

THE OPTIMAL STRUCTURE
OF LAW ENFORCEMENT

Steven Shavell

Discussion Paper No. 112

7/92

Harvard Law School
Cambridge, MA 02138

The Program in Law and Economics is supported by
a grant from the John M. Olin Foundation.

THE OPTIMAL STRUCTURE OF LAW ENFORCEMENT

Steven Shavell

Abstract

When society endeavors to control undesirable behavior, it employs methods that differ in fundamental ways, and one wonders why it has made the choices it evidently has. Why should society sometimes engage in outright prevention of acts -- as when a policeman stops a person from shooting a gun -- and other times employ the threat of sanctions to deter unwanted behavior -- as when the state exacts fines for violation of safety regulations or imposes liability for causing harm? And when sanctions are applied, why should they sometimes be monetary in nature and other times take the form of imprisonment? Further, why should society sometimes rely on private citizens to report violations of law, as happens under the tort system, and other times resort to police or other public enforcement agents for that purpose?

It is suggested in this article that these basic questions about the observed structure of law enforcement can be answered by reference to the theoretically optimal structure of enforcement; the actual pattern of enforcement seems to be broadly consistent with the pattern that is most effective in theory. To this end, the article first defines socially desirable and socially undesirable acts. Next, it sets out what may be considered the principal dimensions of law enforcement, and then describes important means of enforcement (including tort law, criminal law, and safety regulation) in terms of these dimensions. Subsequently, the article considers the determination of the theoretically optimal structure of enforcement and, in light of this, it examines the rationality of the actual structure of enforcement.

THE OPTIMAL STRUCTURE OF LAW ENFORCEMENT

Steven Shavell*

When society endeavors to control undesirable behavior, it employs methods that differ in fundamental ways, and one wonders why it has made the choices it evidently has. Why should society sometimes engage in outright prevention of acts -- as when a policeman stops a person from shooting a gun -- and other times employ the threat of sanctions to deter unwanted behavior -- as when the state exacts fines for violation of safety regulations or imposes liability for causing harm? And when sanctions are applied, why should they sometimes be monetary in nature and other times take the form of imprisonment? Further, why should society sometimes rely on private citizens to report violations of law, as happens under the tort system, and other times resort to police or other public enforcement agents for that purpose?

I will suggest in this article that these basic questions about the observed structure of law enforcement can be answered by reference to the theoretically optimal structure of enforcement; the actual pattern of enforcement seems to be broadly consistent with the pattern that is most effective in theory. To this end, I begin in Section 1 of the article by defining socially desirable and socially undesirable acts. Next, in Section 2, I set out what may be considered the principal dimensions of law enforcement and then describe important means of enforcement (including tort law, criminal law, and safety regulation) in terms of these dimensions. In Section 3, I consider the determination of the theoretically optimal structure of enforcement and, in light of this, I examine in Section 4 the rationality of the actual structure of enforcement.¹

1. Socially Desirable and Socially Undesirable Acts

1.1 Acts and harm. I will identify an act with a set of probabilities of occurrence of possible consequences, and harm with any consequence falling in a socially undesirable category of consequences. Thus, what is normally called an unsafe or dangerous act will be associated with a high probability of harm or, more generally, with a high probability-

*Professor of Law and Economics, Harvard Law School. I wish to thank Gary Becker, Louis Kaplow, A. Mitchell Polinsky, and David Rosenberg for comments and Jesse Fried and Abraham Wickelgren for able research assistance.

¹This article builds on Shavell [1984], dealing with liability and regulation, and on Shavell [1985], briefly comparing liability and criminal law.

discounted or expected harm.²

Harm will ordinarily be interpreted as physical injury or damage to property, although in principle any consequence could be said to be harmful. The theoretical analysis advanced here is not affected by the definition of harm that one adopts.

I will sometimes speak of certain acts as harmful even though, in strict logic, their harmfulness stems mainly from their indirect effects. Theft, for instance, is harmful primarily because it leads to effort to protect property as well as to effort to take property; such effort is inherently unproductive and constitutes the harm due to theft. It will not be necessary, for our purposes, to be explicit about what exactly constitutes the harm due to most acts that are discussed.

1.2 Private and social benefits. The benefits an individual obtains from committing an act will be referred to as private benefits. The benefits that society will be considered to derive from an act may be different from a party's private benefits. Allowing for a divergence between social and private benefits gives the analyst greater freedom to describe society's values. In particular, since the analyst can assume that the social benefits from an act are zero, he is able to study a society that finds some acts (rape, for example) objectionable no matter how high the private benefits may be.

1.3 Socially undesirable and desirable acts. An act will be said to be socially undesirable if the expected harm it brings about exceeds the expected social benefits, and socially desirable if the opposite is the case. Thus, certainly where the expected social benefits are zero, a potentially harmful act will be socially undesirable. Where the expected social benefits are positive, however, whether an act is undesirable will depend on a comparison with the expected harm. Thus, it might be desirable for a person not to shovel ice from the sidewalk if the danger is not great and the burden of the job is substantial.

2. Fundamental Dimensions of Law Enforcement

Here I describe three basic dimensions according to which methods of law enforcement can differ and I then characterize the central methods of enforcement in terms of these dimensions.

2.1 Stage of legal intervention. One important dimension of law enforcement is the timing of legal intervention. Specifically, intervention may take place at the earliest stage, before an act is committed, by means of prevention of the act. Prevention obviously occurs when a policeman stops a person from carrying out some action like shooting at another, but also, on reflection, in manifold other situations: when a truck loaded with explosives is barred

²For example, the expected harm due to an act that will cause harm of 50 with probability 60% and harm of 300 with probability 40% is $60\% \times 50 + 40\% \times 300$ or 150. This may be interpreted as the average harm that would occur were the act repeated many times in approximately independent circumstances.

at the tollgate from entering a tunnel; when a fence is erected around a city reservoir to prevent people from dumping things in it; when a license to operate an enterprise is denied; or when a person is imprisoned and thereby incapacitated.

Prevention rests on physical force or something close to it. The policeman uses physical force to stop the act of shooting; the tollgate and the fence provide barriers against those who would commit undesirable acts; the denial of a license is enforced by the powers of the state, for instance by a sherriff who would lock the doors to a business if it attempted to operate without a license.

Legal intervention may also result after an act has been committed but before harm results (or independently of whether it does so). In this case, intervention is by means of act-based sanctions, that is, sanctions occasioned by the commission of acts. If we punish a person for shooting at another, regardless of whether he hits him, we have imposed a sanction based on the commission of the act of shooting. Likewise, if we impose a fine on a hotel for failure to maintain its sprinkler system in working order, independently of whether there is a fire, we have used sanctions based on the commission of an act. In neither case, note, have we prevented an act (we have not prevented the hotel from operating because it did not have working sprinklers). Rather, we have imposed sanctions in order to deter unwanted behavior.

Finally, legal intervention may come about after harm has occurred, by means of harm-based sanctions. This happens under tort law, under which a person can bring suit and collect damages from a party only if that party has actually caused harm; if the party acts negligently but does not cause harm, he cannot be sued. Harm-based sanctions, like act-based sanctions, have the potential of deterring undesirable behavior but by their nature do not prevent it.

2.2 Form of sanctions. The second dimension of legal intervention is the form of sanctions.³ Sanctions may be of two forms, monetary or nonmonetary. When I speak of nonmonetary sanctions, I shall usually mean imprisonment, and I will generally assume that imposition of monetary sanctions is cheaper than imposition of imprisonment. Imposition of monetary sanctions, while not without cost, amounts to a transfer of purchasing power, of command over resources, not the actual use of resources. In contrast, imprisonment requires the use of resources, since building and operating prisons is expensive, and may as well mean that the imprisoned individual does not engage in productive activity.⁴

³This dimension of intervention is, by its definition, relevant only for act-based and harm-based sanctions, not for prevention.

⁴Furthermore, the imprisoned individual suffers a loss of utility. Assuming that this is counted in the social calculus, there is another reason for imprisonment to be treated as more costly than monetary sanctions. (When a monetary sanction is imposed, it is true that the party who loses money suffers a decline in utility, but this should be balanced, in an approximate sense, by a gain in utility due to the receipt of money by others.)

2.3 Private versus public enforcement. The third dimension of legal intervention involves the role of private parties versus public agents in enforcement. Private individuals may supply the information that results in legal intervention, or enforcement agents hired by the state may do this. When someone seeks an injunction against a neighbor to prevent him from maintaining a noxious compost heap, or when a person sues another for having negligently caused damage to his car, it is a private party who is providing information to the state and in this way instigating legal intervention. In contrast, when a police detective tracks down a murderer or an elevator inspector spots a violation of safety requirements, it is a public enforcement agent who is contributing information to the state for the purpose of legal intervention. The distinction between private parties and public enforcement agents is not always clear, however. If the state offers rewards to private individuals for provision of information, then are we to consider them public enforcement agents? A more detailed analysis than that here would address this issue, but I will simply assume below that anyone who devotes the major fraction of his work effort to enforcement and who is paid by the state -- whether in the form of rewards or in a salary -- is a public enforcement agent.⁵

2.4 Tableau of enforcement methods. While oversimplifying to a degree and repeating some of what was said above, it may be helpful to consider a tableau describing important general methods of law enforcement in terms of the fundamental dimensions of enforcement just discussed.

General method of enforcement	Dimensions of enforcement		
	Stage of intervention	Form of sanction	Private versus public
tort law	harm-based	monetary	private
safety regulation	prevention and act-based	monetary	public
injunction	prevention	-	private
criminal law	prevention, act-based, and harm-based	monetary and nonmonetary	public
corrective taxation	act-based	monetary	public

The table is for the most part self-explanatory. Tort law is said to constitute harm-based legal intervention since, as mentioned above, suit in tort cannot be brought unless harm has

⁵Thus I would consider a bounty hunter to be a public enforcement agent and a "private detective" to be a private party.

occurred. Further, the second and third entries in the row are clear enough, as the tort sanction is ordinarily monetary, namely, the judgment awarded to the plaintiff, and the plaintiff is usually a private party. There are exceptions, to be sure. For instance, the plaintiff might be a public enforcement agency rather than a private party, but the row in the table describes the predominant character of tort law.

Safety regulation is represented as having elements of both prevention and act-based sanctions. Notably, safety regulation is preventive in nature when, as often is the case, regulatory authorities grant licenses to operate only after satisfaction of regulatory requirements. For example, a nuclear power plant or an amusement park may not be allowed to operate until it has passed inspection. Of course, safety regulation may also involve act-based sanctions, for enforcement is frequently carried out by means of penalties imposed for violation of regulatory requirements. (Although these penalties are not usually supposed to depend on the actual occurrence of harm, it may sometimes happen that violations are noticed only because harm came about and enforcement agents then investigated whether requirements were satisfied; hence my description of the penalties as act-based is not entirely accurate, but I shall ignore this qualification below.) The form of sanctions for violation of regulation is normally monetary; in any case, I shall treat regulatory infractions that call for imprisonment or criminal fines under the heading of criminal law. Regulation is stated in the table to be publicly enforced, for ordinarily this is the case, even though it is occasionally true that individuals play a role in indicating violations to the state.

The injunction is described as a means of prevention in which the role of private parties is primary. This is because private parties have the right to report information to the state and then to enlist its power to prevent certain acts from occurring, such as maintenance of a compost heap. Again, this description is not perfect because, for example, a public agency may seek an injunction.

Criminal law is shown as involving intervention at all three stages. Under criminal law, a policeman may not only prevent someone from shooting another but also make an arrest if he fired and missed or if he succeeded in hitting his target. Criminal sanctions are both monetary and nonmonetary, and public enforcement agents play a necessary role in criminal enforcement, although private parties often are also important in reporting information to the state.

The last row in the table pertains to the corrective tax, by which I mean the tax that is usually discussed in the economics literature as a primary way of reducing the level of harmful activities. The example often offered in textbooks is of a tax reflecting the expected harm due to a polluting activity. It is natural and consistent with this example to define the corrective tax in general as an act-based sanction equal to the expected harm due to the act

and paid to the state.⁶

Missing from the table are rows corresponding to other methods of enforcement that one can imagine and to some extent can observe, such as fines for harm done. For the purposes of this article, however, it will be sufficient to consider the methods of enforcement in the table, and I will return to them in Section 4.

3. Theoretically Optimal Law Enforcement

The state has to choose techniques of law enforcement to control the variety of harm-producing acts that individuals may commit. To analyze the optimal means of enforcement, it will be convenient to examine the optimal decision of the state with respect to each of the three dimensions of enforcement. In so doing, I will assume that the measure of social welfare is the social benefits associated with commission of acts less the harm due to them and the costs of law enforcement. These costs of enforcement include the costs of identifying parties to whom the law ought to apply, the costs of applying the law, and also the costs of imposing sanctions.⁷

3.1 Determinants of the optimal stage of intervention. There are several factors that bear on the optimal stage of legal intervention.

Magnitude of possible sanctions. The magnitude of possible sanctions is relevant to the best stage at which to intervene because it affects the ability to deter. If the magnitude of possible sanctions is too low, then sanctions cannot be used to deter, and prevention must be employed to control unwanted behavior. As the magnitude of potential sanctions rises, however, act-based sanctions can be used to deter, and if the magnitude of sanctions becomes sufficiently high, harm-based sanctions can be employed to deter.

To illustrate, consider a person who would obtain a benefit of 50 from committing a socially undesirable act that would cause great harm with probability 20%. Suppose as well that the highest possible sanction is 100, and that the odds of imposing an act-based sanction are 30%, as are, for simplicity, the odds of imposing a harm-based sanction given that harm has occurred. Then the person cannot be deterred by either type of sanction: the highest expected act-based sanction is $30\% \times 100 = 30$, which is less than the benefit of 50 the person

⁶The only alternative definition of the corrective that it is likely a reader would contemplate is a harm-based sanction paid to the state. This does not fit the pollution example, however, since the actual harm due to pollution is often not observable or, if so, easily attributed to the generator of pollution.

⁷This measure of social welfare is used for convenience. In a formal and more complete treatment, I would express social welfare as a function of individuals' expected utilities; and their expected utilities would rise as their benefits increase and fall as harm or costs of enforcement rise (the latter would enter into the calculus by increasing individuals' tax bills).

would obtain from the act; and the highest expected harm-based sanction is⁸ $30\% \times 20\% \times 100 = 6$, which is even less able to deter. Thus, the only way to ensure that the person does not commit the undesirable act is to prevent it. If, however, the maximum possible sanction is higher than 100, the situation changes. If the maximum possible sanction is higher than 166.66, it should be possible to deter the person by use of an act-based sanction (for $30\% \times 166.66 = 50$). And if the maximum possible sanction is higher than 833.33, it should be possible to deter the person by use of a harm-based sanction (for $30\% \times 20\% \times 833.33 = 50$).

What determines the potential magnitude of sanctions? There are, obviously, natural limits on the magnitude of sanctions. For monetary sanctions, the natural limit is the wealth of a party, and for imprisonment, the natural limit is the remaining life of a person. There may be other constraints as well on the magnitude of sanctions. The public may consider it unfair for the sanction for an act to exceed a particular level felt to be appropriate given the gravity of the act (life imprisonment for car theft would probably be thought unfair). In addition, it may be advantageous for the sanction for lesser harms to be low, and for the magnitude of sanctions to rise with the size of harm, so that those who commit bad acts will have an incentive to refrain from doing greater harm (to only rob a person, not kill him as well). Whatever the reasons may be for limits on the potential magnitude of the sanction for an act, it may influence, as described above, the relative ability to control acts through legal intervention at the different stages.

Probability of prevention or of application of sanctions. The probability of preventing a person from committing an undesirable act, or the probability of imposing a sanction, as the case may be, also bears on the choice of the stage of legal intervention. The lower the probability relevant for intervention at a particular stage, the less attractive intervention at that stage will be, other things being equal. Thus, if it is difficult to intercept people who are about to commit an act, prevention will be problematic; and if it is hard to identify people who have committed acts or done harm, in order to impose sanctions on them, the threat of sanctions will be diluted.⁹

There does not seem to be any appealing a priori assumption that can be made about the relative magnitudes of the probability of prevention, the probability of imposing act-based sanctions, and the probability of imposing harm-based sanctions. It may be easy in some circumstances to prevent people from committing an act but difficult to impose sanctions. People may be effectively prevented from dumping materials into a public reservoir by fencing it in but if the reservoir were not fenced in and someone dumped material in it, he

⁸The figure 30% is multiplied by 20% because the harm-based sanction is imposed only if harm occurs, which happens with probability 20%.

⁹I am implicitly assuming that there is a given probability of applying the law at this or that stage of intervention, but in fact the probability is to a degree determined by the effort and resources devoted by society to the task. Thus, a reader may wish to think of the probability here as approximating the probability that would be relevant were society to spend the optimal amount on enforcement at a stage of intervention.

might be hard to apprehend. In other contexts, however, it may be difficult to prevent people from committing acts but easy by comparison to catch them after they have done so. It would be essentially impossible to prevent people from making improper left turns when driving; the only way to do this would be to have police officers riding in cars alongside drivers. But if drivers make improper turns or cause accidents, this may be observed and they may be sanctioned for their behavior. Continuing, one can readily imagine situations in which there is a significant difference between the ability to identify and impose sanctions on people who have committed acts, versus on those who have done harm. Sometimes, what allows a person to be identified is the actual doing of harm. If a firm leaves a holding tank dangerously full of chemicals but the tank does not overflow and pollute a river, no one may know the better, whereas a spill might be immediately recognized as due to the firm because the chemical could only have been released by that firm. Other times, what allows a party to be identified is an act, not the harm. Suppose that the holding tank at the firm is easy to inspect, but that if the tank overflows, this will be impossible to attribute to the particular firm because the chemical is already present in the river and is discharged into it by a multiplicity of other sources.

Thus, the probability of prevention or of imposition of sanctions is a factor that, depending on circumstances, could weigh in favor of one or another of the three possible stages of legal intervention.

Information of the state about acts. As a general matter, it appears that the state will have more information about the character of acts the later the stage of legal intervention. Thus, the factor of the state's information should tend to favor harm-based sanctions over act-based, and the latter over prevention. Suppose that a social authority not only knows that an act has been committed but also that it caused harm. This fact, that harm came about, usually constitutes information about the dangerousness of the act. If a man shoots a gun at another and the bullet kills him, it often can be inferred that the act was more dangerous and thus more important to control than if the shot missed (in which case it would be more likely that the shot was, for instance, meant only to scare). Similarly, if a social authority learns that an act has been committed, it will ordinarily know more about the dangerousness of the act than if the state had intervened to prevent the commission or completion of an act. If a man is reaching for a gun and is stopped at that point, the social authority will not know whether he would have fired or where he would have aimed.

Information of parties about their acts. The knowledge that individuals have about the dangerousness of their acts is also germane to answering the question about the optimal stage of legal intervention. The more information that individuals have about the dangerousness of their acts, the more appealing will be later intervention.

If individuals have a good understanding of the nature of their acts, harm-based sanctions should function well in the absence of other problems. If individuals realize that use of a pesticide is likely to result in substantial harm and they recognize its magnitude, then if there is an appropriate harm-based sanction, the individuals will be adequately deterred

from using the pesticide. If individuals do not recognize how much harm their acts might cause, however, they would not tend to be adequately deterred by harm-based sanctions. They still might be deterred well by act-based sanctions, though. If individuals do not have direct knowledge that a pesticide is dangerous, they nevertheless might learn that its use will result in sanctions and thus be dissuaded from using it by the threat of an act-based sanction. Nevertheless, individuals who do not recognize the dangerousness of an act might be unaware that committing the act could result in an act-based sanction or a harm-based sanction -- so that prevention would be required to control their behavior. They might not be aware that using a pesticide that they think is safe would result in sanctions. Hence, banning sale of the pesticide (or limiting its sale to licensed parties) might be required to control its use.

I should note that in saying that prevention or act-based sanctions instead of harm-based sanctions may be appealing where individuals lack knowledge of the dangerousness of their acts, I am implicitly making two assumptions: that a social authority may have superior knowledge of risk; and that the social authority is not able to apprise individuals about the risk -- for if people can learn about the risk, harm-based sanctions ought, in principle, to work well. With respect to the first assumption, a social authority might possess superior information because information (such as about the toxicological properties of a pesticide) takes effort to develop and the private incentive to generate information is not adequate.¹⁰ The second assumption is also sometimes justified, for people are limited in the amount of time and effort they can devote to learning about risk and sometimes are simply unable to absorb information that is technical in nature, described in statistical or epidemiological terms.

Enforcement costs. There may be substantial differences in the costs of enforcement society bears if it intervenes at one stage rather than another. For example, prevention of an act may be cheaper than use of sanctions to deter when the former can be accomplished by use of a physical barrier. As was mentioned, to stop people from dumping in a reservoir, all that may be needed is a fence, and this may be much less costly than hiring guards to patrol the reservoir. But where prevention requires enforcement agents, prevention may be expensive relative to use of sanctions (recall the example of preventing drivers from making improper turns). With regard to enforcement costs and act-based versus harm-based sanctions, harm-based sanctions appear to possess an advantage in that they are applied less often, only when acts actually result in harm, rather than regardless of whether harm occurs.

It should be added that although I have discussed enforcement costs here as an independent determinant of the optimal stage of legal intervention, they are implicitly related to other determinants discussed above. For instance, the more that is spent on enforcement, the higher will be the probability of preventing someone from committing an act, and the greater will be a social authority's information about acts. A more general analysis than that

¹⁰The private incentive to develop information will be less than the socially appropriate incentive if the information can be used by many people and the developer of the information cannot charge for it.

here would treat these interactions.

3.2 Determinants of the optimal form of sanctions. Where sanctions are to be employed and either form of sanction can adequately deter, the optimal form of sanctions will be monetary, given the assumption that imprisonment is a socially more expensive sanction to impose. Thus, only if monetary sanctions cannot deter appropriately may it be desirable to employ imprisonment, and then only if the deterrence gained by use of imprisonment is worth the added costs. This implies that the following factors are relevant to the choice about the form of sanction.

Level of wealth. The lower is a party's wealth, the less likely it is that a sanction equal to his wealth will adequately deter, and thus the more likely that imprisonment will be needed to deter.

Private benefits from committing acts. The higher are the expected private benefits from committing a socially undesirable act, the higher is the monetary sanction necessary to deter, and thus the more likely that monetary sanctions alone will not be sufficient to deter.

Probability of imposition of sanctions. The lower is the likelihood of imposition of sanctions, the higher will be the monetary sanction necessary to deter, and thus again the more likely it is that imprisonment will be needed to deter. A person who would obtain a benefit of \$10,000 from committing an undesirable act could be deterred by a certain sanction of \$10,000, but if he would face a sanction with a probability less than one, the sanction needed to deter would have to be higher. For instance, if he would face a sanction with a probability of 50%, a sanction of \$20,000 might be needed to deter him, and if he would face a sanction with a probability of only 1%, a sanction of \$1,000,000 might be needed to deter him.¹¹ At some sufficiently low probability, the monetary sanction necessary to deter him would exceed his wealth, so that use of monetary sanctions alone could not deter him.

Expected harmfulness of acts. The higher the expected harmfulness of acts, the more important they are to deter, and thus the more likely it is to be worthwhile for society to bear the expense of use of imprisonment to accomplish deterrence, if monetary sanctions are not adequate. An act that cannot be properly deterred using monetary sanctions alone may not be worthwhile using imprisonment to deter because the harmfulness of the act is not great (consider littering); but if the harmfulness of the act is substantial, then use of imprisonment may be justified despite the attendant expense.

3.3 Determinants of the optimal public versus private roles in enforcement. The issue to be addressed here is when society can rely mainly on private individuals to supply the information that results in legal intervention, and when it must resort to use of public

¹¹If the person is risk-neutral, these sanctions would be needed to deter since the expected sanction must equal \$10,000. If the person is risk-averse, lower sanctions would deter.

enforcement agents for that purpose.¹² I will divide analysis of this issue into two cases. The first is where private parties quite naturally come into possession of information about those to whom the law should apply; here I will argue that private parties can be the primary source of information for enforcement. The second case is where information relevant for enforcement takes effort to obtain; here I will suggest that the use of public enforcement agents may be necessary.

Private parties naturally possess information about the identity of injurers. It is often true that potential or actual victims of harm or third parties are easily able to identify those to whom the law should apply. If so, then it is socially desirable for such parties with information relevant for enforcement to supply it to a social authority, rather than have the state spend its resources on enforcement activity. To harness the information that exists, however, those who possess it must have an incentive to come forward. I now review several factors pertinent to this issue and will conclude that, in the main, people should have, or can be given, reasonably good incentives to report information about parties to whom the law should apply.

One factor of relevance is financial gains from reporting. Such gains obviously increase incentives to report, and are under the control of the state: the state can pay a reward itself for reporting or it can require sanctioned parties to pay sums to those making reports. A complication arises, however, concerning bargains not to report between parties who can do so and liable parties. This will not be a problem -- there will not be significant dilution of deterrence -- where bargained-for payments will approximate sanctions. A general circumstance when that will be so is when a reporting party's reward equals the sanction of a liable party.¹³ When, however, the reward to a reporting party is lower than the sanction, deterrence will likely be diluted, since bargained-for payments will tend to be less than sanctions.¹⁴ There are two steps the state can take to remedy this problem: simply increase rewards to reporting parties; or make agreements not to report illegal, deem them "extortion" or "blackmail" rather than permitted "settlements." This will tend to reduce the incidence of private agreements not to report.

A second factor, also leading to reporting, is the desire to avoid suffering harm. This

¹²In stating that the matter to be addressed concerns private versus social supply of *information* about violations of law, I am ruling out consideration of the issue of private versus social *prevention* of undesirable acts. Private parties can directly prevent acts by doing such things as erecting fences around their property, installing locks, and so forth. Analysis of this issue, on which see Shavell [1991], is beyond the scope of the present article.

¹³If I can collect \$100 by reporting you and \$100 is your sanction, then we would tend to settle for about \$100. (Our motive to settle might be to avoid transactions costs.)

¹⁴If I would receive a reward of \$10 by reporting you and you would pay a sanction of \$100, we would tend to settle for an amount between \$10 and \$100. (Note as well that by settling, we save more than merely transaction costs; we also achieve (joint) gains of the difference between the \$100 sanction and the \$10 reward, namely, \$90; so we have a stronger motive to conclude an agreement than in the previous case.)

operates when the person with information is a potential or actual victim and can alter behavior that threatens him by reporting, as when a person who sees his neighbour driving drunk down the block notifies the police to reduce the chance that in the future he himself will turn out to be harmed.

A third factor, one which discourages reporting, is fear of reprisal from the named party. The likelihood of reprisal, though, can be reduced in many cases by the reporting person's keeping his identity secret, either by supplying his information anonymously or by requesting anonymity from a social authority. But anonymity cannot always be achieved, especially when the liable party is able to deduce who must have supplied the information. When reprisal is a concern, it can to some degree be countered by the state's making any form of reprisal illegal.

A fourth factor is the retributive motive, the desire to see people who have acted wrongly suffer sanctions.¹⁵ This leads to reporting independently of other considerations. It is probably most important when the wrongdoer has actually done harm (rather than only acted undesirably) and when the person who has the knowledge of wrongdoing is the victim or potential victim (rather than a third party).

To conclude, then, it seems that the possibility of financial reward coupled with the desire to avoid future harm and the retributive motive create a set of incentives sufficient in principle to induce people who have information about liable parties to report their information in broad circumstances. Nevertheless, fear of reprisal from liable parties and the possibility of agreements not to report qualify this conclusion.

Effort must be expended to identify injurers. In many circumstances, the identity or whereabouts of a party who ought to bear a sanction or ought to be prevented from committing an act is not known to any private party and will take effort and expense to determine.¹⁶ In these circumstances, public enforcement activity may be justified. In essence, this is because I assume that private parties do not possess an inherent advantage in identifying or locating liable parties, and because their incentives to do so are not well aligned with what is socially desirable.

To amplify, when effort is required to determine the identity of, or to find, a person who committed an undesirable act or did harm, I suppose that private parties have no underlying advantage because the same strategies and techniques that are available to them are also available to public authorities. Only to the extent that private parties initially possess information unknown to the state about the identity or location of liable parties would private parties enjoy an advantage, but the maintained hypothesis in this section is that private parties

¹⁵For an economically-oriented analysis of the retributive motive, see Posner [1980].

¹⁶Private versus public enforcement in this case has been discussed by Becker and Stigler [1974], Landes and Posner [1975], and Polinsky [1980].

possess little or no such information.¹⁷

With regard now to the issue of incentives, it seems that the motive of private parties to find liable parties could either fall short of or exceed the socially correct motive to invest resources in that task, depending on circumstance.¹⁸ Private parties might well have a socially inadequate motive to find liable parties. If I determine who dumped a pollutant in a lake, I may not benefit much because I do not use the lake very often, or if I find the person who stole my car, I may not recover it (perhaps it will have been broken down and sold for parts) and may risk reprisal. In general, I am unlikely to benefit personally from the marginal benefit in deterrence created by identifying a liable party, even though this deterrent effect represents the major social reason for identifying liable parties. On the other hand, private parties may have a greater motive to find liable parties than is socially desirable. A person may be led to expend a great effort to find a liable party because he will receive a large reward for so doing, such as a tort judgment, even though the social benefit is small, because the deterrent that would be generated is limited, or because the result would merely be to shift wrongdoing from one area to another.

If private and social incentives to find liable parties are misaligned, cannot the problem be cured by use of appropriate rewards or subsidies to encourage greater private effort, or by suitable taxes to discourage private effort? In some cases, it seems that private incentives can be so influenced. To induce me to devote the correct resources to finding a person who polluted a lake, perhaps it would be enough to offer me a higher reward. But perhaps that would not be sufficient. One problem is that if the reward is available to anyone, rather than to a single enforcing party, there might be a wasteful effort devoted to finding the party, akin to the waste involved in patent races or in fishing in a common fishing area.¹⁹ Another problem is that the best technologies for finding liable parties often require coordination of many individuals, sometimes on a vast scale. Additionally, it is efficient for various information systems -- such as fingerprint records and data banks on offenders -- to be developed, even though the benefits of these systems would be hard for the private sector fully to capture. Such information systems, as well as certain other enforcement technologies, may constitute natural monopolies (because of high fixed costs and low marginal costs). All this suggests that use of financial rewards paid to those who identify liable parties might not lead to as well-functioning a system as that established by a single public entity (or a functionally equivalent regulated monopoly) charged with the responsibility of finding liable

¹⁷I also put to the side the possibility that the private sector may have an advantage in motivating employees (because a private firm may be better able to discharge those who do not perform and may generally be more resistant to political pressures that retard efficient operation).

¹⁸Polinsky [1980] stresses the divergence between private and social incentives to find liable parties.

¹⁹This problem with private enforcement was emphasized by Landes and Posner [1975].

parties.²⁰ For these reasons (though with some tentativeness), I will often assume in what follows that public enforcement agents are useful.

4. Consistency of Observed and Theoretically Optimal Law Enforcement

I will now consider the five methods of law enforcement described in the tableau -- tort law, criminal law, safety regulation, the injunction, and corrective taxation -- and for each ask whether its actual use is consistent with theoretically optimal law enforcement, as discussed in the last section. That is, for each method of law enforcement, I will ask whether its characteristics -- in terms of the stage of legal intervention, the form of sanction, and the public and private role in enforcement -- makes sense given the nature of the acts that that method of enforcement apparently aims to control. My discussion will be frankly speculative in many respects, but I will still come to the conclusion that these methods of enforcement can be explained as rational in a broad and very approximate sense by the hypothesis that they represent optimal law enforcement.

4.1 Tort law. I have characterized tort law as the method of law enforcement that employs harm-based monetary sanctions which are awarded to individuals who suffer harm and report it when they bring suits. Now I want to explain why the method of enforcement with these characteristics should be employed to control the harmful activities that tort law does control. What are those activities? They are virtually all the activities of everyday life and business enterprise, for any of them can cause accidents -- slips and falls, collisions involving moving vehicles, product injuries, harms caused by explosions and fires, and the like.²¹ Of course, acts causing accidents are controlled not only by tort law but also by regulation and other means of law enforcement, as I shall discuss later; for now, however, my focus is on tort law and acts for which these other methods of law enforcement are not of substantial influence. For concreteness, the reader might keep in mind an act like lighting a grill in one's backyard, which could result in a fire that would destroy a neighbor's property, or running to catch a bus, in the course of which one might knock into someone and injure him.²²

Stage of intervention. Consider first the issue of the stage of legal intervention. Why do we not employ prevention rather than sanctions to control the acts of concern, and why should we use sanctions triggered by harm rather than by commission of undesirable acts?

²⁰What I have said also indicates why it is not a contradiction to think that people can be induced by appropriate rewards to *report* what they know to a social authority but cannot be properly induced to *find* liable parties. The former task is a simple one for a private party to carry out; the latter is complicated because, as I have said, it is not generally something that can be done by just one person; it involves development of information systems, and so forth.

²¹I am restricting myself in this discussion to unintentional torts.

²²In neither of these cases is it likely that regulation or another form of legal intervention would play a real role.

As to why not prevention, my response is simply that this would be impractical. As mentioned, to control driving behavior through prevention, enforcement agents would have to be sitting alongside drivers; and to prevent people from recklessly chasing after buses or from grilling in a dangerous manner, enforcement agents would have to be present to police behavior of people around busstops and in their backyards. That a regime in which all this were to be done strikes one as fanciful only reflects the fact that prevention as a general instrument of control of unwanted behavior would be inordinately expensive.

Consideration of the knowledge of the social authority also argues against prevention. How would an enforcement agent (assuming that he were present) know that a person was about to make an improper left turn before it was actually done, or that a person was going to run for a bus until he or she actually did so? In many situations, an enforcement agent would not have the information to prevent an undesirable act before it happened.

Thus, the use of sanctions rather than prevention seems necessary to control the broad category of acts with which tort law is concerned, but now we must say why harm-based sanctions rather than act-based sanctions are applied. The answer, I think, again lies in part with enforcement costs. The sheer number of acts that can cause harm is great; to hope to employ sanctions whenever acts appear dangerous would thus be very expensive. It would seem far better on administrative cost grounds to focus on the small-by-comparison number of acts that turn out to result in harm. Moreover, the actual doing of harm furnishes us with some information about the dangerousness of acts; that they did harm means that they were probably dangerous. Hence, harm-based sanctions seem generally best as the method for the law to employ to control behavior giving rise to what we call torts.

Form of sanctions. But why should the form of sanction for torts be a money sanction rather than imprisonment? The explanation is that money sanctions work tolerably well, so that society usually need not resort to the more expensive form of sanction of imprisonment. The reasons that monetary sanctions work fairly well become apparent when we review the factors discussed in Section 3.2.

In most accident situations that we describe as torts, the likelihood of the responsible party having to pay, if liable, is fairly high (at least relative to what will be described in certain other contexts, such as where criminal acts are committed). If my grill causes a fire and damages my neighbor's home, he will be likely to know how the harm came about and who was responsible; if I run into someone while trying to catch a bus, that person will ordinarily obtain my name (if I run, my act may become criminal, as I will later discuss).

With regard to ability to pay for harm, although there may be problem, it is often not one of significance. The damage to my neighbor's home might be limited, and in any case I am likely to have liability insurance sufficient to cover it. Moreover, the fact that the likelihood of my having to pay for harm is large means that the magnitude of the sanction necessary to deter me will not be inflated. Hence, it seems that the ability to pay the sanctions needed to deter undesirable acts should be reasonably good as a general matter (but

see the qualifications in the comment below).

An additional factor to be noted in this regard is that under the negligence rule, which is the dominant rule of liability in tort, an individual will be induced to exercise proper care even if his assets are lower than the harm he might cause, and plausibly substantially below it. For instance, if a person would cause harm of \$100,000 with probability 2% and could reduce this risk to 1% by an expenditure of \$10, must his level of wealth be near \$100,000 for him to be induced to make the expenditure? The answer is that even if his wealth were only, say, \$10,000 -- far less than \$100,000 -- he would make the \$10 expenditure to avoid liability, for if he does not, his expected liability would be $2\% \times \$100,000$ or \$200. Under the negligence rule, in other words, the incentive to take adequate precautions is sharp since one thereby avoids all liability for harm.²³

Turning next to the expected harm caused by acts, recall that this is relevant because of the consequences of failure to deter. If the expected harm is low, then failure to deter is not as socially costly than if expected harm is high. Now most acts creating risks of accidents do not create very large risks, even though the size of the harm, if it eventuates, may be large. When I make an improper left turn, an accident usually will not occur as a result, and if I run after a bus, usually I will not bump into someone and cause serious injury. The problem, therefore, due to occasional failure to deter such acts is not nearly as serious as is failure to deter other types of acts, notably those called criminal, that we shall discuss.

Public versus private enforcement. Finally, what can be said about the private nature of enforcement in tort law? It seems quite reasonable, in that victims of tortious harm will usually know the identity of injurers. Thus, as generally argued in Section 3.3, private enforcement is best, and allowing victims to collect rewards -- their tort damages -- should supply them with incentives to come forward with their information. Were we instead to rely on enforcement agents of the state to report tortious harms, we would be spending our resources needlessly.

Also, recall that it was observed that agreements between victims and liable parties not to go to court do not dilute incentives when rewards to reporting parties equal the sanctions of liable parties (for then bargained-for payments should reflect sanctions). This is the situation in tort law, for what the liable party pays is what the party who sues successfully receives. Hence, we can see as rational that settlements between victims and potentially liable parties are allowed in the tort context.

Comment. This review of the typical tort has left out, among other things, some important qualifications concerning situations in which deterrence is apparently weakened. In such situations, what we would expect, and what we find, is that the legal system responds so as to remedy the problem of insufficient deterrence.

²³The observation that incentives to take care are sharp under the negligence rule is first developed by Summers [1983]; see also the discussion in Shavell [1987], p. 167.

One type of situation where deterrence might be inadequate is where an injurer tries to conceal his identity, the classic example being hit-and-run accidents. Here, the response of the legal system is often to deem the act that would otherwise only be a tort a criminal act as well, allowing for the application of additional nonmonetary sanctions to alleviate problems with deterrence. Another response of the legal system in such cases is to permit imposition of punitive tort damages; implicitly, these can reflect the probability of escape from liability, and offset the underdeterrence problem.

It is also possible that an injurer would find that he is not identified because of circumstance rather than design. For instance, a firm's product may not be seen as the cause of disease because of other possible causes of the disease or because other firms sold the same product and the victim cannot say which he had purchased. In this case, the law could, and in fact does sometimes, relax the normal causation requirements that force the plaintiff to establish a clear connection between his harm and a particular injurer. Under the market-share theory, notably, any firm that could have caused harm will be liable, but damages will be only in proportion the likelihood that it was the cause. This tends to solve the problem of dilution of deterrence.

There are other examples that could be mentioned, but the point should be clear. The reaction of the tort system to situations in which deterrence would be compromised is along the lines of what would be predicted, namely, some step to bolster deterrence.

4.2 Criminal law. To control the acts to which criminal law applies -- theft, robbery, rape, murder, treason, and so forth -- it was observed that the law intervenes at all three stages, imposes nonmonetary as well as monetary sanctions, and relies on public enforcement agents along with private parties. The question to be addressed here is why criminal law has these characteristics.

Stage of intervention. The heart of the answer as to why legal intervention in criminal law occurs at all three stages is that deterrence is weak for the acts of concern, meaning that it is desirable for society to avail itself of all opportunities to control the acts.

Deterrence of criminal acts is poor for several reasons. One is that the likelihood of apprehending and convicting offenders is low, often significantly less than 50%.²⁴ That the

²⁴According to Federal Bureau of Investigation, U.S. Department of Justice. 1990 Uniform Crime Reports, Crime in the United States, Table 20 p. 165, the probability of an arrest following a reported offense was 21.6%, when averaged over all offenses. For specific types of offense, the probabilities of arrest were as follows: 45.6% for violent crimes; 67.2% for murder and non-negligent manslaughter; 52.8% for forcible rape; 57.3% for aggravated assault; 18.1% for property crimes; 24.9% for robbery; 13.8% for burglary; 20.5% for larceny-theft; 14.6% for motor vehicle theft; and 14.9% for arson. But offenses are not always reported, and arrests do not necessarily result in convictions, so that these probabilities overstate the true probabilities of conviction. According to Bureau of Justice Statistics, U.S. Department of Justice, Report to the Nation on Crime and Justice (1988), p. 34, the probability of reporting is 48% for violent crimes, 26% for property crimes, 69% for motor vehicle theft, 49% for household burglary, and 25% for household larceny. Also, certain states report statistics on the probability that arrests for serious crimes result in

probability of apprehension and punishment of criminal acts is generally small is not surprising, since such acts frequently are carried out with a view toward escaping sanctions. A person who plots a crime, like murder or treason or theft, will generally execute it in such a way as to avoid responsibility. A robber can search for a lone person in a deserted area at night, and in a neighborhood where he would be unlikely to be recognized.²⁵

Deterrence is weak in the criminal realm not only because the likelihood of sanctions is low, but also for two other important reasons. First, many of those who commit crimes are poor, making monetary sanctions ineffective. Second, the benefits that individuals tend to obtain when they commit criminal acts are large, making the sanctions necessary to deter high. I will elaborate on these below when I discuss the form of sanctions.

The weakness in deterrence means, as I said, that society wants to take all opportunities to control the acts in question. In particular, we endeavor to prevent people from committing criminal acts because we cannot rely on sanctions imposed on those who commit such acts to discourage them. Likewise, we punish people for committing undesirable acts even if they do no harm because we cannot rely on sanctions imposed on those who succeed in their harmful acts to adequately deter. Thus, we impose sanctions on those who shoot at others and miss, who pick empty pockets -- generally, those who attempt crimes but fail. And, of course, we punish those who succeed in doing harm, for this is our final way of augmenting deterrence.

Although we intervene at every stage in the criminal context, we usually impose lower sanctions the earlier the stage. If a person is caught when he is planning to commit an act or is interrupted in an attempt, the sanction is generally less than if he commits the act but fails, that is, engages in a complete but unsuccessful attempt. And if he engages in a complete attempt and fails, his sanction is often less than if he succeeds in his act. This pattern of sanctions is consistent with the point that the state ordinarily possesses less evidence of the harmfulness of acts at earlier stages.

Form of sanctions. Turning now to the form of sanction, the main point to be dealt with is why monetary sanctions are generally not adequate to deter. One factor is that those who commit crimes tend to be poor; notably, statistics show that the inmate population is

conviction, p. 60; this probability is 61% in California, 69% in Minnesota, 68% in Nebraska, 67% in New York, 50% in Ohio, 56% in Pennsylvania, 79% in Utah, and 61% in Virginia. If I assume for simplicity that 60% is the likelihood of an arrest resulting in a conviction for all types of offense, the probability of punishment for violent crimes becomes $48\% \times 45.6\% \times 60\% = 13.1\%$ rather than 45.6%, that for property crimes becomes $26\% \times 18.1\% \times 60\% = 2.8\%$ rather than 18.1%, that for motor vehicle theft becomes $69\% \times 14.6\% \times 60\% = 6\%$ rather than 14.6%, and so forth.

²⁵In contrast, note, the typical tort occurs (as the word "accident" suggests) at an unpredictable time and place, and thus only by chance can the responsible party can easily avoid being identified. Moreover, since the tort is often connected with one's everyday behavior, it may naturally occur in circumstances such that those who observe it will know who caused it: if I leave my sidewalks covered with ice, any accident that occurs will automatically be known to be linked to me.

composed of people who have very little income prior to arrest.²⁶ Why should the poor display a greater tendency to commit crime? Most obviously, they will often commit crimes that are economic in character because they have a greater need for money. In addition, lack of wealth is associated with substandard education, drug and alcohol abuse, and social alienation, all of which makes those with low wealth less likely to respond to extra-legal influences that would channel their conduct in desirable directions. For these various reasons, then, criminality is associated with low wealth, and thus with our being unable to accomplish deterrence with monetary sanctions. Having little to lose, the poor need to be deterred through imposition of nonmonetary sanctions.

A second factor suggesting that use of monetary sanctions alone will not be enough to accomplish deterrence is the low probability of apprehending and punishing those who commit crimes. I explained above why the probability is low, and consideration of an example will illustrate the inadequacy of deterrence were we to employ solely monetary sanctions. Suppose that the probability of punishment for motor vehicle theft is about 5% (see note 24) and that the value of a vehicle to a thief is \$500. Then the monetary sanction needed to deter would be in the range of \$10,000, an amount substantially exceeding the ability to pay of the many thieves who have essentially no liquid assets and few possessions with market value. Similar calculations for many other crimes make it evident that reliance on monetary sanctions would not be nearly adequate to deter.

A third factor is that the benefits people obtain from committing the acts that we call criminal are frequently large, which makes them difficult to deter. One of the defining elements of a crime is intent, which often has to do with a person's purpose. Those who commit acts like murder or rape or treason will usually obtain a substantial private benefit from success, which is something that will be relatively likely given their desire not to fail in their acts.²⁷

A fourth factor is that the expected harm done by criminal acts is, in an average sense, substantial. Most obviously, the magnitude of harm associated with acts that we classify as criminal is often great. But this is not always true (as when a person steals a small amount), and at the same time the harm due to many noncriminal acts may be significant (people may be killed due to negligence). What I believe to be the main reason that crimes are socially more costly socially is that they are more likely to cause harm than noncriminal acts. When a person tries to steal, or to rape, or to kill, he is going to succeed with a relatively high

²⁶For example, in Bureau of Justice Statistics, U.S. Department of Justice, Report to the Nation on Crime and Justice (1988), p. 49, it is reported that the median income of inmates is lower than the poverty level, and that almost half of male inmates had been unemployed prior to incarceration.

²⁷Note that the situation is altogether different in the context of the typical tort, where the private benefit an actor usually obtains from acting improperly is avoiding the cost of a precaution, like saving the effort of removing oily rags that could cause a fire. It requires a much smaller penalty to induce a person to give up this sort of gain than it does to induce a person to give up the likely gains from murder, which may involve large sums of money or great personal satisfaction.

probability. In fact, we ordinarily *define* crime in a way that implies the probability of doing harm will be high: the legal meaning of "intent," usually necessary for a criminal conviction, is either that a person acted with the purpose of causing harm -- meaning that harm would be likely -- or, if not, that harm was, from outward appearances of the act, very likely (as when someone acts with extreme recklessness). Hence, the expected consequences of failure to control the acts we call crimes seem large, and this makes it socially worthwhile to employ the more costly nonmonetary sanction of imprisonment to control the acts.

One more point may be added about imprisonment, namely, that it prevents further bad acts by incapacitating individuals. This is significant, because those who commit acts have revealed themselves not to have been deterred by the threat of sanctions. Hence, they constitute a subpopulation whose behavior we often can control only by prevention, and this is accomplished by imprisoning them. Alternative nonmonetary punishments, like whipping or branding, would not prevent further crimes because they do not incapacitate.

Finally, let me comment on situations in which criminal sanctions do not include imprisonment but involve criminal fines. Such sanctions tend to be observed when crimes are not of the most serious nature. In these cases, one suspects that money sanctions coupled with the nonmonetary sanction of humiliation are enough to deter reasonably well, and thus that society need not invest in the more expensive sanction of imprisonment. Also, it may be that the need to prevent further acts is not great, so that the incapacitative benefit of imprisonment is not significant.

We might ask, however, if society does not impose imprisonment as a punishment, why should it categorize an act as criminal? A chief reason may be that this serves as a socially inexpensive way to increase deterrence. The classification of an act as criminal stigmatizes the person subject to sanction, and thus may add to the humiliation just mentioned that accompanies criminal conviction. Indeed, for corporations, this may be one of the only reasons for classifying certain acts as criminal, for corporations cannot be imprisoned. The strength of the stigmatizing effect, however, is limited by, among other factors, the scope of acts labeled criminal. As the range of such acts expands, the stigma will be diluted, for society will no longer view a criminal offense as signaling the commission of such an undesirable act.

Public and private enforcement. The explanation for the use of public enforcement agents to control and identify those who commit crimes is clear. Victims of crimes frequently do not know who has harmed them. A great deal of effort is often required to find those who have committed criminal acts. Hence, by the general arguments put forward in Section 3.3, a corps of enforcement agents is needed. Moreover, in cases where individuals can identify those who have harmed them, the criminals will often flee, so that enforcement agents will be needed to locate them. In addition, where the victims know who have harmed them and could report this, there is the issue of reprisal by the criminals. This problem is made more serious by the severity of criminal sanctions and the character of people who commit criminal acts, and thus may call for enforcement agents to play a role even where the

victims know who have harmed them and those people have not fled.

Private parties do, however, play a role in reporting criminal acts. They often notify the police when they have been victimized or when someone has attempted to harm them. Sometimes they do this to prevent further harm, as when they fear that the person wants to harm them in particular; other times they may make reports to satisfy a retributive urge, out of a feeling of social responsibility, or in order to collect insurance proceeds.

In any case, a private party is not allowed to make a bargain with a criminal not to report his crime, even if the private party is the victim of the crime. The reason that this makes sense is that allowing such bargains would dilute deterrence. The most that the criminal could give the victim is the criminal's wealth, so the cost to the criminal of such a bargain is at most this amount. But the criminal sanction will generally include imprisonment, so that the sanction that the state would impose will be higher. Accordingly, unlike in the tort context, private agreements would result in substantially lower costs to offenders and would weaken deterrence. In addition, one supposes that allowing private settlements would encourage criminals to threaten victims with harm unless they settled, further reducing deterrence.

4.3 Safety regulation. As discussed earlier, by safety regulation I refer to control of behavior by means of prevention, such as by refusing or removing licenses to operate equipment or businesses, as well as by imposition of act-based monetary sanctions, that is, fines for violation of rules.²⁸

Stage of intervention. Why should the stage of legal intervention for the behavior controlled by safety regulation be before harm comes about? The reason is essentially that use of sanctions for harm done might not provide sufficient deterrence. This is plausibly often true where we tend to see safety regulation, in controlling the risks of fire, foods and drugs, or transport of dangerous substances, to take important examples. A fire at a restaurant could harm a large number of people and create losses far exceeding the net worth of its owner. Likewise, the harm caused by contamination of food could be widespread, easily surpassing the assets of the owner. Because parties would not have sufficient assets to cover losses in such cases, they might not be led by the prospect of liability to take adequate steps to reduce risks. In addition, there may be problems due to parties escaping liability for harm done in some areas of regulation. In particular, many health-related and environmental risks are difficult to trace to their origin. Finally, it may be that harm is dispersed, as is often true with pollution-caused losses, so that individual victims might not find bringing suit worthwhile.

Given, then, that deterrence would often be inadequate if harm-based sanctions alone were relied upon in areas of safety regulation, there is a need for prevention or act-based

²⁸ Although in fact violation of regulations sometimes results in criminal sanctions, I will focus here, as I indicated earlier, only on regulations punished by civil monetary sanctions.

sanctions, and this need is what safety regulation satisfies. Safety regulation appears to take the form of prevention where that form of legal intervention is administratively easy. A general circumstance in which this is so is where a party is applying for a license to undertake an activity, or for renewal of a license, and where what is regulated is the presence of some physical device, like a sprinkler system in a hotel. In such cases, a regulatory authority can deny the license if the sprinkler system has not been installed. The appeal of prevention over act-based sanctions is that the former by definition stops unwanted behavior, whereas act-based sanctions rely on deterrence. Act-based sanctions, though, are often employed where prevention would be expensive to accomplish. This is frequently true when what needs to be controlled is human behavior, which can be modified from one occasion to the next. Thus, if there is a regulation concerning the number of people monitoring controls at night in a nuclear power plant, or the clearing of ice from sidewalks in front of commercial establishments, we would expect enforcement to take the form of act-based sanctions.²⁹

Consider next the question whether a regulatory authority will have the knowledge needed to formulate good rules. It seems typical of much regulation that its requirements can be justified by common knowledge or something close to it, so that informational demands on regulators are minimal. When regulation calls for trucks carrying explosives not to enter tunnels or for people not to hunt in hiking areas, regulatory authorities can be reasonably confident that their requirements are justified. It is true, though, that regulation sometimes requires substantially more than common knowledge. In such cases, regulatory authorities often develop information on their own, such as where government agencies determine health and environmental risks. In some such instances, the knowledge of a regulator may be superior to that of private parties (and this may provide a reason for regulation independent of inadequate deterrence from harm-based sanctions). When, however, information becomes a problem for a regulators, we may see relaxation of regulatory requirements. Thus, we may see that although fire safety regulation stipulates use of sprinkler systems and fire retardant materials, it does not go so far as to say exactly what can and cannot be stored in closets, how often kitchen equipment must be cleaned, and so forth. The limits of regulation appear to reflect the quality of information of regulators.

To summarize the discussion here, it is suggested that safety regulation is explained primarily as an answer to the need for control of behavior where harm-based liability might not create proper deterrence. Further, regulation seems to take the form of prevention rather than act-based sanctions when the former is administratively simpler, and it appears that regulatory rules are sensitive to the adequacy of information possessed by regulators.

²⁹There is, though, a complication that should be noted. If a commercial establishment has failed to clear ice from the sidewalk, it can be prevented from doing this in the future by having its license to operate removed. Analysis of this issue would involve the following points. If an activity like operating a store is prevented, then there may be a loss in social welfare exceeding that associated with preventing a particular unwanted act, like failing to clear ice. Hence, one would expect that act-based sanctions would be preferred to removal of licenses unless the regulatory violations were sufficiently serious and it was decided that act-based sanctions would not function reasonably well in the future.

Form of sanctions. Let us turn now to the question why enforcement of safety regulation through act-based sanctions is by means of money sanctions, that is, why criminal sanctions are not needed. The reasons why money sanctions usually work well enough are several. First, because the sanctions are act-based and therefore may be applied whether or not harm comes about, they can be effective even if their size is relatively small (recall the discussion in Section 3.1). If, for example, regulation calls for a party to spend \$50 on an exit sign to warn people how to escape a public place should a fire occur, a fine of only \$100 would suffice if the probability of checking for fire exit signs is 50%. (In contrast, under harm-based sanctions, sanctions would be imposed only if a fire actually occurred, and if this is unlikely, the size of the sanction necessary to induce installation of the sign could be far larger.³⁰)

Another factor helping to explain why monetary sanctions will ordinarily be enough to lead to adherence to safety regulation is that the benefits obtained by violating regulations are often, like those in the area of tort, the savings from not taking precautions, like that from not purchasing an exit sign. In particular, the benefits seem lower than the gains often obtained by persons committing criminal acts. An additional factor is that regulated parties often have substantial wealth, especially when they are corporations.

These arguments notwithstanding, imposition of monetary sanctions sometimes is not enough to enforce safety regulation, and when so we tend to observe two things. On one hand, we may see resort to prevention. Thus, if in some area of regulation, expected harms are large and a firm's assets are insufficient to pay fines for violations, we may see the firm shut down unless and until it satisfies regulatory requirements. On the other hand, we may see regulatory violations punished under criminal as well as civil law.

Public versus private enforcement. Finally, let us ask why enforcement of safety regulation is primarily public in nature. The answer seems to be mainly that private parties cannot be counted upon to be aware of risks. If a restaurant kitchen is unclean yet I do not get sick when I go to the restaurant, how will I know about the status of the kitchen? If a nuclear power plant is not following proper procedures, it is unlikely that I will be aware of this. If a person fails to vaccinate his dog against rabies, there is no way for me to be aware of it in the normal course of events. Thus, the absence of harm seems a generally strong factor explaining why, in regulated areas, we would not expect people to be aware of who needs to be sanctioned for violating regulations or of how to prevent their violation.

Nevertheless, it could, of course, be the case that someone becomes aware of a regulatory infraction. A person might have occasion to go into a restaurant kitchen, hear about kitchen conditions from a restaurant employee, or be an employee himself. In these cases, the person might have a motive to report, which might be that the person wants to prevent harm to himself or the community or that he would be rewarded by the state for so

³⁰If a fire would occur with probability .05% for instance, then the sanction necessary to induce installation of the \$50 sign would be at least \$100,000, since $.05\% \times \$100,000$ is \$50.

doing. But it does not seem that people have nearly enough information for us to rely generally on private enforcement in regulated areas. Society thus needs to make an effort to discover violations of regulation and therefore needs public enforcement agents.

4.4 Injunction. The injunction is the legal device whereby private parties may enlist the power of the state to prevent harm. Injunctions may be brought in the context of the general class of nuisances, for example, to stop an activity that generates noxious odors, like maintaining a compost heap in a suburban neighborhood, or to make someone chain a vicious dog. Injunctions may also be brought to prevent activities that hold the risk of large harms, not normally called nuisances, such as where people enjoin a factory from continuing its operations because they fear that it is introducing a carcinogen into the water supply.

Stage of intervention. Prevention is the mode of legal intervention where the injunction is employed for reasons that are in some respects similar to those offered in relation to safety regulation. On one hand, where the injunction is used, our knowledge of the undesirability of the acts in question is often good; and on the other hand, use of sanctions might not accomplish deterrence, though, as I shall indicate, the relevance of this latter factor is not entirely clear.

With regard, first, to a social authority's knowledge, consider the typical nuisance. It will be obvious, virtually by definition of a nuisance, that the behavior is undesirable. Moreover, when a party seeks an injunction, the harm will often be ongoing -- as in the case of a compost heap that is already producing foul odors -- so that little need be demonstrated to prove that more harm will eventuate in the future. Or, if the harm is not ongoing, there will frequently have been past harm. In the case of a vicious dog, there will frequently have been previous incidents of problems with the dog; indeed, this may well be a prerequisite for obtaining an injunction concerning the dog. With regard to injunctions secured in situations other than that of nuisance, similar arguments can be made; sometimes, there is ongoing harm; and generally, courts insist on having clear evidence of future danger before granting injunctions.

Consider next the question whether the use of sanctions would not accomplish deterrence. Would sanctions fail to deter in the typical nuisance situation? Is it unlikely that imposition of damages would dissuade a person from maintaining a compost heap or from keeping a vicious dog? I can imagine factors that would reduce the deterrent of liability in such situations, such as difficulty in proving that a particular dog bit a child, but most do not seem to me to constitute a general basis for favoring prevention by means of injunctions. A factor that may be of significance, though, is difficulty in determining the proper level of damages.³¹ In any event, in some instances where injunctions are granted, there are fairly

³¹When components of loss (such as disutility due to a noxious odor) are hard to calculate, courts often exclude them from tort damages, leading to dilution of deterrence and to the appeal of the injunction. This notion is consistent with the fact that to obtain an injunction, it must usually be true that the damage amount cannot be adequately proved.

clear arguments supporting the view that liability would not adequately deter. In the case of a firm that is polluting the water supply with a carcinogenic substance, that might be true. The possibility that the firm would not be able to pay for the cancers caused or that the firm could not be linked to the cancers could explain lack of confidence in deterrence and the utility of an injunction.

Public versus private enforcement. Why is the injunction a privately employed device to prevent undesirable behavior? The answer is evidently that the injunction is used where private parties themselves become aware of dangers. When this is so, it is socially desirable, as argued generally in Section 3.3, for society to use their information. And giving private parties the legal right to prevent the harm that they themselves would suffer provides them with an incentive to supply information.

When we think of nuisances, we can see that it is indeed the case that private parties have information about harms that may come about. Maintenance of compost heaps in backyards and the keeping of vicious dogs are practices whose dangerous characteristics are ordinarily obvious to private parties. Outside the area of the typical nuisance, it may also happen for one reason or another that private parties become aware of a risk. For instance, a neighborhood organization concerned about the safety of the water supply may test the water and discover that a firm is discharging a carcinogen into it. However private parties come upon their information, providing them an opportunity to prevent harm is helpful to society since it makes use of their information.

4.5 Corrective taxation. The corrective tax -- a tax equal to the expected harm caused by an activity -- is, recall, a tool usually described in economics textbooks as one of the most important for the control of harmful activities. The corrective tax is an act-based sanction because it is not related to harm actually done but to predicted harm. The corrective tax for emitting a pollutant into the atmosphere is set equal to the harm that is estimated to be done by the pollution, not to the harm actually done by the pollution.

Despite its prominence in economic literature, the corrective tax is in fact rarely employed to control undesirable activities; its use is restricted to a few instances of polluting activities. For the most part, society relies on regulation, liability, the injunction, or criminal law for controlling undesirable acts. After the rationality of use of the tax to control pollution is examined, I will briefly address the question why use of the tax is so infrequent.

Stage of intervention. Let us first ask why a harm-based sanction may not be desirable to control pollution. The general answer is that it may be difficult to identify and link harm caused by pollution to responsible parties. The ways in which pollution can cause harm are manifold and complicated; and harm may take years to eventuate. If, for these reasons, polluters would often escape liability for harm, deterrence from harm-based sanctions would be diluted.

If harm-based sanctions would not perform well to cure pollution problems, why

should the corrective tax be employed rather than fines for violation of regulations? And why should a corrective tax be preferred to prevention? A tax may be preferable to regulation because a tax requires less information to apply. To determine the proper corrective tax a social authority needs only to know the expected harm due to the activity. If the authority sets the tax equal to the expected harm, the taxed party will engage in an activity if and only if his benefit exceeds the expected harm, which is to say, if and only if the activity is socially desirable.³² But to formulate proper regulation, a social authority must know not only the expected harm due to the activity, but also the benefit from the activity; for the authority has to compare benefits with expected harm to determine whether the activity is socially undesirable and should be disallowed. Similarly, a social authority must know benefits as well as expected costs to know when to intervene and prevent an act. Hence, both regulation and prevention require the social authority to know more than it needs to know to impose a tax, making a tax potentially superior.

Another issue of relevance concerns the cost of enforcing a tax. If the activity to be taxed is fairly easily monitored, and the expected harm is approximately equal to a simple multiple of the scale of the activity, then the tax will be easy to administer. This is plausibly the case for many pollutants, such as the quantity of an effluent that is released into the atmosphere.

Form of sanctions. Since the corrective tax is set equal to the expected harm caused by an activity rather than to the actual harm, the tax will often be lower than the actual harm and more likely to be within the capacity of the taxed party to pay than liability for harm. Also, since taxed parties are often corporations, they usually will have assets sufficient to pay the taxes. This is not to deny that nonmonetary sanctions might be necessary to increase deterrence. It is rather to say that in the few situations in which taxes are employed, they seem to be within the capability of parties to pay.

Public versus private enforcement. That the corrective tax for pollution should be publicly enforced seems easily explained. Private citizens will not naturally have the information about the amount of pollutants that firms or individuals are discharging. Hence, enforcement agents must obtain the information.

Limited use of taxation. Having attempted to say why the corrective tax seems to work particularly well in many cases of pollution, I now want to comment on the question why the tax would not be advantageous in much greater realms. Consider why the tax would not work well in the classic area of tort. Why not tax people for running after buses without looking where they are going or for leaving flower pots dangerously close to the edge of balconies? That is, why not tax negligent behavior, or, for that matter, all behavior on the basis of the expected losses? The explanation may lie in the enormous administrative expense that would entail. In contrast, the tort system, as I have emphasized, is

³²Also, the party will be led to take appropriate precautions to reduce harm, given that he engages in his activity, but I omit this point for simplicity.

administratively cheap, for it is employed only in the relatively few cases in which harm comes about.

Administrative expense also appears to explain why the corrective tax may often be inferior to regulation where regulation is used. Consider, for instance, regulations requiring the display of exit signs in buildings to which the public has access, or regulations concerning the number of lifeboats carried by large vessels. To enforce such regulations, there will only have to be a transaction involving money when regulations are violated, which is to say, only when the number of exit signs or of lifeboats is inadequate. Under corrective taxation, however, there will be transactions involving payments of taxes by all owners of public places and of ships. Hence, tax collection may well be more expensive to administer than regulation. This disadvantage of taxation may offset its informational advantage over regulation.³³

5. Concluding Remarks

(a) Although the theme in Section 4 has been explanatory, the general theory of enforcement discussed in this article also suggests ways of altering the current system toward greater efficiency. For example, it may well be that where private parties have information about violations, we should more vigorously reward the reporting of violations. There are usually people in large organizations who become aware of violations of regulations, yet my impression is that "whistle-blower" rewards are not common and that protection of the identity of informants is not a well-developed strategy. It also appears to me that use of the corrective tax could be expanded. There must be additional areas where it would be administratively feasible to impose a tax based on expected harm, such as for purchase of noisy machinery.

(b) I want also to acknowledge a basic omission from the analysis, namely, that there are sometimes important extra-legal influences on behavior -- market forces and social sanctions such as loss of reputation in the community. These may serve to varying degrees as substitutes for legal sanctions. If, for instance, a firm sells tainted food, consumers will be likely to decide to take their business elsewhere, so the need for control by the legal system will be reduced. Or the prospect of loss of face were a person negligently to cause a fire that harms his neighbors may help to make him careful. Although these factors reduce the need for legal sanctions, they do not seem to do so generally, for they require that private parties be aware of the identity of wrongdoers. Hence, market forces and informal social sanctions are unlikely to be of much help where public enforcement agents would be needed to enforce the law.

³³Weitzman [1974] offers another argument in favor of regulation over taxation that does not involve administrative costs, but it rests on the assumption that a tax must be of a simple form, namely (in the context of pollution), the quantity of an effluent multiplied by a per unit tax. If, as here, it is assumed that the tax equals expected harm -- which may be a nonlinear function of the quantity of effluent -- the tax is unambiguously superior to regulation, in the absence of administrative cost considerations.

References

- Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, Vol. 76, 169-217.
- Becker, Gary S., and George Stigler, "Law Enforcement, Malfeasance, and the Compensation of Enforcers," *Journal of Legal Studies*, 1974, Vol. 3, 1-19.
- Landes, William, and Richard Posner, "The Private Enforcement of Law," *Journal of Legal Studies*, 1975, Vol. 4, 1-46.
- Polinsky, A. Mitchell, "Private versus Public Enforcement of Fines," *Journal of Legal Studies*, 1980, Vol. 9, 105-127.
- Posner, Richard A., "Retribution and Related Concepts of Punishment," *Journal of Legal Studies*, 1980, Vol. 9, 71-92.
- Shavell, Steven, "Liability for Harm versus Regulation of Safety," *Journal of Legal Studies*, 1984, Vol. 13, 357-374.
- , "Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent," *Columbia Law Review*, 1985, Vol. 85, 1232-1262.
- , *Economic Analysis of Accident Law*, 1987, Harvard University Press, Cambridge.
- , "Individual Precautions to Prevent Theft: Private versus Socially Optimal Behavior," *International Review of Law and Economics*, 1991, Vol. 11, 123-132.
- Summers, John, "The Case of the Disappearing Defendant: An Economic Analysis," *University of Pennsylvania Law Review*, 1983, Vol. 132, 145-185.
- Weitzman, Martin, "Prices vs. Quantities," *Review of Economic Studies*, 1974, Vol. 41, 477-491.