of criminal law, see H.L.A. Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (1968).

46. This explains why an illegal sale is not in law a conspiracy between the seller and the buyer, and why a bribe is not a conspiracy between the person paying and the person receiving the bribe. A crime *defined* as requiring the cooperation of two people need not be socially more costly than if it could be committed by one person. There is a

in the sense of committing more serious crimes, there would be no need for extra punishment; the punishment would be more severe anyway. But actually they are more dangerous in being able to commit more crimes (just as a firm can produce more goods or services than an individual) and perhaps do so more efficiently (in a private, not social, sense) by being able to take advantage of the division of labor—for example, by posting one member of the conspiracy as a sentinel, another to drive the getaway car, and another to fence the goods stolen. Although these advantages are offset to some extent by the fact that a conspiracy is more vulnerable to being detected because of the scale of its activities, that scale may also enable the conspiracy to avoid punishment through corruption of law enforcement officers. And some of the most serious crimes, such as insurrection, can be committed only by conspiracies. All this implies that the optimal punishment of conspiracies is indeed more severe than that of individuals.

A conspiracy that does not succeed is still punished. It is a form of attempt. The principal legal difference is that the conspiracy—which is to say the agreement to commit the crime—is punishable even if the conspirators do not get anywhere near the scene of the crime but are caught in the earliest preparatory stage. But again, if conspiracies are more dangerous than one-man crimes, the expected harm may be as great as in the case of the one-man attempt even if the probability of the completed crime is lower because the preparations are interrupted earlier.

4. *Aiding and Abetting*. — Related to the concept of conspiracy is the concept of aiding and abetting a crime. Consider the following cases:

(a) A witness fails to report a crime to the police.

(b) A merchant sells a fancy dress to a woman he knows to be a prostitute.

(c) A merchant sells a gun to a man who tells the merchant that he is planning to use it in a murder.

In all three cases there is an argument for imposing criminal liability: it will raise the expected costs of the (principal) criminal. In the first case, however, the avoidance costs will be very great; people who have information about crime but do not volunteer it at first will be scared to do so later. In the second case the benefits of criminal liability will be rather trivial, and this is only in part because the crime is pretty trivial (and victimless); in addition, the prostitute will incur little added cost by shopping at stores that don't know her occupation. In the third case the benefits in criminal liability seem substantial—and it is the only one where the law (occasionally) imposes such liability.⁴⁷

curious analogy here to the differential treatment of “horizontal” and “vertical” agreements in antitrust law—the former being punished more harshly.

47. For an interesting discussion, see R. Perkins & R. Boyce, *supra* note 45, at 745–47. Under the increasingly popular formulation of the aiding and abetting offense that requires the aider and abettor to share the principal's purpose, rather than just have

5. *Entrapment*. — A concept closely related to attempt is entrapment, even though the former is a crime and the latter is a defense to a criminal prosecution. Conventional legal scholarship finds the concept ill-defined and enigmatic.⁴⁸ Maybe economics can make some sense out of it. Often the police solicit or assist a person to commit a crime. The most common form of this tactic is sending an undercover agent to buy narcotics from a drug dealer who is then prosecuted for an illegal sale. It may seem odd that the law should punish a harmless act, for obviously the sale of narcotics to an undercover agent, who then destroys the narcotics, harms no one. The only important thing, it might seem, would be to get the money used for the purchase back from the seller. But the rationale is again prevention. *This* act is harmless, but it is altogether likely that the dealer, unless prevented, will make illegal sales, and we arrest and convict him now because it is much cheaper to catch him in an arranged crime than in his ordinary criminal activities. The benefits of imprisonment are virtually as great, the costs of apprehension and conviction much lower.

This sort of "entrapment" is perfectly lawful. The defense of entrapment comes into play only if the entrapped person lacked a "criminal predisposition." This fusty legal term can be given the following economic meaning: the defendant would have committed the same crime, only in circumstances that would have made it harder for the police to catch him, if he had not fallen into the police trap. But suppose that instead of just simulating the target's normal criminal opportunities, the police go further and induce him to commit crimes that he would never commit in his ordinary environment. The police offer a poor man who has no criminal record one thousand dollars to steal a bicycle; he does so, and is arrested. The resources used to apprehend and convict the man of bicycle theft are socially wasted, because they do not prevent any crimes. Had it not been for the police offer, he would not have stolen a bicycle (only doing so at a time when they were not looking); the expected benefits of theft were negative to him. Nothing is achieved by the police conduct except deflecting scarce resources from genuine crime prevention, and a defense of entrapment will lie. Police inducements that merely affect the timing and not the level of criminal activity are socially productive; those that increase the crime level are not.

knowledge of that purpose, even the third case would not be actionable. But the better economic view, for which there is some legal support, would not require the sharing of the principal's purpose if the crime is very grave. See *id.* at 746; W. LaFave & A. Scott, *supra* note 45, at 509.

48. See, e.g., Seidman, *The Supreme Court, Entrapment, and Our Criminal Justice Dilemma*, 1981 *Sup. Ct. Rev.* 111. W. LaFave & A. Scott, *supra* note 45, at 372, suggests, rather unhelpfully, it seems to me, that entrapment is contrary to public policy because it is "reprehensible."

B. *Criminal Intent*

1. *In General.* — The subjective intentions, or state of mind, of the accused criminal are a pervasive consideration in the criminal law. This is puzzling to the economist: one can read many books on economics without encountering a reference to “intent.” But in fact the concept of intent in criminal law serves three economic functions: identifying pure coercive transfers, estimating the probability of apprehension and conviction, and determining whether the criminal sanction will be an effective (cost-justified) means of controlling undesirable conduct.

If I take from a restaurant an umbrella that I mistakenly think is mine, I am not a thief; if I know the umbrella is not mine and take it anyway, I am. The economic difference is that in the first case I would have to expend resources to avoid taking the umbrella and the probability of my taking the wrong umbrella is low, so the disparity between B and PL in Hand formula terms is not great, and the risk of overdeterrence through a criminal penalty is great. In the second case, where I expend resources in order to take someone else’s umbrella (maybe I went to the restaurant for the sole purpose of stealing an umbrella), B is negative and P is high (one is more likely to bring about a harm if one wants to bring it about).⁴⁹ The problem is that the external acts involved in these two transactions are the same; only the state of mind with which they are done provides a clue to the difference in their economic character. I admit that this is not an entirely satisfactory explanation. Unless the criminal defendant confesses or makes damaging admissions, his state of mind has to be inferred from external acts; so why not infer the existence of a pure coercive transfer from those acts directly, without the intermediate step of hypostasizing criminal intent? Maybe criminal intent is just a locution that laymen use to describe a pure coercive transfer.

We must be careful to distinguish intent from awareness. Otherwise we could fall into the trap of thinking that the managers of a railroad are murderers because they know with a fair degree of confidence that their trains will run down a certain number of people at railroad crossings this year. They know, but they derive no benefit from killing. They only derive a benefit from saving the resources necessary to prevent the killing, and the benefit, social as well as private, may exceed the cost. Criminal intent is the intent to bring about a forbidden object by investing resources in its attainment.

Although the cost of trying criminal cases would be reduced by not bothering to draw a sharp line between the pure coercive transfer and the accident that it externally resembles, the result would be excessive criminal punishment, leading to all sorts of serious social costs from avoidance of lawful activity—checking umbrellas in a restaurant’s cloakroom, for example. Yet sometimes the line wavers. A well-known

49. This is emphasized in Landes & Posner, *supra* note 27, at 132–33.

example is statutory rape. The girl may look sixteen (let us assume sixteen is the age of consent), but if she is younger, a reasonable mistake will not excuse the male. Another example is felony murder: if death occurs in the course of a felony through no fault of the felon's, still he is liable as a murderer. Again we do not care about deterring activity bordering on the activity that the basic criminal prohibition is aimed at. Because we do not count the avoidance of that activity as a social cost, it pays to reduce the costs of prosecution by eliminating the issue of intent (more precisely, *an* issue of intent). The male can avoid liability for statutory rape by keeping away from young girls, and the robber can avoid liability for felony murder by not robbing, or by not carrying a weapon. In effect we introduce a degree of strict liability into criminal law as into tort law when a change in activity level is an efficient method of avoiding a social cost.⁵⁰

A related idea explains why a person who robs a federally insured bank, thereby violating federal as well as state law, is not excused from federal criminal liability just because he had no reason to know the bank was federally insured. He knew he was committing a crime, and it is but a detail that he did not know the full penal consequences—a detail offset by the savings in resource costs from not having to prove his knowledge of the bank's insured status. Conceivably, this analysis explains the punishment of "mercy killings," even in circumstances where it is likely that the killing averted more suffering than it created. Since it will often be difficult to distinguish the true mercy killing from the murder that is dressed up as a mercy killing, we cast the net of prohibition somewhat wider than the particular conduct we want to deter.⁵¹

The second function of intent in the criminal law is illustrated by the degrees of murder. First-degree murder requires premeditation. Second-degree murder requires a high degree of recklessness; a familiar example is that of the fleeing robber who shoots in the general direction of his pursuers and kills one. Voluntary manslaughter would be committed by one who killed "in the heat of passion," for example, under serious (but not adequate) provocation. The probability that the attempt to kill will succeed is clearly greater in the case of premeditated murder than in the other cases. In addition, the probability of apprehension and conviction is greater in the case of voluntary manslaughter than in the first two cases. The pattern of severity therefore appears to make economic sense. It would make even more sense if within the class of first-degree murders a distinction were made between the planned and the unplanned. Generally this is not done, but there is a hint of the distinction in many murder statutes that single out killing

50. For the tort analysis, see Shavell, *Strict Liability versus Negligence*, 9 J. Legal Stud. 1 (1980); see also Landes & Posner, *supra* note 5, at 871, 875-76, 905-12 (economic comparison of negligence and strict liability standards).

51. An alternative approach might be to require the mercy killer to prove beyond a reasonable doubt that it was a true mercy killing.

while "lying in wait" or by means of poison as examples of first-degree murders.⁵² These are advance-planning murders, where not only is the probability of death maximized, but the probability of apprehension and conviction is minimized, as one who plans a murder in advance will also take steps to escape detection afterward.

There might seem to be another reason for punishing the impulsive crime less severely than the deliberated one: the impulsive crime is less deterrable; punishment is less efficacious, less worthwhile, and therefore society should buy less of it. But this analysis is incomplete. To begin with, the fact that a given increment of punishment will deter the impulsive criminal less than the deliberate one could actually point to heavier punishment for the former. Suppose that a twenty year sentence is enough to deter virtually all murderers for hire, but to achieve the same deterrence of impulsive murderers would require a sentence of thirty years if it were not for differences in the probability of apprehension and conviction. The additional sentence is costly, but if the cost is less than the benefits in additional deterrence it may still be a good investment. And we must not forget the incapacitative effect of imprisonment. The fact that certain criminals may not be deterrable argues for greater emphasis on their incapacitation, which implies long prison terms.

2. *Insanity*. — The conflict between deterrence and incapacitation as objectives of the criminal sanction is keenest in relation to the defense of insanity. If a person is insane in the sense either that he does not know that what he is doing is criminal (he kills a child, thinking it is a gerbil) or that he cannot control himself (supernatural voices command him to kill), the threat of criminal punishment will not deter him. So if the only purpose of the criminal sanction were deterrence, it would be doubtful that such people should be punished as criminals. The resources consumed in punishing them, including the disutility of the punishment to the "criminal" himself, would buy very little deterrence—not zero, though. The existence of an insanity defense attracts resources to proving and disproving it, and deterrence is impaired to the extent either that criminals succeed in faking insanity or that a reduction (for whatever reason) in the number of people punished reduces the deterrent signal that punishment emits.

Once the preventive or incapacitative goal is brought into play, however, an insanity defense is weaker. It increases the cost of the criminal process without reducing the need to incapacitate the defend-

52. See, e.g., 18 U.S.C. § 1111(a) (1982) (federal murder statute) ("Every murder perpetrated by poison, lying in wait, or any other kind of willful, deliberate, malicious, and premeditated killing; or committed in the perpetration of, or attempt to perpetrate, any arson[,], escape, murder, kidnapping, treason, espionage, sabotage,, [sic] rape, burglary, or robbery; or perpetrated from a premeditated design unlawfully and maliciously to effect the death of any human being other than him who is killed, is murder in the first degree.").

ant. However, part of the "kick" of criminal punishment comes from its stigmatizing effect, an effect enhanced by such sanctions as taking away a convicted felon's right to vote, but more deeply grounded in the sense that criminal punishment is reserved for serious wrongdoing—for what in economic terms is socially more costly conduct than is characteristically dealt with by tort law. Stigma has no incapacitative effect; it is therefore wasted on the undeterrable. Incapacitation will still be cost-effective if the expected costs inflicted by the insane person exceed the costs to society of incapacitating him, but there is no reason to brand him a criminal; civil commitment will therefore be preferable. An additional reason for incapacitating the criminally insane is to raise the costs of faking an insanity defense; the defendant will not get off scot-free even if the defense succeeds.

To say that criminal punishment should be reserved for people who are deterrable may seem inconsistent with my emphasis on the preventive function of criminal punishment, and with my remarks about strict liability in the criminal law. But there is no inconsistency. Because criminal sanctions are so costly, they have to be set at levels that do not deter everyone, but it does not follow that a person who is not deterred is not a wrongdoer. He is just someone for whom criminal activity is utility maximizing. As for strict liability, it *will* deter, by inducing a change in activity level.

By arguing that insanity should be a defense to criminal liability to the extent that it isolates a class of undeterrables, I necessarily reject the *Durham*⁵³ test of insanity, which requires only that the criminal act be shown to have been a product of the defendant's insanity. The test amounts to asking whether, but for being insane, the defendant would not have committed the act. Even if he would not have, this tells us very little about whether he could have been deterred by threat of punishment. If he could have been deterred, it is efficient to punish him. This incidentally is why it is no defense to a criminal charge that the defendant would not have committed the crime but for a bad upbringing, racial discrimination, or some other condition beyond his control. Provided the criminal act is socially undesirable, and the criminal deterrable by threat of punishment, it makes economic sense to punish him even if his incentive to commit the crime might have been reduced by some other measure, especially if the alternative measures would be very costly.

The relevant meaning of insanity, then, is that the defendant is so mentally diseased as to be undeterrable; the reason for requiring proof of mental disease, and not just proof of undeterrability, is to focus inquiry and to reduce the risk of a legal error in the defendant's favor. This seems the approximate meaning of the *M'Naghten*⁵⁴ rules as sup-

53. *Durham v. United States*, 214 F.2d 862 (D.C. Cir. 1954).

54. *M'Naghten's Case*, 8 Eng. Rep. 718 (1843).

plemented by the concept of "irresistible impulse."

If mental incapacity is self-induced, as where a person kills in a drunken fit, the law does not excuse the crime, but does reduce the severity of the punishment through the concept of diminished responsibility. The threat of punishment for acts committed while drunk reduces the incentive to become drunk in the first place. However, punishment is lightened in recognition of the fact that the drunk is less likely to succeed in doing harm than if he were sober and also is more likely to be apprehended and convicted.

Insanity is rarely recognized as a defense in tort law⁵⁵—and in general the defendant's state of mind is much less likely to be considered an excuse or mitigation of civil than of criminal liability. Criminal sanctions, as I have emphasized, are more costly than tort sanctions, and this alters the tradeoff between the costs of factfinding and the costs of imposing a sanction on conduct outside the intended domain of the sanction. Thus, it is no defense to civil trespass that the trespasser did not know and could not at reasonable cost have found out that he was on the plaintiff's property,⁵⁶ but it is a defense to criminal trespass. Since the sanction for civil trespass is less severe, the costs of a difficult inquiry into the defendant's state of mind are less likely to produce an offsetting benefit in avoiding the costs of imposing a sanction on conduct that no one wants to deter—a trespass unavoidable in an economic sense—than where the sanction is criminal.

3. *The Costs of Information.* — It often is unclear whether a buyer of stolen goods knows they are stolen. The test of criminal liability is whether, suspecting they were stolen, the buyer "consciously avoided" acquiring the knowledge that would verify or dispel his suspicions. The test places on the buyer a legal duty, enforceable by criminal punishment, of investigating the provenance of the goods when the costs of investigation are extremely low.⁵⁷ Something similar may be at work in the hoary maxim, which still retains much vitality, that ignorance of the law is no defense to criminal liability. Because unclear criminal laws can impose substantial "steering clear" costs, these laws are generally rather clear, less by being clearly drafted than by being confined to a type of conduct that everyone knows is antisocial. The cost of acquiring knowledge of one's duties under criminal law is thus made extremely low.

C. *Recklessness, Negligence, and Strict Liability Again*

Focusing on one type of sanction cost, the "steering clear" cost, will help to explain why accidental conduct is much less likely than in-

55. See W. Keeton, D. Dobbs, R. Keeton & D. Owen, *Prosser and Keeton on the Law of Torts* 1072-75 (5th ed. 1984).

56. See *id.* at 74-75.

57. See J. Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality* 178 (1979).

tentional conduct to be made criminal. A characteristic of accidental conduct is that it cannot be avoided with certainty but can only be made less probable. To be absolutely certain of never hitting another car as a result of negligence, one must forgo driving altogether. Since criminal sanctions are severe, to attach them to accidental conduct would create incentives to avoid what may be a very broad zone of perfectly lawful activity in order to avoid the risk of criminal punishment.

But there are many exceptions to this generalization; here are the principal ones.

(1) There is an argument for criminal liability whenever B in the Hand formula is low relative to PL and where L is high. If B and PL are close together there is a substantial risk of erroneously imposing liability, and the social costs of that risk are greatly magnified when the liability is criminal. Even if B is much smaller than PL , if L is small there is no reason why the matter cannot be left to the tort system. But suppose both conditions are satisfied, as where by driving extremely carelessly one creates a substantial risk of killing someone. Granted, B will be larger and P smaller than if one is trying to kill someone, but that means only that the case for criminal liability is stronger in the intentional case. The reckless, or grossly negligent, case still fits the basic model for criminal liability, and one is therefore not surprised to find that reckless and grossly negligent life-endangering conduct is criminal.

Another example is killing in the honest but unreasonable belief that it is necessary in self-defense. This is a deliberate killing; hence both P and L are high. B is also high; the killer by definition fears for his own life. Nevertheless the gap between PL and B may well be substantial, which together with the fact that L is large would establish the conditions for criminal punishment of conduct that is in an important sense accidental. In the example, the crime would be manslaughter, not murder; the gap between PL and B is smaller than in the case put earlier of reckless killing that is punishable as second-degree murder.⁵⁸

Consistent with this analysis, simple negligence, and gross negligence that does not endanger life, are rarely made criminal.⁵⁹

(2) There are, of course, strict liability crimes—where neither intent nor even simple negligence is an element of the crime. The most important as a practical matter is driving above the speed limit. But

58. "Recklessness" is confusingly used in two distinct senses in criminal law: indifference to consequences (shooting into a crowded room for fun, not caring whether or not you hit anybody), and extreme carelessness. Reckless conduct in the first sense is assimilated to intentional conduct in accordance with my discussion of criminal intent and information costs; a reckless murderer in this sense would in most states be guilty of second-degree murder. Recklessness in the second sense is assimilated to gross negligence, and in a death case would make the defendant guilty of manslaughter. For a good discussion of the importance of the actor's state of mind in distinguishing these two senses of unintended murder, see Note, Defining Unintended Murder, 85 Colum. L. Rev. 786 (1985).

59. See W. LaFave & A. Scott, *supra* note 45, at 209-13.

except in extreme cases, where it becomes part of category (1) above, this is not a crime in a functional sense, as it is punished by a small and nonstigmatizing fine, the practical equivalent of tort damages.

The interesting question about the speeding offense, and about other strict liability crimes, such as selling liquor to a child and selling adulterated foods, is why it is thought necessary to supplement tort remedies with any sort of publicly enforced sanction. My answer draws on the analysis by Wittman and Shavell of *ex ante* versus *ex post* sanctions.⁶⁰ In the case of life-endangering conduct, a feature of virtually all strict liability crimes, the fixing of L in a tort suit (that is, after the accident has occurred) is difficult to do because it is difficult to estimate the value of a human life, and may be a futile act because the tortfeasor may lack the money to pay a large judgment. L , then, is both uncertain and large. The alternative to regulation through the tort system is to have the government step in and (ideally) make the speeder pay a fine equal to PL , in order to induce the taking of the right precautions (which is approximated by complying with the speed limit). PL , the expected cost of being endangered, can be estimated from studies of compensating wage differentials for dangerous work and personal investments in safety, and since PL will be a much smaller number than L , there is unlikely to be a solvency problem.⁶¹

Strict liability is a misnomer in this setting. The speed limit is a rough estimation of B , and so with the other regulatory rules the breaking of which establishes strict criminal liability. Because B and PL may be close together, costly criminal sanctions would not be optimal even though L is high, but much cheaper transfer payments to the government may be optimal.

(3) An important doctrine of strict liability in tort law is respondeat superior: the employer is liable regardless of his personal fault for torts committed by an employee within the scope of the employment. The basic justification for the doctrine is that employees rarely can pay substantial money judgments, and therefore tort liability will have little effect on their incentives. If the employer is liable, his incentives will be productively affected—he will take greater care in hiring, supervising, and where necessary, firing employees. Since the criminal law does not rely primarily on monetary sanctions, since imposing criminal sanctions on the employer would duplicate tort sanctions, and particularly since criminal sanctions can induce too much care because they are so heavy,

60. See Shavell, *Liability for Harm Versus Regulation of Safety*, 13 J. Legal Stud. 357 (1984); Wittman, *Prior Regulation Versus Post Liability: The Choice Between Input and Output Monitoring*, 6 J. Legal Stud. 193 (1977).

61. The analysis is not entirely satisfactory. If PL is known, L can be calculated, for purposes of figuring tort damages, and potential injurers can be made to insure (which is practical since we are speaking of accidental rather than intentional misconduct), thus solving the problem of insolvency.

it is no surprise that the criminal law has not adopted respondeat superior.

The major exception is the criminal liability of corporations.⁶² If a crime at least ostensibly on the corporation's behalf is committed or condoned at the directorial or managerial level of the corporation, the corporation is criminally liable. This means that the shareholders will bear the burden of the fine. They are analogous to employers of the people who did the actual deed. Since a corporation can only be fined, since corporations are either risk neutral or if risk averse less so than individuals, and since there is little stigma to corporate punishment (a corporation can act only through individuals, and there is a constant turnover of these individuals), corporate criminal punishment is much less costly than individual punishment. So there is much less danger of causing the shareholders to be too careful in hiring, supervising, and terminating directors (and through the board of directors, the managing employees).

These circumstances make corporate criminal liability sensible. Assume to begin with that the corporation's managers are perfect agents of the shareholders, so that any revenue obtained from criminal activity inures to the shareholders. Then if the shareholders bear no responsibility for a manager's crime, they will have every incentive to hire managers willing to commit crimes on the corporation's behalf. Of course the shareholders will have to compensate the managers for the expected costs of criminal punishment, but, given the limitations on the severity of criminal sanctions emphasized earlier, they may be able to do this and still have an expected gain from corporate criminal activity.

Now assume that the managers are not perfect agents of the corporation—that in fact they use their corporate positions to facilitate criminal activity intended to enrich themselves. Even so, the corporation has supplied the facilities they are using, and its owners should be given incentives to select and supervise managers more carefully.⁶³

The real puzzle about corporate criminal liability, it might seem, is why it has to be *criminal* liability. The entire rationale of the criminal law is that the optimal tort remedy is sometimes too large to be collectible, and how can that be a consideration with an entity that can only be subjected to monetary sanctions? But corporations are not infinitely solvent, and two of the fundamental techniques of criminal law are fully applicable even to an entity that cannot be punished other than by a

62. This subject is comprehensively surveyed in K. Brickey, *Corporate Criminal Liability: A Treatise on the Criminal Liability of Corporations, Their Officers and Agents* (1984); see also Metzger, *Corporate Criminal Liability for Defective Products: Policies, Problems, and Prospects*, 73 *Geo. L.J.* 1 (1984) (discussing whether the deficiencies of product liability law are so great as to warrant imposing criminal sanctions on corporations).

63. Notice the analogy to collective punishment in primitive societies. See *supra* note 21 and accompanying text.

nonstigmatizing fine: the use of public resources to raise the probability of punishment above what might be a very low level because of efforts taken to conceal criminal responsibility, and the punishment of preparatory activity in order to reduce the net expected gain from crime.

Since, however, corporate criminal punishment is purely monetary, it is not clear why the corporation should be entitled to the elaborate procedural safeguards of the criminal process. Those safeguards make economic sense only on the assumption that criminal punishments impose heavy social costs rather than merely transfer money from the criminal to the state.⁶⁴

D. *The Defense of Necessity*

The famous case of *Regina v. Dudley and Stephens*⁶⁵ involved a murder trial of two men who, *in extremis* in a lifeboat, killed and ate one of their crew mates (the cabin boy). A defense of necessity was raised but rejected. In the modern law the defense of necessity, though still regarded with disfavor except when it takes the form of self-defense, will usually succeed if there is a very great disparity between the cost of the crime to the victim and the gain to the injurer. In our earlier example of “stealing” food from a cabin in the woods in order to maintain life, the “theft” probably would be excused. Notice also that, unlike the case of insanity—a fundamentally different type of defense—no incapacitative goal would be served by rejecting a defense of necessity; we do not *want* incapacitation in this case.

But change the example slightly: I am starving, and beg a crust of bread from a wealthy gourmand, who turns me down. If I go ahead and snatch the bread from his hand, I am guilty of theft, and cannot interpose a defense of necessity. The economic rationale for this hard-hearted result (a good illustration of the difference between efficiency and utility as grounds of criminal punishment) is that, since transaction costs are low, my inability to negotiate a successful purchase of the bread shows that the bread is really worth more to the gourmand, in the strictly economic sense in which value is a function of willingness and hence ability to pay. But transaction costs were prohibitive in the cabin example.

In *Dudley and Stephens*, there was evidence that the cabin boy was near death anyway and that killing and eating him saved the lives of three men.⁶⁶ Yet we know that unless the victim knew he was too far gone to be saved, probably he would not have sold his life to the others at any price. Therefore the case seems similar to that of the starving

64. See R. Posner, *Economic Analysis of Law* 433–34 (2d ed. 1977).

65. 14 Q.B.D. 273 (1884).

66. The third was not charged because he had not participated in the killing. See *id.* at 274.

beggar. Yet something must be wrong. Even though transaction costs were not high in the usual sense in *Dudley and Stephens*, at some point the sacrifice of one person so that others will live must increase social welfare. If in advance of the voyage the members of the crew had agreed to sacrifice the weakest should that become necessary to save the others, there would be an economic argument for allowing the defense of necessity if the agreement had to be performed.

It is only one step beyond that to argue that if we are confident that the members of the crew would have made such an agreement if they had foreseen the contingency that materialized, the defense of necessity ought to have been recognized—always assuming that economic efficiency is to be the guidepost for criminal law doctrine. There actually is some legal support for the related idea that killing and eating a person in conditions of desperation will be excused if lots are drawn to determine who shall be the victim.⁶⁷ Compared to an *ex ante* agreement to sacrifice the weakest, the drawing of lots has the advantage of being both cheaper to administer and a better insurance scheme. An agreement to sacrifice the weakest will give whoever has the least robust constitution the least insurance protection. But drawing lots has the disadvantage that it may result in unnecessary sacrifices, as where the strongest man draws the short straw and is killed and eaten but the weakest dies anyway because he was too far gone to be saved.

CONCLUSION

If the analysis in this Article is sound, the criminal law, though generally considered the domain *par excellence* of moral rather than economic thinking in law, has an impressive economic logic. On reflection this conclusion (which can be reinforced by reference to the literature on the economics of criminal procedure)⁶⁸ is not weird as it sounds. Criminal acts are a source of enormous social costs that no society can ignore, and the modern criminal law is the product of a painstaking evolution powerfully influenced by the explicitly economic approach of Jeremy Bentham. Although judges and legislators do not often speak the language of economics, this Article suggests that they often do reason implicitly in economic terms, and that economic analysis is therefore helpful in explaining the basic structure of law, including the criminal law.

If this Article were not already so long, I would go on and compare the economic approach with its principal rival, the “moral” theory

67. See *United States v. Holmes*, 26 F. Cas. 360 (E.D. Pa. 1842) (No. 15,383); R. Perkins & R. Boyce, *supra* note 45, at 1056; cf. A. Simpson, *Cannibalism and the Common Law* 140, 145, 233–34 (1984).

68. See, e.g., Easterbrook, *Criminal Procedure as a Market System*, 12 J. Legal Stud. 289 (1983) (arguing that important aspects of criminal procedure—prosecutorial discretion, plea bargaining, and sentencing discretion—can be understood in economic terms).

of criminal law, which argues that the criminal law should only punish morally blameworthy conduct. Whatever the normative merits of this approach, I doubt that it is as good a positive theory of criminal law as the economic,⁶⁹ since in so many areas conduct is punished that is not blameworthy in the moral sense. But this is a topic for another day.

69. I am quite sure that the economic approach has more power than the approach of the radical criminologists. See *supra* text accompanying notes 24 and 34.