Journal of Public Economics 46 (1991) 291–306. North-Holland

# A model of optimal fines for repeat offenders

A. Mitchell Polinsky

*Stanford University, Stanford, CA 94305, USA, and National Bureau of Economic Research*

Daniel L. Rubinfeld*

*University of California, Berkeley, CA 94720, USA*

Received September 1989, revised version received March 1991

This paper analyzes optimal fines in a model in which individuals can commit up to two offenses. The fine for the second offense is allowed to differ from the fine for the first offense. There are four natural cases in the model, defined by assumptions about the gains to individuals from committing the offense. In the case fully analyzed it may be optimal to punish repeat offenders more severely than first-time offenders. In another case, it may be optimal to impose less severe penalties on repeat offenders. And in the two remaining cases, the optimal penalty does not change.

## 1. Introduction

It is a common practice to punish repeat offenders more severely for the same offense than first-time offenders. One influential study of sentencing, by von Hirsch (1976, p. 84), describes this policy as follows:

> In the American criminal justice system, and in most others with which we are familiar, an offender's record of previous convictions considerably influences the severity with which he is punished. The first offender can expect more lenient treatment than the repeater.

In a similar vein, guidelines promulgated by the United States Sentencing Commission (1989, p. 4.1) state that 'A defendant's record of past criminal conduct is directly relevant to [the purposes of sentencing].' These guidelines also provide for higher sentences for repeat offenders.

*Polinsky's research was supported by the John M. Olin Program in Law and Economics at Stanford Law School and Rubinfeld's research was supported by the National Science Foundation Programs in Law and Social Science and Economics. During the writing of this paper we benefited from conversations or correspondence with Michael Block, Rayner Cheung, Michael Davis, Hugh Gravelle, Louis Kaplow, Paul Rubin, Ariel Rubinstein, Steven Shavell, Edward Sherry, George Stigler, and two referees.

In light of this practice, it is perhaps surprising that there is relatively little discussion in the literature on deterrence concerning the optimal policy for punishing repeat offenders.[1] This paper adds to that discussion by developing a simple model in which individuals can commit up to two offenses and are subject to a fine when they are caught. The fine for the second offense is allowed to differ from the fine for the first offense. Individuals take into account that the decision to commit the first offense may affect the size of the fine imposed on them if they later commit a second offense.

It will be seen that there are four natural cases within the model, defined by assumptions about the gains to individuals from committing the harmful act. In the case fully analyzed it may be optimal to punish repeat offenders more severely than first-time offenders. In another case, illustrated by a numerical example, it may be optimal to impose less severe penalties on repeat offenders. And in the two remaining cases, the optimal policy is to keep the penalty the same. The rest of the introduction focuses on the intuition behind the first case.

At first glance, it might appear that the analysis of optimal penalties for repeated offenses would not differ from the analysis for a single offense. If the penalty is set optimally with respect to the first offense, and the harm caused by the second offense is the same, there is no apparent reason to change the penalty for the second offense. This argument suggests that, if multiple offenses are possible, a uniform sanction would be optimal (with the level of the sanction determined by the conventional one-offense analysis).

The potential superiority of a policy of increasing penalties for repeat offenders can be explained in the following way. Imagine some characteristic of an individual – other than the socially-acceptable gain he would get from engaging in the harm-creating activity – that affects his propensity to engage in the activity.[2] Call this the individual's 'offense propensity'. For example,

[1]In the modern economic literature on deterrence, Stigler (1970, pp. 528–529) was the first person to discuss informally why first offenders should be punished more leniently than repeat offenders. There also is an interesting informal discussion by Posner (1985, pp. 1216–1217; 1986, pp. 213–215). To our knowledge, only Rubinstein (1979, 1980) has formally analyzed optimal penalties in a dynamic model with repeat offenses. His first paper shows how prior offenses should be taken into account in deciding whether to impose any punishment at all; he does not consider whether the *level* of punishment should depend on the number of prior offenses. His second paper demonstrates that there exists a utility function for which a policy of imposing higher penalties on repeat offenders increases deterrence. Landsberger and Meilijson (1982) also develop a dynamic model with repeat offenses, but their concern is with how prior offenses should affect the probability of detection rather than the level of punishment. Keenen and Rubin (1986) analyze a static model that tangentially relates to repeated offenses, but they acknowledge that 'what is apparently needed [to study repeat offenders] is a dynamic model' (p. 14). The focus of their paper is on how the optimal penalty varies with the severity of the harm.

[2]The reason for excluding his socially-acceptable gain (any gain that counts towards the determination of social welfare) in the statement in this sentence is explained in comment (a) in section 4 below.

an individual's offense propensity might be associated with his obtaining an illicit gain from committing the offense (say he receives pleasure from causing harm to others). The higher is his illicit gain, the more likely he is to engage in the harm-creating activity and the higher is his offense propensity.

To optimally deter individuals – that is, to deter them if and only if their socially-acceptable gain is less than the harm caused – it is necessary to make the punishment an increasing function of the individual's offense propensity. Otherwise, individuals with relatively high offense propensities (for example, high illicit gains) might be underdeterred – they might engage in the activity even when their socially-acceptable gains are less than the harm. And individuals with relatively low offense propensities (low illicit gains) might be overdeterred – they might not engage in the activity even when their socially-acceptable gains exceed the harm.

But suppose that the enforcement authority cannot observe individuals' offense propensities. This is a reasonable assumption, for example, if variations in offense propensities are due to variations in illicit gains. Obviously, the enforcement authority then cannot make the penalty depend on an individual's offense propensity.

*Imposing higher penalties on repeat offenders is an indirect way of imposing higher penalties on individuals with higher offense propensities.* To see why, observe that individuals with relatively high offense propensities will be more likely to engage in the harmful activity the first time, everything else equal. Thus, it can be inferred that individuals who are caught having engaged in the harmful activity a *second time* are, on average, those with relatively high offense propensities. For reasons discussed above, to properly deter such individuals, a higher penalty is required.[3]

This result can be reconciled with the intuitive argument discussed initially that led to the conclusion that a uniform penalty policy is best. The essence of that argument was that if the penalty is set optimally with respect to the first offense, and the harm caused by the second offense is the same, there is no reason to change the penalty for the second offense. It is now clear that the *implicit assumption* in that argument is that the consideration of offense propensity is irrelevant (for example, illicit gains are zero for everyone). Then a uniform penalty policy based on harm alone is optimal. But if offense propensities are relevant, increasing penalties might be better.

Section 2 analyzes the model in the case in which an increasing fine policy may be optimal. Section 3 briefly discusses the three remaining cases and

[3]This rationale for imposing higher penalties on repeat offenders is analogous to the rationale for charging an insured a higher premium after a claim is made ('experience rating') and for rewarding a worker with higher pay after a successful outcome ('merit rating'). See, for example, Venezia and Levy (1983) and Viscusi (1986).

It will become clear in section 3 below that the explanation provided here requires that individuals' offense propensities are fixed and that their socially-acceptable gains are stochastic.

provides a numerical example in which optimal fines decrease. Section 4 contains some concluding comments.

## 2. Optimal fines in a two-offense model

To make the distinction between the first and the second offense meaningful, it is assumed that the offenses are committed sequentially. Reference will be made to a 'first period' – during which at most one offense can be committed – and to a 'second period' – also during which at most one offense can be committed. Because time per se is not essential to the analysis, it is assumed that the discount rate is zero.

All individuals are risk neutral. An individual's gain from engaging in the harmful activity is the sum of two components: a socially-acceptable gain, which is included in social welfare, and an illicit gain, which is not included.[4] It is assumed for simplicity that there are two possible levels of the acceptable gain and two possible levels of the illicit gain.

In the case analyzed in this section, an individual's acceptable gain in each period is stochastic – but revealed to him at the beginning of the period – while his illicit gain is fixed (and invariant from period to period). For example, a driver may obtain a socially-acceptable gain from driving fast that depends on circumstances – whether rushing to a hospital or a golf game – whereas he may obtain an 'illicit thrill' from recklessly endangering others that does not depend on the circumstances. Section 3 discusses three other cases in which different assumptions are made about whether the acceptable and the illicit gains are stochastic or fixed.

In general, the enforcement authority chooses both the probability of detection and the fine for first and second offenders. Because the principal concern of this paper is with the structure of fines, it is assumed that the probability of detection is fixed and, without loss of generality, equal to one.[5]

Social welfare is defined to be the sum over both periods of the aggregate level of acceptable gains less the aggregate harm. (Because the probability of detection is fixed, enforcement costs will be ignored.) Obviously, the *first-best*

[4]In Becker's (1968) article on crime and punishment, all gains were treated as socially acceptable. However, in response to that article, Stigler (1970, p. 527) wrote: 'The determination of this social value [of the gain to offenders] is not explained, and one is entitled to doubt its usefulness as an explanatory concept: what evidence is there that society sets a positive value upon the utility derived from a murder, rape, or arson? In fact the society has branded the utility derived from such activities as illicit.'

[5]To the extent that the investment in enforcement effort applies to a wide range of harms, it is appropriate to treat the probability of detection with respect to any one type of harm as fixed. See generally Shavell (1991). If the probability of detection were less than one, the only effect on the results would be that the optimal fines would have to be raised (assuming that the wealth constraint of individuals is not binding) so that the expected fines equal those discussed in this paper.

*outcome* is that only individuals with acceptable gains greater than the harm engage in the activity. If this outcome is achieved, potential injurers will be said to be *optimally deterred.*

The following notation will be used:

$h$ = harm if an individual engages in the activity,
$a_1, a_2$ = levels of acceptable gain, with $0 \leqq a_1 < h < a_2$,
$b_1, b_2$ = levels of illicit gain, with $0 \leqq b_1 < b_2$,
$\alpha$ = probability in each period that an individual's acceptable gain is $a_1$ $(0 < \alpha < 1)$,
$\beta$ = fraction of individuals with illicit gain $b_1$ $(0 < \beta < 1)$,
$f$ = fine imposed on an individual who engages in the activity.

Clearly, $a_1 + b_1 < a_1 + b_2$ and $a_2 + b_1 < a_2 + b_2$. Whether $a_1 + b_2$ is less than or greater than $a_2 + b_1$ requires a further assumption.[6] If

$$a_1 + b_2 < a_2 + b_1, \tag{1}$$

then the enforcement authority could achieve the first-best outcome by choosing a *uniform* fine between $a_1 + b_2$ and $a_2 + b_1$; such a fine would deter individuals with a low acceptable gain even if their illicit gain is high, but would not deter individuals with a high acceptable gain even if their illicit gain is low. Therefore, to make the study of optimal fines interesting, it is assumed that

$$a_2 + b_1 < a_1 + b_2. \tag{2}$$

Given (2), there are three relevant ranges of the fine, referred to as 'low', 'moderate', and 'high' fines and denoted $f_l$, $f_m$, and $f_h$, respectively. The three ranges are defined as follows:[7]

$$a_1 + b_1 < f_l < a_2 + b_1 < f_m < a_1 + b_2 < f_h < a_2 + b_2. \tag{3}$$

It can be shown that $f_m$ is inferior to both of the other fines. Relative to $f_l$, it overdeters some individuals with high acceptable gains; and relative to $f_h$, it underdeters some individuals with low acceptable gains.

Thus, there are four possible fine policies that could be chosen by the enforcement authority, depending on whether a low fine or a high fine is imposed for the first offense, and on whether a low fine or a high fine is imposed for the second offense:

[6]The special case in which $a_1 + b_2 = a_2 + b_1$ will not be considered.
[7]It can be shown that a fine less than $a_1 + b_1$ or greater than $a_2 + b_2$ never is optimal.

$$\begin{array}{ll} \text{'low–low'} & (f_l, f_l), \\ \text{'high–high'} & (f_h, f_h), \\ \text{'low–high'} & (f_l, f_h), \\ \text{'high–low'} & (f_h, f_l). \end{array} \tag{4}$$

Before considering the behavior of individuals in response to the enforcement authority's policy choice, it will be useful to define 'types' of individuals in terms of the revealed value of their acceptable gain in the first period and their illicit gain. An individual with an acceptable gain of $a_1$ in the first period and an illicit gain of $b_1$ will be referred to as an '$(a_1, b_1)$ type'; other types will be defined similarly. The gain and fraction of each type are (assuming the acceptable gains and the illicit gains are uncorrelated):

$$\begin{array}{ll} \textit{Gain} & \textit{Fraction} \\ a_1+b_1 & \alpha\beta \\ a_1+b_2 & \alpha(1-\beta) \\ a_2+b_1 & (1-\alpha)\beta \\ a_2+b_2 & (1-\alpha)(1-\beta) \end{array} \tag{5}$$

Total population will be normalized to equal one.

There are potentially sixteen cases to analyze – the behavior of each of the four types of individuals defined by (5) under each of the four possible fine policies defined by (4). Because it would be tedious to examine all of these cases explicitly, only one case will be analyzed here.

Consider the behavior of an $(a_1, b_2)$ type under the low–high $(f_l, f_h)$ fine policy. By definition, an individual of this type has a high illicit gain, $b_2$, in both periods and a low acceptable gain, $a_1$, in the first period. In the second period, with probability $\alpha$ he again will have a low acceptable gain, but with probability $(1-\alpha)$ he will have a high acceptable gain. The individual must decide whether to engage in the harmful activity in the first period and in the second period.

Suppose the individual engages in the harmful activity in the first period. He then will face a high fine $f_h$ if he engages in the activity in the second period. If his acceptable gain in the second period is $a_1$, he will not engage in the second period because $a_1+b_2<f_h$ (see (3)). However, if his acceptable gain in the second period is $a_2$, he will engage in the second period because

$a_2+b_2>f_h$. Therefore, if he engages in the first period, his expected gain net of his expected fine payment is

$$[(a_1+b_2)-f_l]+(1-\alpha)[(a_2+b_2)-f_h]. \tag{6}$$

The first term (in brackets) is his net gain in the first period and the second term is his expected net gain in the second period.

Suppose, alternatively, that the individual does not engage in the harmful activity in the first period. He then will face a low fine $f_l$ if he engages in the activity in the second period. Regardless of his acceptable gain in the second period, he will engage in the second period because $a_2+b_2>a_1+b_2>f_l$. Therefore, if he does not engage in the first period, his expected gain net of his expected fine payment is

$$\alpha[(a_1+b_2)-f_l]+(1-\alpha)[(a_2+b_2)-f_l]. \tag{7}$$

It is straightforward to show from the second-to-last inequality in (3) that (7) exceeds (6). In other words, an $(a_1,b_2)$ type subject to the $(f_l,f_h)$ fine policy will not engage in the harmful activity in the first period but will engage in the second period. Consequently, *social welfare* associated with an individual of this type under this fine policy is

$$\alpha(a_1-h)+(1-\alpha)(a_2-h). \tag{8}$$

The behavior of each type of individual under each possible fine policy can be derived in a similar fashion. The social welfare consequences are summarized in table 1. Thus, for example, in the row corresponding to the

Table 1

The social welfare implications of alternative fine policies, by type of individual.

| Individual | | Fine policy | | | |
|---|---|---|---|---|---|
| Type | Fraction | $f_l, f_l$ | $f_h, f_h$ | $f_l, f_h$ | $f_h, f_l$ |
| $a_1, b_1$ | $\alpha\beta$ | $(1-\alpha)(a_2-h)$ | 0 | $(1-\alpha)(a_2-h)$ | 0 |
| $a_1, b_2$ | $\alpha(1-\beta)$ | $(1+\alpha)(a_1-h)$ $+(1-\alpha)(a_2-h)$ | $(1-\alpha)(a_2-h)$ | $\alpha(a_1-h)$ $+(1-\alpha)(a_2-h)$ | $(1+\alpha)(a_1-h)$ $+(1-\alpha)(a_2-h)$ / $(1-\alpha)(a_2-h)$ |
| $a_2, b_1$ | $(1-\alpha)\beta$ | $(2-\alpha)(a_2-h)$ | 0 | $(a_2-h)$ | 0 / $(2-\alpha)(a_2-h)$ |
| $a_2, b_2$ | $(1-\alpha)(1-\beta)$ | $\alpha(a_1-h)$ $+(2-\alpha)(a_2-h)$ | $(2-\alpha)(a_2-h)$ | $(2-\alpha)(a_2-h)$ | $\alpha(a_1-h)$ $+(2-\alpha)(a_2-h)$ |

$(a_1, b_2)$ type of individual and the column corresponding to the $(f_l, f_h)$ fine policy, the entry is the one derived in (8) above. All of the other entries are self-explanatory except in two cases. Under the $(f_h, f_l)$ fine policy, the behavior of both the $(a_1, b_2)$ and $(a_2, b_1)$ types depends on the parameters of the model (see the appendix for details). In these cases there are two possible outcomes and therefore two possible levels of social welfare; both levels are shown in table 1.

*Aggregate* social welfare under each fine policy can be calculated from the information in table 1 by multiplying the fraction of each type of individual by the level of social welfare associated with that type. For example, under the $(f_l, f_h)$ fine policy aggregate social welfare is

$$\begin{aligned}&\alpha\beta(1-\alpha)(a_2-h)+\alpha(1-\beta)[\alpha(a_1-h)+(1-\alpha)(a_2-h)]\\&\quad+(1-\alpha)\beta(a_2-h)+(1-\alpha)(1-\beta)(2-\alpha)(a_2-h)\\&=\alpha^2(1-\beta)(a_1-h)+(1-\alpha)[2-\beta(1-\alpha)](a_2-h).\end{aligned} \tag{9}$$

Next consider whether a policy of increasing fines, decreasing fines, or uniform fines is optimal. It can be shown that a decreasing fine policy $(f_h, f_l)$ never is preferred in the case studied in this section. The proof of this result, which is complicated by the fact that individual behavior is indeterminate for the $(a_1, b_2)$ and $(a_2, b_1)$ types under the decreasing fine policy, is provided in the appendix.

Under the low uniform fine policy $(f_l, f_l)$ aggregate social welfare is

$$2\alpha(1-\beta)(a_1-h)+2(1-\alpha)(a_2-h), \tag{10}$$

and under the high uniform fine policy $(f_h, f_h)$ it is

$$2(1-\alpha)(1-\beta)(a_2-h). \tag{11}$$

A comparison of the levels of aggregate social welfare under the three relevant policies – see (9), (10), and (11) – leads directly to the conclusion that: *The increasing fine policy $(f_l, f_h)$ is superior to both uniform fine policies, $(f_l, f_l)$ and $(f_h, f_h)$, if and only if*

$$\frac{a_1(1-\beta)\alpha(2-\alpha)+a_2\beta(1-\alpha)^2}{(1-\beta)\alpha(2-\alpha)+\beta(1-\alpha)^2}<h<\frac{a_1(1-\beta)\alpha^2+a_2\beta(1-\alpha^2)}{(1-\beta)\alpha^2+\beta(1-\alpha^2)}. \tag{12}$$

The first inequality follows from the circumstances under which (9) is greater than (10), and the second inequality follows from the circumstances under

which (9) is greater than (11). If $h$ is less than the left-hand ratio in (12), the low uniform fine policy $(f_l, f_l)$ is superior to the others, while if $h$ is greater than the right-hand ratio in (12), the high uniform fine policy $(f_h, f_h)$ is preferred.[8]

To understand intuitively why each of the fine policies might dominate, consider how each policy does relative to the first-best outcome. In the first-best outcome, individuals with low acceptable gains are deterred from engaging in the harmful activity in both periods, and individuals with high acceptable gains are induced to engage in the harmful activity in both periods. The individuals' illicit gains are irrelevant.

First consider how the low uniform fine policy $(f_l, f_l)$ does relative to the first-best outcome. Because the fine does not vary with the number of offenses, an individual's decision whether to engage in the harmful activity does not depend on the probability distribution of his acceptable gain in subsequent periods. Thus, it is straightforward to see from (3) that, with a fine corresponding to $f_l$, this policy never causes overdeterrence but does lead to underdeterrence when an individual's acceptable gain is $a_1$ and his illicit gain is $b_2$. This underdeterrence occurs among all $(a_1, b_2)$ types in the first period, certain $(a_1, b_2)$ types in the second period (those whose second-period acceptable gains remain at $a_1$), and certain $(a_2, b_2)$ types in the second period (those whose second-period acceptable gains become $a_1$).

Similarly, it can be seen from (3) that the high uniform fine policy $(f_h, f_h)$ never causes underdeterrence but does lead to overdeterrence when an individual's acceptable gain is $a_2$ and his illicit gain is $b_1$. This overdeterrence occurs among all $(a_2, b_1)$ types in the first period, certain $(a_2, b_1)$ types in the second period (those whose second-period acceptable gains remain at $a_2$), and certain $(a_1, b_1)$ types in the second period (those whose second-period acceptable gains become $a_2$).

The increasing fine policy $(f_l, f_h)$ eliminates some of the underdeterrence that occurs under the low uniform fine policy as well as some of the overdeterrence that occurs under the high uniform fine policy. However, the increasing fine policy still leaves some underdeterrence – among $(a_1, b_2)$ types whose second-period acceptable gains remain at $a_1$ – and some overdeterrence – among $(a_2, b_1)$ types whose second-period acceptable gains remain at $a_2$.

This discussion shows why each of the three fine policies can be the preferred one. If the principal concern is with eliminating overdeterrence, the low uniform fine policy is the most desirable policy; and if the main concern is with eliminating underdeterrence, the high uniform fine policy is the most desirable one. In some circumstances, however, partially reducing both

[8]Because the ratios in (12) are arithmetic weighted averages of $a_1$ and $a_2$, the left-hand ratio is greater than $a_1$ and the right-hand ratio is less than $a_2$. In addition, it can be demonstrated that the left-hand ratio is less than the right-hand ratio.

overdeterrence and underdeterrence by using the increasing fine policy provides a better compromise.

The argument for increasing fines discussed in the introduction can be illustrated in the two-offense model. The essence of that argument was that individuals with high offense propensities were, everything else equal, more likely to engage in the harmful activity. Under the low uniform fine policy, the fraction of offenders in the first period with high illicit gains is $(1-\beta)/(1-\alpha\beta)$, which exceeds the fraction of individuals in the general population with high illicit gains, $(1-\beta)$. And under the high uniform fine policy, all offenders in the first period have high illicit gains. Thus, if a uniform fine policy is used, individuals who commit offenses in the first period are more likely than average to have high illicit gains.

Under the increasing fine policy, however, the fraction of offenders in the first period with high illicit gains is $(1-\beta)$, which is smaller than the fraction under either uniform policy. Thus, the increasing fine policy can be viewed as a response to the disproportionate number of high illicit gainers who otherwise would engage in the activity under a uniform fine policy.

The potential advantage of an increasing fine policy can be illustrated by the following numerical example. Suppose the population is divided equally between individuals with low and high acceptable gains, and low and high illicit gains: $\alpha=\beta=1/2$. Also, suppose that the levels of the acceptable and the illicit gains are $a_1=\$5{,}000$, $a_2=\$10{,}000$, $b_1=\$15{,}000$, and $b_2=\$25{,}000$. Then the three fine ranges, corresponding to (3), are

$$\$20{,}000<f_{\mathrm{l}}<\$25{,}000<f_{\mathrm{m}}<\$30{,}000<f_{\mathrm{h}}<\$35{,}000. \tag{13}$$

The condition for the increasing fine policy to be superior to the others, corresponding to (12), becomes

$$\$6{,}250<h<\$8{,}750. \tag{14}$$

Suppose, for example, that $h=\$7{,}500$. Then, using (9), (10), and (11), it is straightforward to compute that aggregate social welfare under the increasing fine policy is \$1,875, whereas aggregate social welfare under each of the uniform fine policies is \$1,250. Thus, in this example the optimal policy would be to impose a fine of \$20,000 to \$25,000 for the first offense and a fine of \$30,000 to \$35,000 for the second offense.

## 3. The choice of assumptions within the model

A key assumption in the model in section 2 was that individuals' socially-acceptable gains are stochastic, while their illicit gains are fixed. There are three natural alternatives to this assumption: (a) that both types of gain are

stochastic; (b) that both are fixed; or (c) that acceptable gains are fixed and illicit gains are stochastic. This section briefly considers the optimal fine policy in these cases.

If both types of gain are stochastic, then second offenders would not differ systematically from first offenders, and there would be no reason to punish them differently. If both types of gain are fixed, then although second offenders would differ from first offenders – they would on average have higher illicit gains and higher acceptable gains – the differences do not imply that higher fines for repeat offenders are desirable. Essentially, this is because the marginal costs and benefits of raising the fine for second offenders above that for first offenders have not changed.

Suppose, however, that acceptable gains are fixed and illicit gains are stochastic. For example, a driver may attach an approximately constant value to the time saved from driving fast – the acceptable gain – whereas his illicit thrill may depend on who is with him in the car to be impressed, his mood, etc.[9] Then for reasons that are essentially the reverse of those discussed in the previous section, it can be demonstrated that a decreasing fine policy may be optimal (and that an increasing fine policy never is desirable).

To see why decreasing fines might be beneficial, consider starting with a high uniform fine policy. Then, in both periods, only individuals with high acceptable gains and high illicit gains will engage in the harmful activity; individuals with high acceptable gains and low illicit gains will be overdeterred. Now suppose instead that a high fine is imposed on first offenders and a low fine on second offenders, and assume that in the first period only individuals with high acceptable gains and high illicit gains engage in the harmful activity.[10] This decreasing fine policy has, by assumption, the same effect as the high uniform fine policy in the first period. But it improves welfare in the second period: since all potential second offenders are individuals with high acceptable gains, by lowering the fine for second offenders some individuals who would have been overdeterred because their illicit gain in the second period is low can be encouraged to engage in the activity.

The numerical example at the end of section 2 can be used to illustrate the potential superiority of a decreasing fine policy. Let the parameter values be the same as before (with $h = \$7{,}500$), but now assume that the acceptable gains are fixed and the illicit gains are stochastic. Under a decreasing fine policy, the behavior of certain types of individuals depends on the particular levels of the high fine $f_h$ and the low fine $f_l$. (This complication was ignored

[9]This example was suggested by Louis Kaplow.

[10]Under a decreasing fine policy, other types of individuals also may engage in the first period depending on the parameter values and the levels of the fines. The assumption made in the text simplifies the discussion.

in the example in section 2 because, under the assumption there, the decreasing fine policy was dominated by one of the other policies.) Thus, for concreteness, suppose $f_h = \$34{,}000$ and $f_l = \$24{,}000$. It can then be calculated that social welfare under a decreasing fine policy is \$1,562, whereas social welfare under each of the uniform fine policies is \$1,250, and social welfare under an increasing fine policy is \$625.

## 4. Concluding remarks

This section contains comments about the exclusion of socially-acceptable gains as a basis for offense propensities; generalizations of the model; alternative bases for offense propensities; and another deterrence rationale for punishing repeat offenders differently.

(a) *The exclusion of socially-acceptable gains as a basis for offense propensities.* In section 1 it was noted that an individual's socially-acceptable gain could not serve as the basis for his offense propensity (as that term is used in this paper). At first glance this may seem peculiar because socially-acceptable gains vary among individuals, affect their decisions whether to engage in harm-creating activities, and may be difficult for an enforcement authority to observe. However, if individuals differ only in terms of their socially-acceptable gains, optimal deterrence can be achieved by a penalty that is based solely on the harm caused and, therefore, that does not change with the number of offenses. For if individuals are made to take into account the harm caused, they will engage in the harmful activity if and only if their socially-acceptable gains exceeds the harm. This claim is not valid if individuals differ in other ways that affect their propensity to engage in the harmful activity.

(b) *Generalizations of the model.* There are two natural ways to generalize the model. First, both the socially-acceptable gains and the illicit gains could be made continuous rather than discrete. When this generalization was attempted, the model became analytically intractable. However, we believe that if the model were generalizable along these lines, results similar to those discussed in sections 2 and 3 would occur. A second generalization of the model would be to allow the number of possible offenses to be greater than two. Our conjecture is that in a model with more than two offenses the optimal fine would strictly increase with the number of offenses in the case studied in section 2 (provided an increasing fine policy is preferred), and strictly decrease in the reciprocal case.[11]

[11]Another generalization would be to allow for the possibility of false convictions. In the conventional analysis of deterrence in which only one offense is considered, false convictions reduce deterrence because the incremental cost of becoming an offender is reduced. In the repeat offense context, there may be an opposing tendency if a false conviction raises the fine that a potential offender would face in the future, and a reinforcing tendency if it lowers the fine. Also, Rubinstein (1979) has shown that if individuals can be found liable by mistake, it is optimal to punish an individual 'only if his long-run record is "unreasonably" bad' (p. 407).

(c) *Alternative bases for offense propensities.* In this paper, variations in offense propensities were attributed to variations in illicit gains. There are at least two other plausible sources of variations in offense propensities. One is the pure disutility of time spent in jail. The lower this disutility, the higher the offense propensity. Another source of variation might be due to risk aversion. The less risk averse an individual, the more likely he is to engage in a harm-creating activity, and therefore the higher is his offense propensity. In each case additional complexity arises because of the need to include another factor in the determination of social welfare – either the disutility of time in jail or risk-bearing costs. It is not clear without further analysis whether the insights derived from the illicit-gain model would apply when variations in offense propensities are due to these sources.

(d) *Another deterrence rationale for punishing repeat offenders differently.* It was assumed throughout the paper that the probability of detecting a repeat offender is the same as for a first-time offender. More generally, however, this probability might be expected to change. For example, because repeat offenders are more experienced in committing offenses, they may be less likely to be detected. Then, to achieve optimal deterrence, it would be necessary to make the level of punishment rise with the number of offenses to make up for the declining probability of punishment. Conversely, repeat offenders may be more likely to be detected and punished – perhaps because the enforcement authority already has some information about them[12] – in which case the optimal penalty would fall with the number of offenses.

## Appendix

The appendix demonstrates that the decreasing fine policy $(f_h, f_l)$ is strictly inferior to the other policies when the socially-acceptable gains are stochastic and the illicit gains are fixed. This will be done by showing that either the low uniform fine policy $(f_l, f_l)$ or the high uniform fine policy $(f_h, f_h)$ leads to a higher level of aggregate social welfare.

Under the decreasing fine policy there are two possible outcomes for both the $(a_1, b_2)$ and the $(a_2, b_1)$ types. An $(a_1, b_2)$ type of individual either will engage in the harmful activity in both periods (if $(1+\alpha)(a_1+b_2) < f_l + \alpha f_h$), in which case the top entry in the corresponding box in table 1 applies; or he will not engage in the first period, but will engage in the second period if his acceptable gain is high, in which case the bottom entry applies. Similarly, an $(a_2, b_1)$ type of individual either will not engage in the first period and will engage in the second period if his acceptable gain is high (if

[12]See Stigler (1970, p. 530). Also, Posner (1985, p. 1216, n. 43) has pointed out that 'a previous offender is easier to convict than a first offender, because if he takes the stand the prosecution can introduce his record of convictions to try to undermine his credibility.'

$(2-\alpha)(a_2+b_1)>(1-\alpha)f_l+f_h)$, in which case the top entry applies; or he will not engage in either period, in which case the bottom entry applies.[13]

It will be useful first to show that under the decreasing fine policy only two of the four combinations of these outcomes are possible: the top entry for the $(a_1, b_2)$ types combined with the bottom entry for the $(a_2, b_1)$ types; or the bottom entry for the $(a_1, b_2)$ types combined with the top entry for the $(a_2, b_1)$ types.

The argument involves a proof by contradiction. Assume initially that the top entry applies for both the $(a_1, b_2)$ and the $(a_2, b_1)$ types, that is,

$$(1+\alpha)(a_1+b_2)<f_l+\alpha f_h \tag{A.1}$$

and

$$(2-\alpha)(a_2+b_1)>(1-\alpha)f_l+f_h. \tag{A.2}$$

It will be shown that (A.1) and (A.2) cannot hold simultaneously. Recall from (5) that

$$a_1+b_1<f_l<a_2+b_1<a_1+b_2<f_h<a_2+b_2. \tag{A.3}$$

Let

$$f_l=a_1+b_1+\theta_1(a_2-a_1), \tag{A.4}$$

where $0<\theta_1<1$, and

$$f_h=a_1+b_2+\theta_2(a_2-a_1), \tag{A.5}$$

where $0<\theta_2<1$. Substituting (A.4) and (A.5) into (A.1) and (A.2) and rewriting yields:

$$b_2-b_1<(a_2-a_1)(\theta_1+\alpha\theta_2) \tag{A.6}$$

and

$$(a_2-a_1)(2-\alpha-(1-\alpha)\theta_1-\theta_2)>b_2-b_1. \tag{A.7}$$

Note that (A.3) implies that $(b_2-b_1)>(a_2-a_1)$. Therefore, for (A.6) and (A.7) to hold, it must be that

$$\theta_1+\alpha\theta_2>1 \tag{A.8}$$

and

$$(1-\alpha)\theta_1+\theta_2<1-\alpha. \tag{A.9}$$

Multiplying both sides of (A.8) by $(1-\alpha)$ and comparing the resulting inequality to (A.9) shows the contradiction. A similar argument also can be used to demonstrate that the two bottom entries cannot hold simultaneously.

To distinguish between the two mutually exclusive combinations that can occur under the decreasing fine policy, let Case 1 refer to the situation in which the top entry for the $(a_1, b_2)$ types and the bottom entry for the $(a_2, b_1)$ types applies. Similarly, let Case 2 refer to the situation in which the bottom entry for the $(a_1, b_2)$ types and the top entry for the $(a_1, b_2)$ types applies.

In Case 1 aggregate social welfare under the decreasing fine policy is (see table 1)

$$2\alpha(1-\beta)(a_1-h)+(1-\alpha)(2-\alpha\beta)(a_2-h). \tag{A.10}$$

A comparison of (A.10) with (10) shows that aggregate social welfare under the low uniform fine policy is greater (since $2-\alpha\beta<2$).

Similarly, in Case 2 aggregate social welfare is

$$\alpha(1-\alpha)(1-\beta)(a_1-h)+2(1-\alpha)(1-\beta)(a_2-h). \tag{A.11}$$

A comparison of (A.11) with (11) shows that aggregate social welfare under the high uniform fine policy is greater (since $a_1-h<0$).

Thus, the decreasing fine policy always is inferior to one of the uniform fine policies.

## References

Becker, Gary S., 1968, Crime and punishment: An economic approach, Journal of Political Economy 76, 169–217.

Keenen, Donald and Paul H. Rubin, 1986, The economics of evil deeds, unpublished manuscript.

Landsberger, Michael and Isaac Meilijson, 1982, Incentive generating state dependent penalty system: The case of income tax evasion, Journal of Public Economics 19, 333–352.

Posner, Richard A., 1985, An economic theory of the criminal law, Columbia Law Review 85, 1193–1231.

Posner, Richard A., 1986, Economic analysis of law, 3rd edn. (Little, Brown and Company, Boston).

Rubinstein, Ariel, 1979, An optimal conviction policy for offenses that may have been committed by accident, in: S.J. Brams, A. Schotter and G. Schwodiauer, eds., Applied game theory (Physica-Verlag, Wurzburg) 406–413.

Rubinstein, Ariel, 1980, On an anomaly of the deterrent effect of punishment, Economics Letters 6, 89–94.

Shavell, Steven, 1991, Specific versus general enforcement of law, Journal of Political Economy (in press).

Stigler, George J., 1970, The optimum enforcement of laws, Journal of Political Economy 78, 526–536.

United States Sentencing Commission, 1989, Guidelines manual (U.S. Government Printing Office, Washington, D.C.).

Venezia, Itzhak and Haim Levy, 1983, Optimal multi-period insurance contracts, Insurance: Mathematics and Economics 2, 199–208.

Viscusi, W. Kip, 1986, Moral hazard and merit rating over time: An analysis of optimal intertemporal wage structures, Southern Economic Journal 52, 1068–1079.

von Hirsch, Andrew, 1976, Doing justice: The choice of punishments (Hill and Wang, New York).