AVOIDANCE COSTS IN OPTIMAL PENALTIES

Jeffrey S. Parker, George Mason University School of Law

07-37

George Mason University Law and Economics Research Paper Series

Corrected: 1/19/90

Preliminary Draft

AVOIDANCE COSTS IN OPTIMAL PENALTIES

by

Jeffrey S. Parker*

January 1990

---

*While this paper is still very preliminary, I already have benefited greatly from conversations with Michael Block, Mark Cohen, John Lott, and Mitch Polinsky.

1. Introduction

The purpose of this paper is to expand upon an argument I have made in the context of penalties against organizations:1 that avoiding a legal prohibition is not free, but rather involves "avoidance costs" that must be considered in determining an optimal public law enforcement policy. In the case of organizations, it is relatively straightforward to identify avoidance costs in the form of expenditures on monitoring and controlling agents of the firm, and to observe that the level of expenditure is likely to be a function of the expected penalty. However, there are more subtle types of avoidance costs in the form of lost opportunities for productive activity, which are difficult to observe directly but may have a very significant aggregate effect on social welfare.

I believe that the significance of avoidance costs is more general than the case of business firms and "white-collar" violations, and extends to all types of

---

1 See Parker, Criminal Sentencing Policy for Organizations: The Unifying Approach of Optimal Penalties, 26 Am. Crim. L. Rev. 513, 556-61 (1989).

potential offenders and offenses. The existing literature does not explicitly consider the problem of avoidance costs in determining optimal enforcement, and in some instances offers solutions that depend critically on the implicit assumption that avoidance costs are zero. If avoidance costs are considered, the optimal enforcement policy is affected in certain circumstances. In particular, the policy of setting expected penalties equal to the violator's gain, in order to reduce public enforcement costs, is shown to be sub-optimal even where it can be assumed that violators' gains are always less than the external harm created by the violation. More generally, the existence of avoidance costs and their dependence on the expected penalty level explains why optimal penalties always should be based on external harm, and never on the violator's gain.

2. The Optimal Penalty Literature

For the most part, the existing literature on optimal public law enforcement does not encounter the

problem of avoidance costs,2 because the standard solution is to set the expected penalty equal to external harm. The conventional explanation for this result is to permit a potential violator to compare the prospective gain directly with the social harm, and to commit the "optimal" violation when gain exceeds harm. Although I would suggest that this explanation is incomplete for omitting explicit consideration of avoidance costs, that suggestion does not appear to change the optimal enforcement policy in the general case.

However, under certain conditions -- specifically, where there is a wealth boundary on the absolute size of the penalty and gain is always less than harm -- it has been suggested that the optimal solution is to base the expected penalty on gain, in order to save on public enforcement

---

2Some references to avoidance costs are found in Posner, Economic Analysis of Law, 206-09 (1986) and Posner, An Economic Theory of the Criminal Law, 85 Colum. L. Rev. 1193, 1205-06 (1985). However, Posner considers avoidance costs only as an adjunct to the possibility of erroneous convictions, and does not develop a systematic result. As I will discuss below, I do not think that avoidance costs exist only with error, although they are analytically similar in that non-violators bear a cost.

costs. In this instance, the failure to consider avoidance costs leads to a sub-optimal result.

The optimal penalty theory, first developed by Becker,3 considers the problem of law enforcement as one of minimizing the total social costs associated with both violations and enforcement effort. Potential violators are assumed, on average, to behave rationally, in the sense that violations occur only when the violator's gain exceeds the expected penalty, i.e., the absolute penalty times the probability. Taking the simplest case -- assuming a socially costless form of punishment (a monetary fine), risk neutrality among potential offenders, and the absence of erroneous convictions4 -- Becker showed that the optimal fine was equal to the external harm of the violation divided by the probability that the fine will be imposed, given that

---

3Becker, Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968).

4Unless otherwise stated, I will continue to use this simplified model throughout this paper. Obviously, relaxing the assumptions of the simple model will affect the optimal policy. However, the simple model is adequate for my purposes.

a violation has occurred.5 This result can be restated as setting the expected fine (amount times probability) equal to the external harm caused by the offense.

Setting the expected fine equal to external harm is described as "optimal" in the sense that only welfare-diminishing violations are deterred. A welfare-enhancing violation (conventionally described as one where gain exceeds harm) is not deterred, because to do so will make society worse off than if the violation takes place. Both types of violations are subject to the same expected penalty equal to external harm, thus "internalizing" the harm to the harm-producing activity.

However, finding the optimal expected penalty does not determine the optimal enforcement policy, because the expected penalty is made up of two different variables -- the magnitude and the probability -- which are to be selected simultaneously. Here, and still under the assumption that fines are the form of penalty, the basic argument is that while fining is socially inexpensive if not costless, enforcement is costly. As the expected fine can

---

5This formulation ignores public enforcement costs or implicitly includes them in external harm.

be composed of any combination of magnitude and probability, the optimal solution was said to be the selection of a very high magnitude and a very low probability, thereby saving on the costs of enforcement.6 While this prescription would lead to virtually infinite fines, obviously there is a finite wealth boundary that constrains the absolute magnitude of a fine. But within that boundary, the standard solution is select a high magnitude and low probability.

This tradeoff between the probability and magnitude of fines is developed further by Polinsky and Shavell, who consider the effect of a wealth boundary on the magnitude of the fine, under alternative assumptions of risk neutrality and risk aversion.7 Taking the simple case of risk neutrality, they endorse the Becker result by arguing that the optimal magnitude of a fine is as high as possible -- equal to individuals' wealth.8 In determining the

---

6See Becker, supra note 3, at 183.

7Polinsky & Shavell, The Optimal Tradeoff between the Probability and Magnitude of Fines, 69 Am. Econ. Rev. 880 (1979).

8The alternative assumption of risk aversion produces a lower optimal magnitude and a higher optimal probability, on the basis of the tradeoff between

(Footnote Continued)

optimal probability, however, Polinsky and Shavell do not simply divide external harm by wealth, but rather suggest that further savings in public enforcement costs can be achieved by using gain rather than harm, if gain can be assumed to be less than harm.9 Their reasoning essentially is that a lower expected fine equal to gain is sufficient to deter, and using the lower expected fine permits a lower probability and therefore lower public enforcement costs.10 As developed further by the example discussed in the following section, in my opinion this analysis runs afoul of the problem of avoidance costs.

My own prior article approaches the problem of avoidance costs from the somewhat different perspective of considering the deficiencies of a gain measure as the basis

---

(Footnote Continued)

public enforcement costs and individuals' risk-bearing costs. See id. at 884-85, 886-87; see also Polinsky, Introduction to Law and Economics, ch. 10 (1983).

9The Polinsky-Shavell analysis considers a bimodal distribution of gains, in which some offenders gain amount a, less than external harm, while others gain amount b, which is greater than external harm. They use the lower amount a to determine the optimal probability, reasoning that public policy does not in any event wish to deter violators in category b.

10See Polinsky & Shavell, supra note 7, at 882-84, 888.

for penalties against organizations.11 It seems obvious that organizations do incur positive avoidance costs, at least in the form of expenditures intended to prevent the firm's agents from committing a violation. If so, and if the firm is in a better position than the public to avoid the offense, then an expected fine equal to gain is sub-optimal whenever gain is less than harm, because the firm is given no incentive to incur avoidance costs and therefore is not deterred from a violation that would be optimal to prevent. My argument there is that setting the expected penalty precisely equal to harm solves this "underdeterrence" problem by giving firms the socially optimal incentive to incur avoidance costs, and also eliminates an "overdeterrence" problem that would arise if firms were faced with an expected penalty greater than harm. However, I was not considering a situation where both probability and magnitude were selected simultaneously.

In this paper, I will argue that the same considerations developed in the context of firms are equally applicable to individuals, and that the simultaneous

---

11See Parker, supra note 1, at 556-63.

selection of the probability and magnitude of the penalty does not change that result.

3. A Suggestive Example

My thinking on the generality of avoidance costs was stimulated by an example posed to me by Professor Polinsky in response to my argument against the use of gain-based penalties. His example, which draws upon the prior Polinsky-Shavell analysis, is as follows:

Assume a violation that invariably causes $10 of external harm and produces only $5 of gain to the violator. To simplify the analysis, assume that each potential violator is a risk-neutral individual with a total wealth of $100, and that the only form of penalty available is a monetary fine. In this situation, the optimal magnitude of the fine is $100. Under the Parker argument, we would have to catch and convict 1 out of every 10 offenders, in order to produce an expected penalty equal to the external harm of $10 ($100 x .1). But we can make society better off by lowering the probability of conviction to 1 in 20. The violation will still be deterred by an expected penalty of $5 ($100 x .05), because no one will have any incentive to

commit the violation. At the same time, we will save the difference in public enforcement costs of catching only 1 in 20 versus 1 in 10, which is likely to be substantial.

My reply to the example is that it implicitly assumes that avoidance costs are zero. What if it costs the individual something (in addition to the foregone gain) to avoid committing the violation? To illustrate how avoidance costs can affect the analysis, let us put some more flesh on the bare bones of the example. Assume that the violation in question is the classic Polinsky-Shavell example of double parking.12 Assume further that in order to avoid double parking, each potential violator has to drive around the block, which will cost that person $1 worth of lost time and aggravation.13 In this situation, no one will be deterred, because no one will drive around the block. Everyone will violate, because compliance costs them $1, while violating

---

12See Polinsky & Shavell, supra note 7, at 886-87.

13To advance the analysis, I assume here that saving the cost of driving around the block is not part of the $5 gain achieved by violators. This assumption is relaxed, and shown to be unimportant, in section 4 below.

costs zero ($5 gain less $5 expected fine).14 Whatever was spent on enforcement was wasted, because no one was deterred.

To check on the social welfare implications here, let us add the assumptions that there are 100 potential violators and enforcement costs are fixed at $10 for each percentage point increase in the probability of detection above zero. With positive avoidance costs and a $5 expected fine, everyone violates, resulting in total harm of $1,000 (100 x $10), less gains of $500 (100 x $5), plus $50 enforcement costs ($10 x 5%). The net social loss is $550. Now compare the situation where the probability is raised back to .1 and the expected fine is $10. Under this scenario, no one violates but everyone incurs the $1 avoidance cost, which total $100, as do enforcement costs ($10 x 10%). The net social loss is $200. Although twice as much has been spent on enforcement costs, society is better off with the higher probability.

Obviously, the outcome of this particular numerical example depends upon the assumed relationship

---

14Actually, only 5% of the violators were convicted, and each paid a $100 fine.

between the rate of change in enforcement costs and the other variables. But that is a separate subject, deserving of its own treatment.15 My point is made so long as it is recognized that, under plausible enforcement cost assumptions, the consideration of avoidance costs changes the optimal enforcement policy even where it is known that all violations cause more harm than gain.

Why not simply increase the expected fine to $6, and at least save some enforcement costs by selecting the correspondingly lower probability? If we knew that avoidance costs never could exceed $1, this solution might be attractive. However, I do not think that it is plausible to assume that avoidance costs are subject to some absolutely fixed constraint. Instead, it is more plausible to assume that avoidance costs, like other costs, will continue to increase so long as there is some incentive to incur an additional burden. Furthermore, in the case of

---

15At least some portion of enforcement costs is likely to be imposed as part of the penalty, therefore requiring an even higher probability. One of the implications of my analysis is that the optimal solution may be affected by a tradeoff between public enforcement costs and avoidance costs, depending upon their relative rates of change.

avoidance costs, the primary incentive is supplied by the level of the expected penalty.

To illustrate in the context of the example, it seems implausible to assume that potential violators will drive around the block only once. If given the incentive to do so, they may drive around the block many times. How many times is determined by the strength of that incentive, which is supplied by the expected penalty. If the penalty becomes too large, it creates the problem of "overdeterrence" -- deterring welfare-enhancing violations. In the presence of positive avoidance costs, overdeterrence can occur even when gain never exceeds harm, which has been conventionally considered as a situation presenting no danger of overdeterrence.

Suppose that the parking regulatory authorities have decided that, since gains from double parking never exceed harms, there is no danger of deterring a welfare-enhancing violation. Therefore, the expected fine is set at $20 as part of the "absolute deterrence" policy. In that case, the failure to consider avoidance costs will be very costly to society, as every potential offender now has the incentive to drive around the block 15 times (at $1 per circuit) rather than violating. Even if this policy

succeeds in eradicating double parking, it costs society more (100 potential violators x $15 avoidance cost = $1,500, plus enforcement costs) than if society did nothing to control double parking (100 violators x $10 external harm = $1,000, with no enforcement costs). As this example suggests, there is a finite limit to the amount of avoidance cost that a potential violator should be encouraged to bear. Setting aside several refinements -- including the consideration of enforcement costs -- that limit is given by the difference between harm and gain.

This conclusion becomes even more important if we relax the basic example's assumption that gain is fixed at a level below external harm. The more realistic assumption is that gains as well as avoidance costs will vary across potential violators. One potential violator may gain nothing from a violation, and another may gain more than the external harm. Similarly, one person may be able to find a parking place after driving around the block once, while another may not find a place after 15 circuits. Under optimal enforcement, society wants the non-gainer to drive around the block more times than the person who gains more, and it does not want anyone to drive around the block 15 times.

The analysis in this section brings out two main points of significance to optimal enforcement policy. First, it shows that a consideration of avoidance costs changes the optimal enforcement policy, by refining the definition of an "optimal" violation to be one where the sum of gain plus avoidance costs (and not merely gain) exceeds the external harm. Second, it shows that the level of avoidance costs, unlike gain, generally will be endogenous with the expected penalty level. Before discussion of some of the implications of these points, I will turn back to a consideration of the nature of avoidance costs, particularly as compared with gain.

4. What Do We Mean by "Gain"?

One possible interpretation of my argument so far is that all I have accomplished is to identify an ambiguity in the concept of "gain." Thus, relaxing the assumption of the previous section, one could argue that my example of double-parking avoidance costs -- driving around the block -- is nothing more than a negative expression of at least part of the gain to violators. Violators do not drive around the block; they double park, and one of their "gains"

is saving the time and aggravation of looking for a lawful parking place.

In response to this argument, I would concede that there is in fact some overlap between the concepts of avoidance cost and gain, in the sense that avoidance costs often can be characterized as "negative gain," that violators' saved avoidance costs often are characterized as "gain," and that avoidance costs and gain often are traded off against each other -- in fact, this is one of the principal conclusions of my analysis. I also would agree that there is ambiguity in identifying instances of avoidance costs, or for that matter, gains as well. Despite its common usage in discussions of enforcement policy, the concept of gain is highly ambiguous and largely unexplored in the literature. However, I do believe that avoidance cost is analytically distinct from gain, for the simple reason that avoidance costs are incurred by non-violators (and possibly violators as well), whereas gain, by any definition, accrues only to violators.

Whatever "gain" means, it is uniformly considered to exist only when a violation takes place. Going back to the double parking example, one may want to debate what comprises a violator's gain. For example, if the

double-parker has gone into an office building to close a multi-million dollar business transaction, is the double-parking "gain" the profits on that transaction, or is it merely the price of parking in the building's garage? The concept of "gain" inevitably involves this type of ambiguity as to causation and remoteness, among other issues. However, there is never a discussion of a non-violator's "gain," or "negative gain." By definition, a non-violator's gain is zero. In contrast, a non-violator certainly incurs avoidance costs. As illustrated by the discussion in the preceding section, there can be substantial avoidance costs even where there are no violations at all, and therefore no gain.

Given the general ambiguity of "gain," I suppose that one could redefine it to include "negative gains" incurred by non-violators. Doing so might eliminate the terminology of avoidance costs, but not their analytical effect on optimal enforcement policy. Under the basic optimal penalty analysis, gain is assumed to be applicable only to violators, and to be determined exogenously to the penalty system. If avoidance costs could be called "negative gains," then non-violators' "negative gains" would have to be included in the social welfare analysis, and to

be considered endogenous with the expected penalty level. The implications for optimal enforcement policy would be exactly the same as if avoidance costs were considered explicitly.

Ironically, one of the practical advantages of considering avoidance costs in optimal enforcement may be to reduce the importance of ambiguities in identifying gain and avoidance costs. The major result of considering avoidance costs as part of the analysis is to eliminate any vestige of significance to the use of gain as a datum for setting penalties. The optimal solution is simply to set the expected penalty equal to external harm. Wading through the ambiguities of gain therefore becomes unnecessary. Similarly, under the simple case I have been considering -- which implicitly assumes that avoidance costs are perfectly responsive to the expected penalty -- avoidance costs also may "fall out" of the optimal solution.

5. The Generality of Avoidance Costs

So far, I have given only one concrete example of avoidance costs to individuals -- driving around the block

to avoid double parking. In fact, most avoidance costs are likely to be far more subtle and pervasive.

Extending the double parking example, individuals may incur avoidance costs long before they find themselves circling the block. They may literally "avoid" the violation by avoiding the congested area altogether. Or they may use an alternative means of transportation. Whatever they do, anyone who changes their behavior on the basis of the double-parking prohibition has incurred a cost, in terms of a lost opportunity. Whenever that occurs, there has been a loss in social welfare that ought to be considered by enforcement policy.

In that view, avoidance cost is a completely general phenomenon that occurs to some extent whenever any legal prohibition is enforced. So long as people respond to the incentives produced by enforcement, there will be avoidance costs. Recognizing that there are avoidance costs is not a criticism of legal prohibitions generally, or any one in particular. The law in question may be highly desirable, in terms of the harmfulness of the conduct prohibited. In that case, a high level of avoidance cost also will be desirable under optimal penalties. But even where the social benefits of the law are large, optimal

enforcement cannot be achieved by ignoring the associated costs.

6. Conclusions

The analysis developed in this paper is simply an application of the generally recognized point that laws have costs as well as benefits. Heretofore, the optimal penalty literature has not fully taken account of this fact, by failing to consider non-violators' avoidance costs.

The most general result of considering avoidance costs is to redefine a welfare-enhancing violation as one where the sum of the gain plus avoidance costs exceeds external harm. Because avoidance costs vary with the expected penalty level, this result does not necessarily imply a change in the optimal level of the expected penalty.

However, the existence and variability of avoidance costs suggest that there is no a priori basis for determining the optimal number of violations merely by considering the level of gain, or the difference between gain and harm. Even if we were fairly certain that a particular violation never created more gain than harm, or indeed that it never produced any gain at all, we still

could not conclude that the optimal number of violations was zero. Instead, the preferred course is the conventional solution of setting the expected penalty equal to external harm, and allowing the penalty system to determine the optimal number of violations through individuals' comparisons of their costs and benefits of compliance or violation.