Prices and Sanctions

Robert Cooter

Berkeley Law
PRICES AND SANCTIONS

Robert Cooter*

INTRODUCTION

Scholars of jurisprudence traditionally view law as a set of obligations backed by sanctions, or commands backed by threats.1 In contrast, economists tend to view law as a set of official prices.2 Associated with each of these viewpoints is a characteristic blindness. The jurisprudential perspective blinds lawyers to the fact that officials cannot regulate the economy efficiently by giving orders.3 Instead, they must rely upon legal instruments similar to prices. Conversely, the economic perspective is blind to the distinctively normative aspect of law, viewing a sanction for doing what is forbidden merely as the price of doing what is permitted. In brief, the economic analysis of law lacks a clear account of sanctions, and the jurisprudential tradition lacks a good account of prices. This Article attempts to bridge the two traditions by developing a theory about the difference between the effect of prices and sanctions upon behavior.

Part I of this Article offers working definitions of the two key concepts, defining a sanction as a detriment imposed for doing what is forbidden, and a price as money extracted for doing what is permitted. Officials should create prices to compel decisionmakers to take into account the external costs of their acts, whereas officials should impose sanctions to deter people from doing what is wrong. Part I then explains the difference in the way sanctions and prices typically control behavior. A sanction typically creates an abrupt jump in an individual’s costs when he passes from the permitted zone into the forbidden zone where behavior is sanctioned. This abrupt

* Professor of Law, University of California, Berkeley. B.A. 1967, Swarthmore College; M.A. 1969, Oxford; Ph.D. 1975, Harvard University. I would like to thank Bruce Ackerman, Lucian Bebchuk, Alvin Klevorick, Mitch Polinsky, Joseph Raz, Dan Rubinfeld, Steve Shavell, Judy Thompson, and my colleagues on Berkeley’s law faculty for helpful comments.

1. The view that law is the command of the sovereign or the obligations that the community imposes upon its members is so old that its origins cannot be established. John Austin stated the imperative theory with clarity in 1 Lectures on Jurisprudence (R. Campbell 5th rev. ed. 1911) and in The Province of Jurisprudence Determined (2d ed. 1861).

2. Two lawyers who have adopted the economic view that laws are prices have stated: [M]anagers do not have an ethical duty to obey economic regulatory laws just because the laws exist. They must determine the importance of these laws. The penalties Congress names for disobedience are a measure of how much it wants firms to sacrifice in order to adhere to the rules; the idea of optimal sanctions is based on the supposition that managers not only may but also should violate the rules when it is profitable to do so.


jump in costs disappears if the sanction is replaced by a price. Behavior is more elastic with respect to changes in prices than to changes in sanctions.

This fact is used in Part II to develop a prescription for guiding lawmakers’ choice between prices and sanctions. Since behavior is relatively responsive to prices, an error in setting an official price will have a large effect upon behavior; since behavior is relatively unresponsive to the level of a sanction, modest errors in the level of sanctions will have little effect upon behavior. Finally, since most people will obey a reasonable legal standard backed by a reasonable sanction, an error in setting the legal standard will cause an error in most people’s behavior.

These facts lead to a simple prescription. If lawmakers can identify socially desirable behavior, but are prone to error in assessing the cost of deviations from it, then sanctions are preferable to prices. However, if officials can accurately measure the external cost of behavior, but cannot accurately identify the socially desirable level of it, then prices are preferable to sanctions.

Part III develops a simple test for preliminary classification of laws according to whether they impose sanctions or create prices. Sanctions usually increase if the actor’s offense is intentional or repeated, while prices do not.

In Part IV, this test is used to classify areas of tort, contract, criminal, and regulatory law. For example, strict liability imposes a price, whereas negligence imposes a sanction. Furthermore, the types of torts governed by strict liability and negligence conform to the prescription developed in Part II.

The Article is followed by an appendix in which the fundamental arguments are mathematically formulated as propositions and proved.

I. CONTRASTING PRICES AND SANCTIONS

A. Definitions

The mere existence of an obligation or issuance of a legal command may provide insufficient motivation for obedience. In addition, a sanction or threat may be needed to induce conformity. A sanction is a detriment imposed for doing what is forbidden, such as failing to perform an obligation. For example, a defendant in a tort dispute may be ordered to pay compen-

4. A norm is a general name for rule-like guides to action, including rules, laws, and principles. See J. Raz, Practical Reason and Norms (1975); G.H. von Wright, Norm and Action (1963). This Article discusses norms that impose an obligation, such as “All citizens have a duty to vote,” or “Working on the Sabbath is forbidden.” A command is an act in which the speaker describes what he wants done with the intention that this description should cause the hearer to do it. See J. Searle, Speech Acts (1970). Although philosophers have dwelt upon the distinction between commands and norms, it is unimportant for my purposes so I shall limit my discussion to norms backed by sanctions, as opposed to commands backed by threats.
satory damages for an injury caused by his negligence, or a convicted crimi-
nal may be sentenced to jail.

In contrast, a price is payment of money which is required in order to
do what is permitted. For example, a company may buy goods in the market-
place, but it must pay the seller's price. Similarly, individuals are permitted
to earn income, but obliged to pay taxes on their earnings.

These definitions of sanction and price are not always consistent with
ordinary speech. Tax evasion is forbidden, but in casual speech people
often say a fine is the price of tax evasion, when by these definitions it is a
sanction. Furthermore, these definitions are unlikely to satisfy philosophers
in search of necessary and sufficient conditions for "sanction" and "price." To illustrate, paying an arsonist not to burn down a store is extortion, yet
protection money fits the definition of a price—an exaction for doing what
is permitted. Despite such shortcomings, these definitions lay the basis for
the behavioral model developed in the next section.

B. Incentive Effects of Prices and Sanctions

Assume that someone engages in an activity that imposes costs upon
others—external costs. The external cost might be the harm caused by an
accident, the nuisance created by a polluter, or the injury that arises from
breach of contract. External costs can usually be reduced by the injurer at
some expense to himself—accident costs can be reduced by precaution, pol-
lution can be reduced by abatement, and the probability of breach can be
reduced by care in performance. This Article uses "precaution" as a ge-
neric term to refer to any costly activity that reduces external costs.

The relationship between precaution and external costs is summarized

---

**FIGURE 1**

**SOCIAL COSTS**

---
in Figure 1. Social costs, which refer to the sum of external costs and the cost of precaution, are shown on the vertical axis, and the actor's precaution is shown on the horizontal axis. This Article assumes that a small increase in precaution reduces external costs by a large amount when precaution is low, and a small increase in precaution reduces external costs by a small amount when precaution is high. Consequently, the graph of social costs is U-shaped and there is a level of precaution, denoted $x'$, which minimizes social costs.\(^5\)

1. **Incentives in the General Case.** — Now consider two rules for allocating the costs represented in Figure 1. Rule A partitions levels of precaution into a permitted zone and a forbidden zone. Specifically, Rule A creates a legal standard of precaution $x^*$, permitting the actor to take precaution as great or greater than the legal standard ($x \geq x^*$), and forbidding the actor to take less precaution than the legal standard ($x < x^*$). Furthermore, under certain conditions, Rule A exacts a sum of money as a sanction for being in the forbidden zone.

Figure 2 illustrates Rule A.\(^6\) It shows precaution on the horizontal axis

---

5. The social cost ($SC$) of accidents can be written $x + \rho(x)r$, where $x$ is expenditure on precaution by the potential injurer, $\rho = \rho(x)$ is the probability of an accident, and $r$ is the money value of the harm caused by an accident. Choosing $x$ to minimize social costs yields the equation: $0 \geq 1 + \rho(x)r$. The solution of this equation for $x$ yields $x'$. For a more general treatment, see infra Appendix, pp. 1554-61.

6. This figure first appeared in the law and economics literature in Brown, Toward an Economic Theory of Liability, 2 J. Legal Stud. 323, 341 (1973), but the discontinuity was not analyzed.
and the actor's private costs on the vertical axis. There are two components of the actor's private costs: precaution and sanctions. When precaution is below $x^*$, the actor bears the cost of his precaution and he must also pay the sanction; when the actor's precaution is above $x^*$, he pays only for his own precaution.\(^7\)

To keep Figure 2 simple, this analysis assumes that the sanction equals the external cost of the act—the actor is liable for the full harm which he causes. Thus, the curve in Figure 2 is identical to the curve in Figure 1 at those values of $x$ below $x^*$. As precaution increases, there is a jump or discontinuity in the cost curve at the legal standard $x^*$, because that is the cutoff point for imposing the sanction. Above $x^*$, the sanction is not exacted, so the actor's private costs equal the cost of his precaution.\(^8\)

A person who is rationally self-interested will choose his level of precaution to minimize his private costs. If he faces costs like those created by Rule \(A\), he will search for the lowest point on the cost curve. This point occurs when his precaution equals the legal standard $x^*$. Thus, a self-interested person subject to Rule \(A\) will take just enough precaution to satisfy the legal standard and escape liability.\(^9\)

It is not essential that the sanction equal the harm caused by the act to induce the self-interested actor to take precaution $x^*$. It is only essential that the sanction be large enough so that his private costs are minimized by conforming to the legal standard $x^*$. This point is illustrated in Figure 3, which shows the actor's private costs under four different assumptions about the sanction. The crucial feature of Figure 3 is that the lowest point in the graph is at $x^*$, regardless of whether the sanction corresponds to the high level \(A\), the medium level \(B\), or the low level \(C\). Any one of these three sanctions will induce conformity with the legal standard $x^*$. However, if the sanction falls to the very low level indicated by \(D\), then $x^*$ is not the lowest point on the cost curve, so the actor will minimize his private costs by taking less precaution than the legal standard.

Another kind of rule achieves similar ends by different means.\(^10\) In-

---

7. Using the same notation, see supra note 5, to the left of the discontinuity, the potential injurer's expected costs are $x + \mathcal{P}(x)sr$, where $sr$ stands for the injurer's expected liability in the event of an accident. To the right of the discontinuity, the potential injurer's costs are $x$.

8. See Cooter, Economic Analysis of Punitive Damages, 56 S. Cal. L. Rev. 79 (1982). This will hold true even if enforcement of the sanction is lax. Imperfect enforcement will reduce the size of the discontinuity in the cost curve, but this Article's analysis will be correct as long as the amount of enforcement is large enough to maintain some discontinuity.

9. Someone who acts from duty and obeys the law out of respect will also satisfy the legal standard, but not in order to minimize private costs.

10. This Article distinguishes two kinds of rules and identifies one of them with prices and the other with sanctions. Of course, it is possible to expand the definition of a price to cover sanctions by defining a sanction as a discontinuous price, or to expand the definition of a sanction to cover prices by defining a price as a continuous sanction. However, the important point is not to argue about names but to understand the differences in the behavior
instead of dividing precaution into permitted and forbidden zones, Rule B requires the actor to pay the external costs that his activity imposes upon others. Under Rule B, the private costs of the actor are equal to the social costs of his activity; thus the social cost is internalized. To illustrate Rule B, relabel the vertical axis in Figure 1 to read "private costs = social costs." Because the actor's private costs are lowest at \( x' \) when he is liable for the external cost of his acts, his private costs are minimized at the same level of precaution that minimizes social costs.

The incentives created by the two rules work in different ways. With Rule B, the actor balances the cost of additional precaution against the resulting reduction in the external costs for which he is responsible. His private costs are minimized when the benefits and costs of a small change in precaution offset each other.\(^{11}\) By contrast, with Rule A, the benefits and costs of a small change in precaution are not offset, because the actor's private costs are much lower when he just satisfies the legal standard than they would be if he just failed to satisfy it.\(^{12}\)

The incentive effects of prices and sanctions are different because benefits and costs are usually equipoised with prices but not with sanctions.\(^{13}\)

---

**Figure 3**

RESPONSE TO DIFFERENT SANCTIONS

---

\( ^{11} \) This assumes the actor is not on a flat portion of his cost curve. Flat portions of such curves are unusual but possible.

\( ^{12} \) In technical language, the actor under Rule A is at a corner on the cost function where the cost of an incremental change exceeds the benefit.

\( ^{13} \) I say "usually" because there are exceptional persons and special classes of rules that do not fit my model. Although many types of commands and obligations can be modeled by
When benefits and costs are equipoised, a small change in them causes a change in behavior. Thus, under Rule B, the actor’s precaution is responsive to changes in the price. If the price equals a fraction of the external harm, then reducing the fraction will cause the level of precaution which minimizes private costs to be reduced. This fact is illustrated in Figure 4, where lowering the price causes the lowest point on the cost curve to shift to the left.14 By contrast, when benefits and costs are not equipoised, moderate changes in them do not cause behavior to change. Thus, as was shown in Figure 3, the actor’s precaution is unresponsive to modest changes in the sanction or the frequency of its application.15

The difference in responsiveness can be summarized in a sentence: The amount of precaution an actor takes is more elastic with respect to changes in prices than

---

To prove this point, minimize \( x + \rho(x)r \) with respect to \( x \), then take the derivative of the cost-minimizing value of \( x \) with respect to \( r \). By using the second order conditions, it is easy to see that the derivative is positive, proving that the cost minimizing value of \( x \) falls when liability \( r \) diminishes. See infra Appendix, pp. 1554–61.

In order for a change in the sanction to change the actor’s behavior in Figure 3, the sanction would have to fall at least the full distance of the jump or discontinuity in the cost function at \( x^* \), as occurred with sanction \( D \).
to changes in levels of sanctions. On the other hand, an actor's precaution is responsive to a change in the obligation backed by a sanction, which corresponds to a change in the legal standard $x^*$ under Rule $A$.

2. Exceptional Actors. — Figure 3 represents the typical decisionmaker facing a reasonable obligation backed by a reasonable sanction. In such a case, a fall in the sanction from $A$ to $C$ causes no change in the actor's precaution level. In reality, there are some atypical decisionmakers and some unreasonable laws which require modifying this behavioral theory.

Unlike the typical decisionmaker, some individuals will be in a situation where nonconformity with the legal standard is cheaper than conformity. To illustrate, consider someone whose compliance costs are extraordinarily high, as depicted in Figure 5. The effect of exceptional compliance costs is to raise the straight-line portion of the graph, representing costs in the zone of compliance, relative to the curved-line portion of the graph, representing costs of noncompliance. Thus, when the penalty for noncompliance is modest in Figure 5, the lowest point on the graph is below

**Figure 5**

**The Actor with Exceptional Compliance Costs**

16. For example, drivers who ordinarily obey traffic laws may speed when they are in a big hurry. In this example, the cost of compliance, or the benefit from noncompliance, depends upon the value of the driver's time, which in turn depends upon his personal circumstances.
so that complying with the legal standard is more costly than not complying. Deterring an unusual person like the one depicted in Figure 5 requires a severe sanction.

In the usual case, where a reasonable obligation is backed by a reasonable sanction, most people will find conforming strongly advantageous, as described in Figure 3, and a few people will find nonconformity advantageous, as illustrated in Figure 5. Very few people will find themselves on the margin between compliance and noncompliance, where conforming to the legal standard costs the same as not conforming. However, those on the margin are the only ones whose behavior changes when the sanction changes by a small amount. If the sanction is reduced or applied less frequently, then the people on the margin will tip in the direction of not conforming.

Although few people tip, those who do will probably change their behavior a lot, rather than a little. The change in behavior will be substantial because the choice is between taking enough precaution to avoid the sanction entirely, or paying the sanction and taking little or no precaution. Cost minimizing behavior is either equal to the legal standard $x^*$ or significantly short of $x^*$, but it is never a little short of $x^*$. Once it pays to fall short of the legal standard, it pays to fall significantly short of it.

Thus, a small change in a sanction usually causes a small number of people to change their behavior a lot. In contrast, pricing the behavior causes most people to balance benefits and costs at the margin. Since many people would then be on the margin, a small change in a price will cause many people to change their behavior a little. In aggregate, then, behavior is more elastic with respect to prices than sanctions.

---

17. I am assuming that a reasonable law is one for which the tipping point with respect to compliance is in the tail of the distribution where the density is low. It is possible that the tipping point for a law could be in the center of the distribution where it is dense, although many people would consider such a law unreasonable.

18. More technically, the response to a change in sanctions is on the extensive margin, not the intensive margin. In economic jargon, the margin for prices is intensive, involving many people, and the margin for sanctions is extensive, involving a few people.

19. This result is a consequence of the nonconvexity in the cost function. The fact that the change is substantial can be seen by examining Figure 3. When the sanction falls to the low level $D$, the cost minimizing level of behavior jumps substantially below $x^*$. In fact, the lowest point on any cost curve with a discontinuity at $x^*$ will be located at some distance from $x^*$.

20. A similar argument applies when the obligation or sanction is unreasonable, in which case the law may be disobeyed by many people. Regardless of the extent of disobedience, it is likely that only a few people will be on the margin between obedience and disobedience, assuming that compliance costs are distributed smoothly, rather than bunched at the exact level where incentives are equipoised. See infra Appendix, pp. 1554–61. Only those people will respond to a modest change in the sanction. Thus, the conclusion about the elasticity of behavior with respect to a sanction is unchanged by the unreasonableness of the law.
II. A Normative Theory of Lawmaking

The difference in the incentive effects of prices and sanctions is important for lawmakers deciding whether to price or sanction behavior. If officials always possessed perfect information, socially desirable behavior could be induced by either prices or sanctions. Officials could either charge the price which exactly internalizes costs—which would induce individuals to choose the socially efficient level of behavior \( x' \)—or officials could create a legal standard \( x^* = x' \) and back that standard with a sanction strong enough to induce conformity. Regardless of the approach, individuals would respond by choosing the efficient level of precaution.

In reality, however, lawmakers and officials who administer the law often make mistakes because they lack information or the incentives to use information. The behavioral consequences of mistakes are different depending upon whether the law creates a sanction or a price. A normative theory for choosing between sanctions and prices can be based upon the propensity of lawmakers to make mistakes, which depends in part upon the cost of information to them.

Most people conform to a reasonable obligation backed by a reasonable sanction, even if the legal standard is inefficient or otherwise undesirable. Consequently, lawmakers who create an obligation backed by a sanction must be certain that the partition between permitted and forbidden zones is in the right place. On the other hand, mistakes in computing the level of the sanction or the frequency of its application are not crucial, because most people will conform in spite of these mistakes.  

In contrast, pricing behavior does not require dividing action into permitted and forbidden zones, so lawmakers need not compute the socially optimal behavior. Instead, the lawmakers must choose the price accurately. For efficiency, the price must fully reflect the external harm caused by the behavior. Since individuals are responsive to the magnitude of the price and the frequency of its collection, accuracy is crucial to induce behavior that is efficient or otherwise desirable.

This argument can be restated in algebraic terms. If lawmakers can identify the efficient precaution level, \( x' \), then the legal standard, \( x^* \), can be equated with it. Errors in computing or assessing the sanction will not have a large influence on behavior. In contrast, if lawmakers can obtain accurate

21. Professor Wittman made a similar point, saying "[t]here would be low administrative costs if \( X \) [the decisionmaker] were charged only for acting inefficiently, a type of behavior rare in comparison to acting efficiently." Wittman, Liability for Harm or Restitution for Benefit, 13 J. Legal Stud. 57, 61 (1984). When private persons are uncertain about the legal standard or the evidence regarding their conformity that could be offered in court, the discontinuity in Figure 2 is smoothed. However, the cost-minimizing point will not be very different, nor will the decisionmakers' response to errors in enforcement or damages, after the curve is smooth. Smoothing the curve tends to result in overprecaution relative to the legal standard. In technical terms, uncertainty transforms the discontinuity into a nonconvexity. For a discussion of this point see Cooter, Note: Why Comparative Negligence is Efficient (1984) (unpublished manuscript, on file at the office of the Columbia Law Review).
information on external costs, then behavior can be priced at its external cost. However, behavior is more elastic with respect to prices than sanctions, so using prices to create efficient incentives requires accurate assessment of external costs.

These observations lead to a simple decisional rule for lawmakers: **If obtaining accurate information about external costs is cheaper for officials than obtaining accurate information about socially optimal behavior, then they should control the activity by pricing it; if the converse is true, then they should control the activity by sanctioning it.**

It may seem paradoxical that lawmakers could have information about socially optimal behavior and not have information about external costs: isn’t the latter needed to compute the former? This apparent paradox is resolved by the existence of community standards. A community standard represents a consensus among private individuals about socially optimal behavior. In many circumstances, government officials can observe the community standard, but not the costs and benefits which private individuals took into account when arriving at it. If the community’s judgment is sound, then officials can achieve optimality by adopting the community standard as the legal standard and attaching a sanction to it, even though officials lack information about external costs.

To illustrate, suppose that everyone in a community is exposed to the risk of accidents of a certain type, and everyone can take precautions against them. Over time, the community may develop a negligence standard reflecting its judgment about the balancing of costs and benefits of precaution. If an accident occurs and the victim sues, then a court can accurately identify the community standard of negligence by hearing testimony from third parties. However, the victim may exaggerate the seriousness of his injury in order to increase his recovery, thus preventing the court from accurately assessing the cost.

The preceding example concerns a standard of the general public. Specialized communities may also have standards. For example, an accountant who plans to work in a town for many years will have a stronger incentive to maintain a good reputation by avoiding mistakes than an accountant whose aim is to turn a quick profit and leave town. This is an application of the well known result in game theory that cheating in one-shot games can be reduced or eliminated by replacing the one-shot game

22. This proposition is similar to the claim in the pollution literature that quantitative restrictions are preferable to pollution taxes when authorities know the optimal quantity of pollution but do not know the tax price which would achieve it. It has been suggested that this situation might occur when, for example, smog reaches an emergency level. Conversely, tax prices should be used when authorities know the external cost of pollution, but not the optimal quantity. See W. Baumol & W. Oates, The Theory of Environmental Policy (1975); see also Weitzman, Prices v. Quantities, 41 Rev. Econ. Stud. 477 (1974) (developing a model to determine the comparative advantages of indirect market control by prices versus direct regulation of quantities).
with a sequence of repeated games.\textsuperscript{23} The practices of the accountant of good repute may be taken as evidence of the standards which the court should apply to all accountants. The importance of reputation may overcome market failures which would otherwise prevent accountants from adopting the most efficient standards. Thus, a community standard will be efficient in general when it is based on the long run need for its members to acquire the reputation for taking precautions whose costs are commensurate with the resulting reduction in accident costs.\textsuperscript{24}

In the first example of a community standard, efficiency arises because of symmetry: each person could be either the victim or the injurer in an everyday accident. Symmetry facilitates a consensus regarding the appropriate level of care because each person must consider both the costs and benefits of precaution. In the second example, a specialized community achieves an efficient community standard because the importance of reputation overcomes a potential market failure. In either case, a court may be able to decide that the community standard is optimal by examining the institutional structure which gave rise to it.\textsuperscript{25}

In some circumstances, however, lawmakers may be unable to observe

\textsuperscript{23} See, e.g., R. Radner, Collusion Behavior in Non-Cooperative Epsilon-Equilibria of Oligopolies with Long but Finite Lives (Mar. 1979) (Bell Tel. Laboratory Discussion Paper).

\textsuperscript{24} There are many different circumstances in which such a norm will reflect the balancing of benefits and costs. A full analysis of these circumstances would involve examining the conditions under which people act from motives other than self-interest. To be consistent with its economic foundations, this Article is limited to examining a mechanism by which rational self-interest will yield an efficient community standard.

Another illustration is provided by tests for glaucoma performed by ophthalmologists. Since glaucoma is common in old people and rare in young people, there is an age—say forty—before which the cost of routine testing exceeds the expected benefit. If these facts are known and appreciated by ophthalmologists and by informed members of the public, then an ophthalmologist who wanted to cultivate a good reputation would give the test routinely to patients whose age is over forty, and not give it routinely to patients under forty.

This example is drawn from the famous case of Helling v. Carey, 83 Wash. 2d 514, 519 P.2d 981 (1974). Curiously, the plaintiff accepted the defendant's claim that ophthalmologists in Washington at the time were not giving glaucoma tests routinely. A subsequent survey, however, revealed that more than half of ophthalmologists were doing so. See Wiley, The Impact of Judicial Decisions on Professional Conduct: An Empirical Study, 55 S. Cal. L. Rev. 345, 383 (1981). The community's balancing of costs and benefits suggested that the test should have been given routinely. Thus, it appears that the court could have awarded damages on negligence grounds, using the usual community standard, rather than resorting to a new standard of strict liability.

\textsuperscript{25} I have suggested two structural conditions which give rise to efficient community standards, specifically symmetry and repeated games. A complete analysis of the structural conditions for efficient community standards is beyond the scope of this Article, but an analogy may help the reader to appreciate the usefulness of community standards absent such a unified theory. It is well known that monopoly is inefficient, which is one reason why monopoly is prohibited. To guide the antitrust laws, economists have developed a detailed theory based upon the structure of the market, which bypasses questions concerning the conduct and performance of individual firms. Like the antitrust laws, courts can sometimes decide whether community standards are efficient by examining the structure which creates them, rather than directly evaluating their efficiency.
the community standard, or there may be a special interest standard instead of a community standard, or the problem may be so novel that no standard has arisen. Without a community standard to guide them, officials who attempt to control behavior through sanctions must assess the socially optimal level of behavior themselves. This assessment requires information about benefits and costs that may be difficult to obtain. If officials obtain the necessary information and create an optimal legal standard, then private persons will be required to obey the legal standard upon pain of suffering the sanction attached to it. Private persons will not need to assess the socially optimal behavior—the officials will have done that for them. An advantage of an obligation backed by a sanction, as opposed to a price, is that the computation is made once by officials rather than many times by individuals, thus reducing the costs of decisionmaking.

For example, suppose that scientists prove an ingredient widely used in cooking is carcinogenic. The legislature, using the scientists’ results, balances benefits against costs and determines that the ingredient should never be consumed by humans. If the legislature outlaws use of the ingredient, private individuals will be required to obey the law and abstain from using the ingredient. If the ingredient were not banned, private individuals would have to decide whether to use it. To decide, they would have to collect information and replicate some aspects of the legislature’s decision process. Thus, banning the ingredient by legislative fiat reduces the cost of decisionmaking for the community as a whole.

In reality, such decisions are difficult for officials to make, especially when the affected individuals are heterogeneous. To illustrate, if pollution has many sources and many victims, it will be excessive in the absence of government intervention. If officials try to prescribe efficient pollution standards, they must know the abatement costs and the external harm for every source of pollution. It is impossible for officials to possess that much information, as the recent history of the Environmental Protection Agency has demonstrated. Obligations backed by sanctions will not succeed unless individual differences in compliance costs are comparatively small or receive no social weight.

When individual differences in compliance costs are important, the amount of information officials need can be reduced by relying upon prices instead of sanctions. Assigning the correct price to an activity only requires officials to compute the external cost, as opposed to balancing the costs and benefits. For example, if government creates pollution rights and encourages their exchange in a market, polluters will abate until the abatement

26. This assumes that the existence of many polluters and pollutees will prevent a solution through private negotiation.

27. The section on criminal law, see infra notes 54-60 and accompanying text, discusses illegitimate benefits or costs which receive no social weight. For example, the pleasure that some violent criminals apparently enjoy is not counted as a reason for allowing their crimes.

28. The Supreme Court recently held that the “bubble concept” definition of “source” was permissible under the Clean Air Act Amendments of 1977, Pub. L. No. 95-95, 91 Stat.
cost at the margin equals the price of pollution rights. Thus, the market in pollution rights will reveal the marginal abatement cost of polluters without officials having to know anything about the technology of abatement.

Once the market reveals marginal abatement costs, officials can compare these costs to their computation of the external harm from pollution and adjust the amount of pollution until it is efficient. If the external harm caused by pollution exceeds the marginal abatement cost, then efficiency requires further abatement, so government should buy back some pollution rights to reduce their number. If the relative magnitudes are reversed, government should create and distribute additional pollution rights.\(^{29}\) These observations lead to the conventional conclusion that controlling pollution through prices requires less information than controlling it through obligations backed by sanctions.\(^{30}\)

Table 1 summarizes the argument about information and policy making. If private persons are good observers of social costs and benefits, as indicated by the lower right cell of the table, then a community standard may arise reflecting a consensus about efficient behavior. Optimal incentives can be created by giving official sanction to such standards. If officials can observe benefits and costs more accurately than private persons, as indicated by the upper left cell of the table, then optimal incentives can be created by legislative fiat. In contrast, if private persons can assess the benefits from their own acts and officials can assess the external costs, as indicated by the lower left cell in the table, then optimal incentives can be created by pricing the externality.\(^{31}\) Finally, turning to the upper right cell, if private persons can assess the costs of their own acts and officials can assess the external benefits, then optimal incentives can be created by official subsidies.

---


29. Pollution taxes, rather than a market in pollution rights, can achieve the same result while providing revenues to the government. Assuming transaction costs are minimal, the only difference between a pollution tax and a government auction of pollution rights is that the government chooses the price and the market chooses the quantity with taxes, while the opposite is true with an auction. See Cooter, The Cost of Coase, 11 J. Legal Stud. 1 (1982). Both approaches raise revenues for the government by correcting distortions in the allocation of resources rather than by creating such distortions. The revenues raised by nondistorting taxes can be substituted for the revenues raised by distorting taxes like the income tax, thus increasing efficiency. On the other hand, the government can allocate ownership of pollution rights to individuals, rather than operate a government auction, in which case the government will not derive revenues from pollution control.

30. This is the familiar Pigouvian tax solution to pollution. See W. Baumol & W. Oates, supra note 22.

31. As explained in the discussion of markets for pollution rights, supra notes 27–29 and accompanying text, private actors' responses to the first round of prices will establish the compliance costs and the government can accordingly reset the price to achieve the socially desired level of activity.
TABLE 1

Distribution of Information and the Optimal Policy
(Assuming Market Failure and External Costs)

<table>
<thead>
<tr>
<th>Best Observer of Social Costs</th>
<th>Officials</th>
<th>Private Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best legislative official</td>
<td>legislative standard and sanction</td>
<td>official subsidy</td>
</tr>
<tr>
<td>Observer</td>
<td>Private official price</td>
<td>community standard and sanction</td>
</tr>
</tbody>
</table>

III. DISTINGUISHING PRICES FROM SANCTIONS

The prescription for lawmakers developed in Part II can be tested to see whether it is being followed in various bodies of law. First, however, it is necessary to be able to distinguish prices from sanctions. Sometimes a rule clearly resembles Rule A, in which case it is an obligation backed by a sanction, and sometimes it clearly resembles Rule B, in which case it imposes a price. However, some rules appear opaque upon first examination, so it is useful to have a criterion for preliminary classification.

A criterion can be developed by recalling that sanctions attach to forbidden acts and prices attach to permitted acts. Since the purpose of a sanction is to deter people from wrongdoing,32 the sanction will be adjusted to achieve this goal. Deterring actors whose fault is intentional, deliberate, or repeated requires a more severe sanction than deterring actors whose fault is unintentional, spontaneous, or committed for the first time. Therefore, sanctions increase with certain mental qualities of the act indicating more resistance to deterrence.

The efficient price depends upon the extent of external harm, not the actor's state of mind. If, contrary to fact, prices varied with the actor's state of mind—making the price higher if the act were done intentionally—then people would be deterred from doing the very acts that are permitted. Since a typical purpose of prices is to internalize costs, and since the external cost of an act is unrelated to the actor's state of mind, a price should not increase just because the activity is intentional, willful, or repeated.

These observations point to a simple test for deciding whether a law creates a sanction or a price: Sanctions increase with the need for deterrence, as indicated by the actor's state of mind, whereas prices increase with the amount of external harm.

32. Deterrence is usually regarded as one motive for imposing a sanction. W. LaFave & A. Scott, Jr., Handbook on Criminal Law 23 (1972). Other motives, including punishment, are described in H.L.A. Hart, Punishment and Responsibility (1968).
nal harm caused by the act, which is invariant with respect to the actor’s state of mind.33

IV. RELATION TO VARIOUS AREAS OF LAW

Part II developed a prescription for lawmakers who must choose between prices and sanctions, and Part III developed a criterion for preliminary classification of laws as prices or sanctions. This Part distinguishes prices from sanctions in several bodies of law—specifically, tort, contract, criminal law, and regulation—to determine whether Part II’s prescription is followed.

A. Torts: Strict Liability v. Negligence

Negligence and strict liability rules in tort law present a stark contrast between sanctions and prices. For the sake of concreteness, imagine a manufacturing process in which some defective products inevitably slip through quality control and injure consumers. Precaution against these accidents is costly. To keep the example simple, assume that effective precaution is possible for the manufacturer and impossible for consumers. Social costs are the sum of the costs of the manufacturer’s precaution and the cost of injuries caused by defective products. Thus Figure 1 can be interpreted as a graph relating the social cost of manufacturing defects to precaution. As explained earlier, social costs are minimized at the point $x^*$ in the graph, which is the efficient point.

1. Incentives. — A negligence rule partitions the world into permitted and forbidden zones, as illustrated in Figure 2, and the sanction for being in the forbidden zone is liability for the harm caused by accidents. If the manufacturer’s precaution is below the legal standard $x^*$, then he is liable for accidents; if his precaution is at least as great as the legal standard $x^*$, then he is not liable. If the legal standard, $x^*$, equals the efficient level of precaution, $x^*$; a manufacturer whose precaution is inefficient is liable, and a manufacturer whose precaution is efficient is not liable. Thus, a negligence rule corresponds to Rule A—representation of an obligation backed by a sanction.

   To understand the incentive effects of a negligence rule, consider the private costs of the manufacturer. If his precaution $x$ is below the legal standard $x^*$, then he bears the cost of precaution and liability for accidents. If liability covers the full cost of the harm suffered by the victim, the injurer’s private costs will equal the social cost in the zone of liability, as depicted in Figure 2. However, when the manufacturer’s precaution satisfies

---

33. This proposition is not a statement of necessary or sufficient conditions. Other criteria could be mentioned, such as: Is shame attached to the activity? Can an injunction be obtained to prevent its recurrence? Is the cost deductible from the actor’s tax liability? These ancillary criteria have some value, but this test proves especially useful in classifying laws.
the legal standard, he escapes liability under a negligence rule, so he only bears the cost of his own precaution in the zone of nonliability.

There is a jump, or discontinuity, in the manufacturer’s costs at the legal standard of care because his liability plummets from one hundred percent to zero percent at the point where his precaution satisfies the legal standard.\(^{34}\) The size of the jump is determined by the amount of damage caused by defective products when precaution exactly satisfies the legal standard.\(^{35}\) As a consequence of the jump, it is far cheaper for the injurer to be nonnegligent than to be negligent.\(^{36}\)

Since it is far cheaper to be nonnegligent than negligent, changes in the legal standard will induce equivalent changes in the manufacturer’s precaution. The manufacturer will satisfy the legal standard, \(x^*\), even if it is pegged at a level above or below the socially efficient level, \(x'\). On the other hand, the manufacturer’s precaution will not change if there is a moderate change in the sanction or the probability of its application.

The situation is quite different if a negligence rule is replaced by strict liability. If the manufacturer is strictly liable for the actual harm caused by defective products, he will bear the social cost of accidents. When social costs are borne by the manufacturer, minimizing his private costs is equivalent to minimizing social costs. Figure 1 can be interpreted as a graph of the manufacturer’s private costs under strict liability, which are minimized at the efficient precaution level \(x'\). Since there is no jump in the private cost function of the manufacturer under strict liability, the manufacturer minimizes private costs at a level of precaution where marginal benefit equals marginal cost.

Unlike negligence then, strict liability makes the manufacturer very responsive to imperfections in assessing and computing damages. If the courts sometimes allow a manufacturer to escape liability for harm caused by a defective product or award damages less than the harm caused by the

\(^{34}\) There will be a discontinuity in the cost curve whether a traditional negligence standard or a comparative negligence standard is applied. The curves will be identical when the plaintiff is entirely without fault. Where the plaintiff has been negligent, the size of the discontinuity in the cost curve will be reduced under comparative negligence but will still exist.

\(^{35}\) It is sometimes said that a person who is negligent is only liable for the harm that would have been avoided by nonnegligent care. This statement is misleading for many types of accidents, because nonnegligence just reduces the probability of an accident, rather than eliminating its possibility. To illustrate with a classic example, suppose that railroad trains emit sparks which sometimes set fire to farmers’ fields. Nonnegligent operation of the train requires the use of spark arresters, which reduce, but do not eliminate, spark emissions. Under a negligence standard, a railroad using spark arresters would escape liability from fires caused by its trains, whereas a railroad failing to use spark arresters would be liable for fires caused by its trains. The point to notice is that failing to use spark arresters increases the probability of fires, but makes the railroad liable for all fires, including the fires which spark arresters would not have eliminated. This fact gives rise to the discontinuity in the railroad’s cost function.

\(^{36}\) In economic jargon, the level of care which minimizes the manufacturer’s private net costs is a corner solution where marginal private costs of a change far exceed marginal private benefits.
product, or if some consumers injured by defective products do not bring suit, then it will be profitable for the manufacturer to reduce his precaution below the efficient level. This conclusion is illustrated in Figure 4, where shifting the private cost curve below the social cost curve causes the cost minimizing level of precaution to fall.

2. Prescription for Lawmakers. — Since a negligence rule makes injurers responsive to the legal standard of care, and unresponsive to modest errors in computing damages or assigning liability, a court operating under a negligence rule must compute the legal standard with accuracy, but it can make modest errors in computing damages and assigning liability. In contrast, since a rule of strict liability makes injurers responsive to errors in computing damages or assigning liability, a court operating under strict liability must compute damages and assign liability with accuracy. Thus, a negligence rule should prevail when officials make relatively small errors in identifying the efficient level of care and relatively large errors in computing damages or assigning liability; and a rule of strict liability should prevail where officials make relatively large errors in identifying the efficient level of care and relatively small errors in computing damages or assigning liability.

To see whether this prescription is being followed, it is necessary to examine the frequency and magnitude of three types of court-made errors in tort disputes: (1) imperfections in setting the legal standard, (2) imperfections in determining liability, and (3) imperfections in setting damages.

a. Imperfection in Setting the Legal Standard. — To achieve efficiency by a negligence rule, a court must be able to determine the efficient level of precaution. If there is a community standard governing the activity, and the community has balanced incremental costs and benefits in arriving at the standard, then the court can achieve efficiency by enforcing the community standard. In many cases the community will have done so. The public experiences a broad class of everyday dangers such as household or automobile accidents. Since most people maintain a household and drive a car, the public has experience with the cost of precaution, as well as the harm caused by such accidents. Thus it has enough information to balance benefits and costs in deciding the appropriate standard of behavior required to be a safe driver or to properly maintain a house.

In court, testimony about the usual practice of community members with good reputations is accepted as evidence about the community standard.37 A good reputation is acquired by proper behavior over a long period of time. Proper behavior may require giving similar weight to the cost of one's own precaution and to the resulting reduction in risk for others, as required for efficiency. It may also require taking the long run view of costs and benefits. To illustrate, recall the example of an accountant planning to

work in a town for many years, who will have a stronger incentive to maintain a good reputation by avoiding mistakes than an accountant whose aim is to turn a quick profit and leave town. The practices of the accountant of good reputation may be taken as evidence of the standards a court should apply to all accountants.

The community referred to in “community standard” is different for different activities. For example, in the case of driving, “community” refers to the general public; in the case of accounting, the term refers to a special group with its own private interests. The general public has extensive experience with driving and little experience with auditing. Naturally enough, the public is suspicious about whether specialized groups acting on their own will develop standards of precaution commensurate with the resulting savings in accident costs enjoyed by the public. In the specific case of defective products, concern with reputation may be insufficient to cause firms to internalize the full benefits of precaution, especially if consumers are unaware of the true risk of accidents. Thus the courts cannot automatically assume that the standard of care spontaneously adopted by specialized groups will be efficient.38

If the market failure prevents an efficient community standard from arising and the courts nonetheless decide to adopt a negligence rule, officials must compute the optimal standard directly or proceed by successive approximations. To proceed by approximation, the court can adopt the following rule: If the incremental benefits of precaution exceed the defendant’s incremental costs, then precaution is held to be insufficient and the defendant is liable, whereas if the relative magnitudes are reversed, then the precaution is held to be sufficient and the defendant is not liable. A sequence of such determinations, each by courts with partial information, will eventually reveal the exact level of care that is efficient.39 Proceeding by approximation requires less information than making the calculation directly, but the information requirements are high in either case. Reaching a satisfactory and consistent standard by this method is so difficult that lawmakers would usually do better to adopt a rule of strict liability.

In reality, courts generally enforce community standards and are reluctant to impose novel obligations; where there is no satisfactory community standard, strict liability rules are often employed. This claim has been illustrated by contrasting tort liability for everyday accidents with tort liability for injuries to consumers. A reasonably efficient community standard spontaneously arises for everyday accidents and courts enforce it, thus conforming to the prescription outlined in Part II. If market failure prevents an efficient standard from arising for consumer products, then the adoption of a rule of strict liability for consumer injuries conforms to the prescription.

38. The T.J. Hooper, 60 F.2d 737, 740 (2d Cir.) (L. Hand, J.) (refusing to adopt industry standard of care), cert. denied, 287 U.S. 662 (1932).
b. Imperfections in Determining Liability. — A second source of error by officials is imperfections in determining liability. Rational accident victims will bring suit when the expected court award exceeds the plaintiff's cost of trial. A negligence rule imposes a difficult burden of proof upon the injured consumer. The injury has to be severe and the stakes large before the expected damage award exceeds the cost of a suit. The low probability of successfully proving negligence prevents many injured consumers from suing. Thus there is an imperfection in the enforcement of a negligence rule for defective products.

This imperfection is often cited as a reason for abandoning a negligence rule, but close inspection reveals some flaws in the conventional reasoning. Part I showed that precaution is inelastic with respect to the level of a sanction. It follows that imperfections in enforcing a negligence rule against manufacturers would probably not have much effect upon their behavior. Conforming to the legal standard would be cheaper than breaking it for most firms, in spite of imperfect enforcement. By contrast, mistakes in assigning liability would have a large effect upon the behavior of firms under a rule of strict liability. Fortunately, under strict liability the consumer's burden of proof is relatively light, so few injurers escape liability.

This theory suggests that traditional arguments about burden of proof confuse cause and effect. A correct statement is: that proof is hard under negligence does not cause suboptimal precaution; that proof is easy under strict liability is necessary for optimal precaution.

c. Imperfections in Assessing Damages. — A similar argument applies to the third source of error—imperfections in assessing damages. An injury often has aspects that are difficult to prove, as when property with sentimental value is destroyed or a business venture is disrupted before it becomes profitable. Proof is so difficult with subjective and speculative injuries that they are often excluded from damages, causing a bias toward undercompensation.

In recent years the courts have expanded the scope of compensable injury to include formerly excluded emotional and speculative losses. Insofar as this change has eliminated a bias towards undercompensation, it has eliminated a significant problem with a rule of strict liability, making such rules more efficient. By contrast, systematic undercompensation would not have much effect upon injurer's precaution under a negligence rule.

In summary, the prescription that negligence rules should be favored when courts can observe the socially desirable level of precaution more accurately than the external harm, enjoys limited success in explaining the contours of tort law. Community standards for everyday accidents represent legitimate judgments about the socially desirable level of precaution. Courts observe these standards and enforce them through negligence rules for tort liability. However, the standards of some industries are marred by market failures. For example, since courts cannot observe the socially desirable levels of precaution in manufacturing without difficulty, a strict liabil-
ity standard is applied to consumer injuries. Thus, differences in the information available to courts and the errors to which courts are prone, provide part of the explanation for applying strict liability for consumer injuries, and negligence for everyday injuries.

3. Punitive Damages in Tort. — Economists sometimes think of sanctions as prices, which prevents them from understanding the normative character of the law. To illustrate this confusion in the area of tort law, suppose an individual believes that compensatory damages are the price of negligence. When accused of negligence, he might explain to the court that he deliberately took actions proscribed by the legal standard in order to avoid the high cost of compliance, and, in doing so, he was fully prepared to compensate anyone who was injured. The court might reply that the injurer has a duty to take reasonable care, not permission to be careless and pay compensatory damages. The court might add that deliberate violation of the legal standard, as admitted by the injurer, provides grounds for punitive damages. Finally, the court might suggest that punitive damages be set high enough to deter the injurer in the future.

This illustration suggests that punitive damages have the earmarks which distinguish a sanction from a price: (1) they are attached to a forbidden action; (2) they depend upon the actor's state of mind; and (3) their purpose is to deter. The law is murky, however, concerning standards for awarding punitive damages. This Article's behavioral theory can be used to clarify these standards.

According to this theory, most potential injurers will try to conform to standards of reasonable care. For them, nonconformity is accidental. For some exceptional injurers, however, compliance costs or the benefits from noncompliance are unusually high. For these injurers, violating the standard and paying compensatory damages is cheaper than conforming. When deterrence is appropriate, punitive damages can be used to deter such exceptional injurers.

Intent distinguishes the usual injurer and the exceptional injurer, but intent is difficult to observe and prove. Intentional violation of the legal standard exposes the injurer to liability for compensatory and punitive damages. Consequently, his expected costs will jump when his precaution intentionally falls below the legal standard. As illustrated in Figure 5, exceptional injurers will choose a precaution level that is far short of the legal standard, not a little short of it. Thus their fault will be gross. Yet if puni-

40. Sometimes exceptional compliance costs or exceptional benefits from noncompliance are sound reasons for applying a different standard of care to the individual. In such cases it is appropriate to allow the actor to violate the usual legal standard. In other cases, the exceptional compliance costs or exceptional benefits from noncompliance are viewed as illegitimate, and deterrence by assessing punitive damages is appropriate. To illustrate the latter, a driver who loves to take risks may find complying with traffic laws subjectively difficult, yet he will be held to the same driving standards in court as someone who is instinctively cautious.
tive damages are sufficiently severe, even the exceptional injurer will mini-
mize his expected costs by conforming to the legal standard.

According to many commentators, courts have appropriately assessed
punitive damages when fault is gross, repeated, or willful and wanton.41
This Article’s behavioral theory explains the connection between gross fault
and intent. Intentional fault is gross because once the injurer decides to
disobey the legal standard it pays to fall far short of it.

B. Contracts

There is currently a dispute among theorists about whether breach of
contract can be justified merely on the ground of economic efficiency. This
Section argues that this dispute concerns whether damages for contract
breach are sanctions or prices. Moreover, this section demonstrates that the
analytical framework of Parts I and II can be used to resolve the dispute.

1. The General Case. — In America, the standard remedy for breach of
contract is expectation damages.42 The purpose of expectation damages is
to put the victim of breach in as good a position as he would have been in if
the promisor had performed.43 Expectation damages which achieve this
purpose can be called “perfect.”44

If the promisor’s nonperformance creates a benefit for him that exceeds
perfect expectation damages for the promisee, then nonperformance is with
compensation Pareto efficient—that is, it makes at least one person better
off without making anyone worse off. Nonperformance under these circum-
stances creates a surplus equal to the amount by which the promisor’s bene-
fit from nonperformance exceeds perfect expectation damages.

The surplus can be divided between the parties in various ways. The
promisor may offer to give the promisee a share of the surplus in exchange
for a release from obligation. If the offer is accepted, the promisor will not
have to breach the contract. Alternatively, the promisor can breach the
contract and pay damages to the promisee. If the promisor breaches and
pays perfect expectation damages and nothing more, then the promisee will
get none of the surplus and the promisor will get all of it.

To illustrate, suppose that Farmer promises to sell 100 quarts of
strawberries to Retailer for $.75 per quart, which Retailer will resell for
$1.00 per quart, realizing a total profit of $25. Before Farmer delivers to
Retailer, Buyer, who wants all 100 quarts, offers to pay Farmer $1.35 per

41. See Cooter, Economic Analysis of Punitive Damages, 56 S. Cal. L. Rev. 79 (1982);
Ellis, Fairness and Efficiency in the Law of Punitive Damages, 56 S. Cal. L. Rev. 1 (1982);
Landes & Posner, An Economic Theory of Intentional Torts, 1 Int'l Rev. L. & Econ. 127
42. Unfortunately, the courts seem to have lost sight of this fact. See Cooter, supra note
8.
44. See R. Cooter & M. Eisenberg, Damages for Breach of Contract (Sept. 12, 1984)
(unpublished manuscript, on file in the offices of the Columbia Law Review).
PRICES AND SANCTIONS

quart for immediate delivery. If Farmer diverts the strawberries to Buyer, instead of performing for Retailer, as promised, a surplus of $35 over the original plan will result. By diverting the strawberries to Buyer, Farmer can pay $25 in compensation to Retailer and still enjoy $35 in extra profits.

If negotiating a release from the contract is costly, then efficiency requires a promisor like Farmer to breach. However, it is often the case that a promisor like Farmer can negotiate a release from his contract at little cost. In such circumstances, it is equally efficient to negotiate a release or breach with compensation. If the two alternatives are indifferent from an efficiency viewpoint, should the law be indifferent between them? This is the subject of a dispute among theorists. The proponents of the efficient breach hypothesis argue that Farmer can breach, rather than request release from his promise to Retailer. The opposite view holds that it would be wrong for Farmer not to perform unless he has a valid excuse.

A promisor who breaches whenever it is efficient weighs the benefits against the costs, just as a consumer weighs the benefits of a new overcoat against the price. No weight is given to the promise over and above the cost of perfect expectation damages. Thus, the promisor who breaches whenever it is efficient acts as if breach with compensation is permitted, which implies that damages are the price of breach.

On the other side of the dispute are theorists who maintain that a promisor should not breach merely because his benefits exceed his liability for expectation damages. According to this view, a stronger reason than efficiency is required to justify breach. Efficiency is excluded because the promisor allegedly has an obligation to perform, not permission to breach with compensation. According to this view, the promise breaker does what is forbidden, even though he compensates his victim. Thus, the view that efficient breach is forbidden implies that damages for breach are a sanction.

The dispute over efficient breach can be described as a disagreement about whether damages are a sanction for wrongdoing or the price of breach. This Article suggests a simple criterion for a preliminary resolution of such disputes: Do damages vary with the state of mind of the injurer, or do they vary only with the extent of harm suffered by the victim? Facts about the promise breaker's state of mind do not usually influence the court's assessment of expectation damages. Thus, according to the preliminary criterion, damages for breach of contract are a price, not a sanction.

This conclusion is reinforced by considering punitive damages. The victim of breach cannot ordinarily increase the damage award simply by


46. Some justifications for breach are the traditional defenses, which enable the promise breaker to escape liability, such as impossibility, changed circumstances, or mutual mistake. Other circumstances may justify breach with compensation, such as a situation in which performance would impose extreme hardship on the promisor and produce only a small advantage to the promisee. The list is open ended, but the opponents of efficient breach insist that efficiency is excluded from it.
showing that the promise breaker's conduct was gross, repeated, or willful and wanton.\textsuperscript{47} For punitive damages to be awarded, the contract must involve a special relationship, such as a fiduciary relationship.\textsuperscript{48}

If expectation damages are perfect in the sense of compensating exactly for the lost expectation, then it can be shown by using economic analysis that the cost of breach is internalized, as required for efficiency.\textsuperscript{49} If damages were increased over the expectation level, they would deter breach rather than internalize costs. Damages are rarely increased above the compensatory level, so they usually internalize the cost of breach, rather than deter it, which supports the conclusion that damages are the price of breach.

According to this Article's prescription, lawmakers ought to price an activity when it is easy for them to determine its external costs. In the case of contracts, the promisor makes a definite commitment—for example, to pay a certain sum of money on a particular date, to deliver a specific good, or to provide a specific service. Usually the performance has economic value which is much easier to quantify than, for example, the cost of a bodily injury in a tort suit, so it is relatively easy for the court to determine the harm caused by breach.

The prescription further suggests that lawmakers ought to sanction an activity when they can easily determine its optimal level. Because there are many legitimate circumstances which can interfere with the promisor's ability to perform, or increase the cost of performance, lawmakers would be hard-pressed to determine generally the optimal level of performance or the optimal level of precaution against breach.

To illustrate, suppose that a construction company fails to complete a project on time. The failure may be due to inclement weather, recalcitrant building inspectors, a wildcat strike by plumbers, or a shortage of materials. If damages were a sanction for wrongdoing, then the court would have to determine whether the construction company was at fault. It would be very difficult for the court to obtain information on all of the relevant facts and weigh them. Since it is much easier for courts to determine the external harm caused by breach of contract than to determine whether the promisor's behavior was faulty, the prescription leads to the conclusion that breach should be priced rather than sanctioned.

2. \textit{The Exceptional Cases}. — While the broad contours of contract law correspond to the theory advocated here, there are details that pose difficul-

\begin{itemize}
\item \textsuperscript{47} Insurance contracts are also an exception. See Sullivan, Punitive Damages in the Law of Contract: The Reality and the Illusion of Legal Change, 61 Minn. L. Rev. 207, 241-44 (1976-1977); Comment, Punitive Damages on Ordinary Contracts, 42 Mont. L. Rev. 93, 100 (1981). In these cases punitive damages are a sanction. Note also that compensation for partial performance under the quantum meruit principle may be blocked if the breach is willful. For a discussion of this point, see L. Fuller & M. Eisenberg, Basic Contract Law 316 (3rd ed. 1972).
\item \textsuperscript{48} E. Farnsworth, supra note 43, \S\ 12.8 (1982).
\item \textsuperscript{49} See R. Cooter & M. Eisenberg, supra note 44.
\end{itemize}
ties. First, compensatory damages are the usual remedy for breach, but not the unique remedy. Courts award damages which exceed the loss suffered by the victim of breach in cases involving unjust enrichment. For example, if an art dealer promises to sell a painting to the first buyer, and then sells the painting to the second buyer at a higher price, the art dealer may have to "disgorge" his profit by paying damages equal to the price differential. This is a situation in which a penalty is added to compensatory damages in order to offset the exceptional advantage gained by nonperformance. In cases involving unjust enrichment, the court imposes a sanction for wrong-doing in order to deter it.

Another remedy for breach is specific performance. A leading theory holds that courts should order specific performance when breach involves a failure to deliver a unique good. This theory is consistent with the prescription that courts should sanction an activity when they can determine the efficient social standard with more accuracy than they can determine the external harm. In cases where the good is unique, the courts cannot accurately assess its value to the victim of breach. The buyer of a unique good enjoys a surplus equal to the amount by which his willingness to pay for the good exceeds what he actually pays. The consumer's willingness to pay is impossible to assess for unique goods. Since courts cannot observe the size of the surplus, the remedy for seller's breach is to order performance, backed by a sanction, rather than to award money damages.

Finally, standards similar to negligence are sometimes relevant in deciding whether or not a contract exists. Specifically, a person is liable for the careless use of the symbols by which contracts are formed. To illustrate, a person who raises his hand at an auction to point out an attractive light fixture on the ceiling may find that he has inadvertently made a purchase. If a person does not intend to make a promise, but uses the symbols of contracting in such a way that a reasonable person would believe a promise was made, then the courts may enforce it. Enforcing the contract is sometimes the sanction for careless use of the symbols of contracting.

50. There are also nonlegal remedies for breach. Business practice may correspond imperfectly to the law with regard to whether damages are a sanction or a price. In the usual case of breach, where compensatory damages are the remedy, the court proceeds as if breach were priced, rather than sanctioned. However, social custom may condemn a person whose breach is gross, repeated, or willful and wanton, even though courts do not pronounce on the question of fault. Such a loss of reputation can be costly to a businessman.

51. Deterrence is imperfect because disgorging profits eliminates the actual gain, but there is still an expected gain whenever there is a positive probability that the wrongdoing will go undetected. This is analogous to requiring a tax evader to pay the IRS what he owes, but attaching no additional fine to his liability.


53. This is an example of the well known problem of preference revelation, first identified with clarity by Professor Samuelson. See Samuelson, The Pure Theory of Public Expenditure, 36 Rev. Econ. & Stat. 387 (1954).
C. Criminal Law

The incentive effects of sanctions and prices, as represented by the contrast between Rules $A$ and $B$, assume continuous behavior whose optimal amount is positive for some people. Some crimes fit this model. For example, vehicular manslaughter can be analyzed along the same lines as tortious negligence. However, other crimes cannot fit this model without being distorted. For example, premeditated murder is a discontinuous act whose optimal frequency is zero. In this section the model is adapted to fit such behavior. The substantive question addressed is, "Why should crimes be punished rather than priced?"

This Article's prescription about when lawmakers should sanction behavior, rather than price it, provides an immediate answer to this question.

Figure 2 can be modified to represent behavior whose optimal amount is nil. As illustrated in Figure 6, the horizontal axis has been changed to indicate the seriousness of the crime and $x^*$ has been moved to the origin to

---

54. A theorist determined to construe choice as continuous might argue that the probability of killing someone by a particular action is continuous, and it is optimal—whatever that means—for some people to be murdered. This line of argument strains the connection between my model and ordinary speech to the breaking point. Instead, I attempt a more immediate and natural modeling in terms of discontinuous choices with optima at corners.

55. This section does not consider the few cases of strict liability for crimes that have arisen in the business context. See, e.g., Morissette v. United States, 342 U.S. 246 (1952).
indicate that the legal standard requires that crimes not be committed. To take a concrete example, suppose that the crime in question is theft and that the seriousness of the theft is proportional to the amount of money stolen. The vertical axis shows the external costs of the crime, which include public and private expenditures on preventing, prosecuting, and punishing it, as well as the amount actually stolen and other consequential losses to the victims. The legal standard, denoted $x^*$, is located at the origin of the graph, which indicates that stealing any positive amount of money, no matter how small, is prohibited.

A person who steals exposes himself to the risk of costly prosecution and loss of reputation, even if he steals very little. Consequently, a person's private expected costs will often jump abruptly as soon as he steals a small amount. Thus, the private expected costs of the thief will have a discontinuity at the origin where he violates the legal standard, as shown in Figure 6. That the jump occurs at the origin in Figure 6, rather than at an internal point, does not change the conclusions reached in discussing Rule A. First, the theory of incentives shows that a typical person minimizes his private costs by conforming to a reasonable obligation backed by a reasonable sanction. Since most criminal laws are reasonable, most people will conform to them. Only an exceptional person finds crime advantageous when account is taken of its risks.

Second, according to the prescription, lawmakers should sanction activities when they have better information about community standards than about external costs. There are clear community standards with regard to many crimes such as robbery, murder, rape, burglary, or embezzlement—the standard being that no one should do them. Thus, the existence of community standards concerning many crimes solves one of the information problems lawmakers face. Community standards condemn many crimes unambiguously, but the

56. The reader who is troubled by the idea that victims can assign a dollar value to the cost of crime should bear in mind that many crimes are also torts, for which the jury must assign a dollar value for compensation. Thus, I am only assuming that a task already assigned to juries can actually be performed.

57. This fact is captured by the title of M. Feeley, The Process Is the Punishment (1979).

58. It is also the case that differences in compliance costs across individuals are unimportant in formulating many aspects of criminal law. Conforming to a legal standard can involve restraining an impulse. For example, a person who is very angry at someone may find it difficult not to batter him. Self-restraint is analogous to compliance costs. When the legal standard must be responsive to differences in compliance costs among individuals, the information requirements for sanctions are excessive, as illustrated by the discussion of pollution. See supra notes 25–29 and accompanying text. On the other hand, the legal standard for battery will not be adjusted according to the strength of the injurer's anger. Many psychological costs of compliance—anger, revenge, envy, or lust—receive little or no weight in court. The law treats restraint of an impulse to do wrong as an illegitimate compliance cost which receives no weight, so the information requirements for setting criminal standards are modest. In contrast, if criminal acts were priced, the actor would be permitted to give weight to these illegitimate costs.
public is often divided about the extent of punishment that should attach to crimes. Legal officials are left with the difficult task of constructing schedules of punishment. Fortunately, if the officials make a mistake and attach the wrong sanction to a crime—either too high or too low—the behavior of most people will not be affected by the error, because the cost of crime far exceeds the benefit. The social cost of moderate errors in calibrating punishments is low because behavior is relatively inelastic with respect to sanctions.59

Conversely, if crimes were priced, rather than sanctioned, people would be permitted to commit crimes provided that they paid the price. With prices, self-interested decisionmakers would balance the incremental benefit against the incremental cost of committing a crime. To internalize costs, the price would have to cover harm suffered by the victims and the enforcement activities of the state, adjusted by the probability of its collection. If benefits and costs are balanced at the margin, then the number of crimes would be relatively elastic with respect to the price, so errors in computing damages would affect the amount of crime. It is very difficult to compute the cost of crime to its victims, so large errors would occur. Thus the information requirements for pricing crimes are prohibitive, while the information requirements for sanctioning crimes are relatively low.60

D. Regulation

Most regulations consist of obligations backed by sanctions. Economists have consistently advocated replacing many regulations with prices. This section examines this policy problem using the analytical framework developed in Parts I and II.

1. The Normative Issue. — An example of the policy question is: Should pollution standards be replaced with pollution taxes? To compute the efficient tax, government officials must know the amount of external harm caused by the polluter and nothing more. By contrast, to discover the

59. This statement should not be interpreted to mean that deterrence is ineffective.
60. The usual reason economists offer for not pricing crime is different from the one offered here. According to the traditional argument, it is inefficient, if not impossible, to punish every crime, so potential criminals will discount the punishment by the probability of its application. If the punishment were the requirement of compensating the victim fully, then the discounted punishment would provide too little deterrence from an efficiency perspective, so there would be too much crime. See, e.g., Becker, Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968). Furthermore, the cost of punishing the convicted offender must be taken into account. See Polinsky & Shavell, The Optimal Trade-off between the Probability and Magnitude of Fines, 69 Am. Econ. Rev. 880 (1979). This argument, however, does not go to the root of the problem, because it does not explain what is wrong with pricing most crimes and punishing those who attempt to evade paying the price. There is a sense in which sanctions lie behind prices—the obligation to pay the price is itself sanctioned. Professors Calabresi and Melamed made this same point in somewhat different language when they explained the difference between property rights and liability rights. See Calabresi & Melamed, Property Rules, Liability Rules, and Inalienability One View of the Cathedral, 85 Harv. L. Rev. 1089 (1972).
efficient standard, officials must balance the external harm against the cost of abatement, which requires complete information on each polluter’s abatement technology. Since knowledge of the technology of each polluter is difficult or impossible for officials to obtain, some economists recommend replacing regulations with taxes.

There is, however, a problem with prices that does not arise with standards. It is in the interest of most people to conform to standards backed by sanctions, even though violations often go unpunished. Specifically, if reasonable pollution standards are backed by reasonable sanctions, most polluters will comply, even though pollution sometimes goes undetected. By contrast, if a price sometimes goes uncollected, then most self-interested people will change their behavior. Polluters will respond to a decrease in the frequency of collection of pollution fees by increasing the amount of pollution. In general, a sanction requires lower enforcement costs than a price, even if the two are equal in magnitude, because precaution is elastic with respect to imperfections in collecting prices and inelastic with respect to imperfections in enforcing sanctions.¹¹

Thus the choice between prices and sanctions in regulating pollution is a choice between the low enforcement costs of obligations backed by sanctions and the low information costs of prices.

2. The Descriptive Issue. — It is especially difficult with regulations to decide after the fact whether they create prices or sanctions. To illustrate, is a parking ticket a price or a sanction? What about a ticket for operating an overloaded truck, or a citation for a health violation in a restaurant? Officials often have discretion about whether to increase the penalty for intentional, repeated, or gross offenses. Whether or not they increase it is a good test of whether they regard it as a sanction or a price.

For example, there does not seem to be any additional penalty attached to repeated parking violations. Most judges are prepared to allow individuals to repeatedly break parking regulations provided that they pay the price. Perhaps lawmakers and judges allow individuals to balance the costs of compliance against the benefits of noncompliance because the legal officials do not have good information about these benefits. Apparently, judges treat parking tickets as if their purpose were to internalize the cost of parking congestion.

In contrast, most judges take a different attitude with respect to moving violations such as speeding or drunken driving. With these traffic regulations, the penalty increases with repetition of the offense, indicating that officials regard the fine as a sanction. Sanctions for moving violations probably make sense in light of the fact that obtaining information about community standards with regard to automobile safety is easier than determining the cost of automobile accidents.

¹¹ This assumes that the distribution of compliance costs is such that the elasticity of precaution on the intensive margin, which is relevant for prices, is greater than the elasticity on the extensive margin, which is relevant for sanctions.
Conclusion

A sanction is a detriment imposed for doing what is forbidden, and a price is an amount of money exacted for doing what is permitted. An obligation partitions acts into permitted and forbidden zones, with the partition representing a social judgment about correct behavior. If a penalty is attached to forbidden activities, there is a jump or discontinuity in the private costs of the actor at the point of the partition. Consequently, the actor’s behavior will be unresponsive to moderate changes in the penalty or the frequency of its application. By contrast, prices allow the individual to choose how he will behave, as long as he pays the price. The individual balances the cost of compliance against the benefit of noncompliance, so his behavior is responsive to moderate changes in price.

A prescriptive and a descriptive conclusion can be derived from this theory. If obtaining accurate information about external costs is cheaper than obtaining accurate information about efficient behavior, then price the activity. If the converse is true, then enact a standard backed by a sanction. The supply of an activity is more elastic when it is priced than when it is sanctioned, so the distortion from imposing the wrong price is greater than the distortion from imposing the wrong sanction. On the other hand, when an activity is sanctioned, the partition must be placed in the right place to avoid distortions. Typically, activities should be sanctioned when there is a clear community standard and unclear external costs, and should be priced when the opposite is true.

Sometimes it is difficult to determine whether a law prices or sanctions an activity. To classify such a law, it is useful to note that a price is invariant with respect to the actor’s state of mind and variable with respect to the harm caused by his action, while a sanction is variable with respect to states of mind affecting the seriousness of the fault and the need for deterrence. A price only depends upon the extent of harm because its purpose is to internalize the cost of an activity and to allow individuals to decide whether to engage in it. By contrast, a sanction is often adjusted to the state of mind of the offender because its purpose may include deterrence and punishment as well as compensation.

Many lawyers are blind to the proper use of prices, which contributes to heavyhanded regulation. Conversely, many economists are blind to the distinctively normative element in law, which prevents them from understanding its essential nature. By supplying a unified theory of prices and sanctions, I hope to bring economics into contact with the tradition of jurisprudence and to advance our understanding of how law controls behavior.
A. Variables and Functions

Each of the individual decisionmakers (DM's) control a nonnegative variable, $x$, indicating the intensity of an activity called precaution, that influences the probability, $p$, that others suffer external harm. The extent of the external harm is indicated by the variable $h$. Thus,

- $x$ = DM's precaution, $x \geq 0$,
- $h$ = external harm,
- $p$ = probability density of external harm $h$ occurring, $1 \geq p \geq 0$, where $p = p(h,x)$, $p_2 < 0$, and $p_{22} > 0$.

Precaution is costly to the DM. This cost increases with the amount of precaution. Furthermore, different DM's may pay different prices for precaution. Let $b$ denote a variable which differentiates DM's according to the cost of precaution. Incorporating all elements into the cost function yields the following expression:

$$c = DM's\ text{ cost of precaution} = c(b,x(b)),$$

where $c_2 > 0$, $c_{22} > 0$.

B. First Best Social Optimum

Social costs (SC) created by a $b$-type DM consist of the expected value of the external harm which he imposes upon others and his own costs of precaution, as represented by Equation 1:

$$SC(b) = \int p(h,x(b))hdh + c(b,x(b)),$$

where SC is assumed to be a twice differentiable concave function of $x(b)$ for all $b$.

The number of $b$-type DM's is indicated by the density function $f(b)$. Aggregate social costs are the integral of costs for each type of DM:

$$SC = \int SC(b)f(b)db.$$

Efficiency requires the DM's to choose their precaution $x(b)$ so as to minimize social costs $SC$:

$$\text{efficiency} = \min_{x(b)} SC | x(b) > 0,$$

with first order condition

$$SC = \int p_2 hdh + c_2 > 0 \text{ for all } b.$$

Figure 1 is a graph of Equation 1 for a value of $b$ at which the solution is an equality.

C. Prices and Sanctions

If officials had perfect information and complete control over individuals, then officials could require each individual to choose his level of precaution to satisfy Equation 1. However, officials can only influence the choice
of individual DM’s by changing their incentives. The two mechanisms for changing incentives considered here are prices and sanctions.

If the DM’s behavior imposes external harm h, which occurs with probability \( p(h) \), then officials can attempt to impose a detriment upon the DM, where the detriment is denoted \( r \). If officials impose the detriment \( r \) as a function of the amount of external harm, and no other variable influences the detriment, then the activity is priced: \( r = r(h) = \) activity is priced.

Alternatively, the officials can control behavior by creating an obligation backed by a sanction. The DM is obligated to take precaution \( x \) which is at least as great as the legal standard of care, denoted \( x^* \). The external harm the DM causes is said to be his fault whenever his precaution falls short of the legal standard, whereas the DM is not at fault whenever his precaution is at least as great as the legal standard:

\[
\begin{align*}
&x < x^* = DM \text{ at fault} \\
&x > x^* = DM \text{ not at fault}
\end{align*}
\]

Officials are said to sanction behavior if fault is a necessary condition for the detriment to be positive and the detriment is an increasing function of the extent of the harm:

\[
\begin{align*}
r &= r(h, x^* - x) \\
&= 0 \text{ if } x > x^* = \text{ activity is} \\
&> 0 \text{ if } x < x^*, \text{ where } r > 0 \text{ sanctioned}
\end{align*}
\]

Thus, fault is a necessary condition for the application of a sanction. Two levels of sanctions are distinguished in law. First, the sanction is said to consist of compensatory damages when the detriment equals the external harm: \( r = h = \) compensatory damages. Second, whereas the detriment includes punitive damages to the extent that it exceeds the external harm: \( r - h > 0 = \) punitive damages.

D. Minimizing Private Costs

The individual DM minimizes his private expected costs that consist of the cost of precaution and the expected detriment imposed upon him. Prices and sanctions are particular forms of the detriment function \( r = r(h, x^* - x) \). Regardless of whether the detriment is a price or a sanction, it is minimal in the event that he does not cause external harm. If he causes external harm in such a way that he owes a price or sanction, the debt may not be collected. It may not be collected because the injured party does not bring suit, the state does not prosecute, or the plaintiff cannot assemble enough evidence to win in court. Failure to collect an owed detriment constitutes an imperfection in enforcement. Let the imperfection be measured by the function \( s(r) \), which indicates the probability that the detriment \( r \), which is owed, is actually collected. The expected private costs of the individual DM, denoted \( PC \), can be written as in Equation 2:

\[
PC(b) = \int s(r) r(h, x^* - x) p(h, x) dh + c(b, x),
\]

with the first order condition on precaution given by

\[
\int [(s^2 p + sp)r_2 + sp_2] dh + c_2 > 0.
\]
An activity is said to be perfectly priced for a \( b \)-type \( DM \) if the expected detriment equals the external harm:

\[
[f rs(r)p(h,x)dh = \int hp(h,x)dh] = \text{perfectly priced.}
\]

Perfect pricing results in the internalization of the externality by the \( DM \).
Thus, when an externality is perfectly priced, first-best efficiency is achieved.

1. **Proposition 1:** When an externality is perfectly priced, first-best efficiency is achieved. — To prove this proposition, simply observe that with perfect pricing the first order condition for minimizing private costs, which is given by Equation 3, reduces to Equation 4:

\[
\int p_2(h,x)hdh + c_2(b,x) > 0, \text{ for all } b.
\]

Equation 4 is identical to Equation 1, so first-best efficiency is achieved.

2. **Proposition 2:** If the official price does not exceed the external harm and enforcement is imperfect, or if enforcement is perfect and the official price is less than the external harm, then the amount of external harm will be excessive relative to first-best efficiency. — To prove this proposition, assume that behavior is priced, \( r=r(h) \), so private costs can be written

\[
PC = \int sr(h)p(h,x)dh + c.
\]

Let \( x'' \) denote the value of \( x \) which minimizes \( PC \). Thus, the first order condition for an internal optimum can be written as in Equation 5:

\[
\int sp_2(h,x''')dh + c_2(b,x''') = 0.
\]

Proposition 2 assumes that either \( s<1 \) and \( r=h \), or else \( s=1 \) and \( r<h \). In either case, \( sr(h)<h \). Replacing \( sr \) with \( h \) in Equation 5, and using the fact that \( p_2<0 \), yields

\[
\int hp_2(h,x''')<dh + c_2(b,x''') < 0.
\]

By Equation 1, first-best efficiency can be written

\[
SC' = \int hp_2(h,x')dh + c_2(b,x') = 0,
\]

where \( x' \) indicates the social optimum. By the second order condition, \( SC' \) is increasing in \( x \) at \( x=x'' \). Thus the two preceding expressions imply that \( x''<x' \). Furthermore, the expected external harm is a decreasing function of \( x \), so \( x'' \) results in too much external harm relative to first-best efficiency. This completes the proof.

Conversely, when an activity is sanctioned, the problem of cost minimization by any \( DM \) can be divided into three steps. First, compute the level of precaution which minimizes private costs subject to the constraint that the legal standard is satisfied:

\[
\min c(b,x) \mid x>x^*.
\]

Since \( c \) is an increasing function of \( x \), the value of \( x \) which solves this problem is \( x=x^* \).

Second, compute the level of precaution which minimizes private costs subject to the constraint that the legal standard is unsatisfied:

\[
\min \int s(h)r(h,x^*-x)p(h,x)dh + c(b,x) \mid x<x^*.
\]

Let the solution to this problem be denoted \( PC(x^\oplus) \), where \( x^\oplus \) denotes the
optimal value of $x$.\footnote{62} For the third and final step, compare the two costs to decide whether to conform with the legal standard:

$$[PC(x^*) - PC(x@)] < 0 \implies \text{choose } x=x^*;$$
$$> 0 \implies \text{choose } x=x@.$$ 

As explained, a sanction $r$ stands behind a legal standard $x^*$. A legal standard will be described as perfect if it equals the first-best efficient precaution level for the $DM$ to whom it is applied. Thus,

$$[x^*(b) + x'(b)] = \text{perfect legal standard for } b.$$ 

3. Proposition 3: Assuming that a perfect legal standard ($x^*(b)=x'(b)$) for decisionmaker $b$ is backed by a perfectly compensatory sanction ($r=h$) which is perfectly enforced ($s=1$), then decisionmaker $b$'s precaution will be first-best efficient. — To prove this proposition, consider the three steps discussed above, that guide a $DM$ faced with an obligation backed by a sanction. First, if the $DM$ does not conform to the legal standard, his private cost are identical to social costs as a consequence of the assumptions in the proposition—that is, $PC=SC$ for $x<x^*$. Thus the $DM$'s private costs approach their lower bound at the first-best efficient level $x^*$, which equals $x^*$ by assumption:

$$\inf_{x<x^*} PC = \inf_{x<x^*} SC \implies x=x^*.$$ 

Turning to the second step, if the $DM$ does conform to the legal standard, then his private costs will be minimized by setting $x=x^*$. Thus, when the $DM$ comes to the third step, he will choose $x=x^*$ in either case, so his precaution will be first-best efficient. This completes the proof.

Small imperfections in the price caused the $DM$'s precaution to fall below the first-best efficient level. Small imperfections in sanctions do not have this effect. To prove this, it is necessary to understand why a sanction results in a corner solution for the $DM$. It is often the case that $PC(x^*)-PC(x@)<0$, in which case the $DM$ is at a corner, as depicted in Figure 2. Specifically, assuming perfect enforcement, $s=1$, a perfect legal standard, $x^*=x^*$, and a perfectly compensatory sanction, $r=h$, the discontinuity is given by

$$PC(x^*)-PC(x@) = -\int hpdh.$$ 

The existence of such a discontinuity leads to proposition 4.

4. Proposition 4: Assuming that the legal standard is perfect, ($x^*=x'$), and that the first-best efficient precaution for a $b$-type decision maker is strictly positive, $x'(b)>0$, then if there are imperfections in compensatory damages ($r=h$-epsilon) or in enforcement ($s=1$-epsilon), which are small relative to the perfect values of these variables ($sr=h$), precaution will still be first-best efficient. — To prove this proposition small imperfections in sanctions and enforcement must be shown not to change the outcome of the $DM$'s decision rule. Let $sr$ denote the ex-

\footnote{62}
PRICES AND SANCTIONS

ected sanction, and \( D(sr) \) denote a variation in the expected sanction. Similarly, let \( D(PC) \) indicate the change in the DM’s private costs \( PC \) caused by the variation \( D(sr) \) in the expected sanction.

In the first step of the decision rule, the DM is assumed to conform to the legal standard, so his private costs are

\[
PC(x*) = c(b,x*)
\]

Since the DM is not liable, variations in the sanction and enforcement do not influence his costs. Thus, \( D(PC(x*)) = 0 \). In the second step of the decision rule, assume the DM does not conform. His private costs are therefore

\[
PC(x@) = \int s(r)r(x*-x@,h) \rho(h,x@)dh + c(b,x@).
\]

A first order approximation to the change in these private costs caused by a variation in the sanction or enforcement can be written

\[
D(PC(x@)) = PC'(x@)Dx + \int pD(sr)dh.
\]

Assume that the sanction, enforcement of it, and the legal standard are all perfect in the initial situation, in which case \( PC(x@)=SC'(x*) = 0 \), as demonstrated in the proof to Proposition 3. Substituting this fact into the preceding equation yields

\[
D(PC(x@)) = pD(sr)dh.
\]

Now turning to the third and final step, the variation \( D(sr) \) causes a change in the values of the DM’s decision rule whose first order approximation is Equation 6:

\[
\{PC(x*)+D(PC(x*)) - PC(x@)- D(PC(x@))\} = PC(x*) - PC(x@) - \int pD(sr)dh \text{ evaluated at } x@ = x*.
\]

When \( x*=x@ \), as is true in the initial situation, expansion of the expressions for \( PC \) gives \( PC(x*) - PC(x@) = -\int srpdh \). Thus the change in the decision rule can be written

\[
\{PC(x*)+D(PC(x*)) - PC(x@)- D(PC(x@))\} = -\int p[(sr+D(sr)]dh.
\]

As long as the variation \( D(sr) \) is a fraction of the initial value of \( sr \), then the right side of this equation is negative, which implies by the DM’s decision rule that he should choose \( x=x* \), thus conforming to the legal standard. Obviously, this proposition will remain true if higher order effects are taken into account, so the proof is complete.

According to Proposition 2, small imperfections in a price will cause changes in precaution, and according to Proposition 4, small imperfections in a sanction will not cause changes in precaution. The elasticity of aggregate precaution is the proportional change in aggregate precaution caused by a proportional change in the expected sanction or price. These propositions lead to Proposition 5.

5. **Proposition 5:** When measured at the first best optimum, precaution is relatively elastic with respect to prices and inelastic with respect to sanctions. — When an obligation is backed by a sanction, however, a small change in the legal standard will cause the DM to change his precaution. This leads to Proposition 6.
6. *Proposition 6:* Assuming that the sanction is perfectly compensatory \((r=h)\) and that enforcement is perfect \((s=1)\), the DM will continue to satisfy the legal standard in spite of small imperfections in it \(x^*=x^*+\epsilon\). — Since the proof of proposition 6 follows the same steps as the proof for Proposition 5, it can be abbreviated. Under the assumption in Proposition 6, the DM's decision rule can be written

\[
[PC(x^*) - PC(x^@)] = -\int hpdh < 0 \Rightarrow [\text{choose } x=x^*].
\]

Let \(x^* \) change by an amount \(Dx^*\). A first order approximation to the change in the decision rule can be written

\[
PC(x^*) + D(PC(x^*)) - PC(x^@) - D(PC(x^@)).
\]

It is easy to see that \(D(PC(x^*)) = c_2 Dx^*\) and that \(D(PC(x^@)) = 0\). Hence, Equation 7 reads:

\[
\{PC(x^*) + D(PC(x^*)) - PC(x^@) - D(PC(x^@))\} = -\int hpdh + [c_2 Dx^*].
\]

The first term in square brackets is the *total* external harm, and the second term in square brackets is the *marginal* cost of precaution, both measured as the social optimum \(x=x^*\). Obviously, the *total* harm exceeds the *marginal* cost of precaution at the optimum: \(\int hpdh \gg c_2\).

Thus, the right side of Equation 7 is negative and the decision rule requires the DM to choose \(x=x^*\). The proof is complete.

E. *Second Best Regulations*

In reality, most regulations are insensitive to individual differences. Thus a legal standard may require every DM to exercise the same precaution, even though their compliance costs are different. When regulations are insensitive to individual differences, they are not first-best efficient.

When a regulation imposes the same legal standard \(x^*\) on every DM aggregate precaution is affected. Assume that the individual cost parameter \(b\) has been scaled so that the cost of precaution increases with \(b\):

\[
c_1(b,x(b))>0.
\]

Under certain conditions, there will be a value of \(b\), denoted \(b^*\), which is the switching point between compliance and noncompliance. The effect of a change in the sanction or the frequency of enforcement will be to change the switching point. Thus, aggregate precaution will respond to changes in the sanction by adjustments on the extensive margin and a relatively small number of people will increase their precaution by a relatively large amount. This leads to Proposition 7.

1. *Proposition 7:* Assuming that the individual cost parameter \(b\) can be scaled so that compliance costs, \(PC(x^*)\), increase faster than noncompliance costs, \(PC(x^@)(b)\), an increase in the sanction \(r\) or the frequency of enforcement \(s\), that causes the switching point \(b^*\) to increase by \(Db^*\), will cause an increase in aggregate precaution whose first order approximation equals \(f(x^*-x^@)(b^*)f(b^*)Db^*\). — This proposition, which is little more than an application of definitions, can be proved as follows. Since compliance costs, \(c(b,x^*)\), increase faster than non-compliance costs, \(c(b,x^@)\), there will be no more than one value of \(b\) at which \(PC(x^*) - PC(x^@) = 0\). This value is \(b^*\), the switching point. According to
the decision rule, every DM for whom \( b < b^* \) will conform to the legal standard, and every DM for whom \( b > b^* \) will not conform. Thus, aggregate precaution equals

\[
\int b^* x^* f(b) \, db + b^* \int x@ \cdot (b) f(b) \, db.
\]

Now the increase in the sanction or in the frequency of enforcement cause \( b^* \) to increase by \( Db^* \). Thus, a first order approximation to the change in aggregate precaution is given by \( [x^* - x@ \cdot (b^*)] f(b^*) Db^* \). A relatively small number of people, indicated by \( f(b^*) \), increase their precaution by a large amount, indicated by \( x^* - x@ \cdot (b^*) \). This completes the proof.

F. Punitive Damages

Detriments can be either compensatory or punitive. A detriment is punitive whenever it exceeds the actual harm—that is, \( r - h > 0 \). In law, the application of punitive damages is controlled by the DM’s attitude. Specifically, noncompliance with a legal standard must be intentional or motivated by ill-will before punitive damages are awarded. This section develops this idea mathematically.

Besides controlling his own precaution, assume that the DM also controls some aspects of his attitude towards his own noncompliance. The law is especially concerned with the DM’s intentionality and ill-will. Noncompliance is conventionally graded according to degree of intent as suggested by the following legal categories: inadvertent < reckless < intentional < deliberate. Similarly, ill-will is conventionally graded according to the following categories: well-intentioned < indifferent < reckless < wanton < malicious. Let \( m \) denote a nonnegative variable indicating the DM’s attitude, where higher values of \( m \) represent a higher degree of intentionality or more ill-will:

\[
m(b) = \text{attitude of a } b\text{-type DM}.
\]

In order to relate the DM’s mental attitude to his behavior and make sense of the idea of unintentional noncompliance, it is necessary to introduce a random element into the relationship between the DM’s action and the resulting precaution. Let \( t'' \) be a random variable, let \( x \) be the precaution actually achieved, and let \( y \) be the intended precaution, where \( x = y + t'' \). The decisionmaker chooses \( y \), nature chooses \( t \), and their sum determines \( x \). The intentions of the DM are to be determined with reference to \( y \), whereas the actual performance of the DM is measured by \( x \). Thus, there are three possible conditions of a DM who causes external harm:

- no fault = \( x > x^* \),
- intentional fault = \( x < x^* \) and \( y > x^* \),
- unintentional fault = \( x < x^* \) and \( y < x^* \).

There is a connection between mental attitude \( m \) and the intended action \( y \). A complete analysis of the DM’s behavior when faced with the random term \( t'' \), would be necessary to explain this connection fully. Such an analysis will not be undertaken here. The basic point of such an analysis,
however, is that when private costs of compliance are heterogenous, and the legal standard does not take account of these individual differences, DM's with high compliance costs will choose an intended precaution level $y$ which is less than the legal standard $x^*$. 