How Optimal Penalties Change with the Amount of Harm

Eric Rasmusen

Published: *International Review of Law and Economics* (1995) 15: 101-108.

Abstract

Intuition tells us that the optimal penalty and court care to avoid error should rise smoothly with the harm to the victim. This is not always correct: sometimes the optimal penalty and level of court care increase discontinuously with harm, even when penalties deter harm and court care reduces error continuously. This is shown in a model in which the social cost of crime consists of its direct harm, the cost of court care, and the cost of false convictions.

Draft: 4.1 (Draft 1.1, July 1991).

I would like to thank John Lott, Steven Shavell and two anonymous referees for helpful comments. Much of this work was completed while the author was Olin Faculty Fellow at Yale Law School and on the faculty of UCLA's Anderson Graduate School of Management.

2000: Eric Rasmusen, Professor of Business Economics and Public Policy and Sanjay Subhedar Faculty Fellow, Indiana University, Kelley School of Business, BU 456, 1309 E 10th Street, Bloomington, Indiana, 47405-1701. Office: (812) 855-9219. Fax: 812-855-3354. Erasmuse@indiana.edu. Php.indiana.edu/~erasmuse.

1. Introduction

Should a crime's penalty rise smoothly in proportion to the crime's harmfulness? This seems obvious, and has a sound economic intuition behind it. The optimal penalty is the result of a tradeoff between the penalty's costs and benefits, and when tradeoffs are made, they usually change smoothly. If the harm increases slightly, then so, it seems, should the penalty. If the penalty is imprisonment, then too short a prison term results in too much crime and too long a prison term results in excessive spending on prisons. A similar argument can be made to show that the care the court takes with a particular case should be smoothly increasing in the harmfulness of the crime involved.

This intuition is deceiving: penalties and court care should not always rise smoothly with harm. The model below will formalize the idea that the optimal penalty and court care involve tradeoffs between the penalty's cost and benefit. The penalty and court care will not actually decline as harm increases, but they may jump sharply, even when the harm increases smoothly.

The reason for the discontinuous jumps is that increasing the penalty has the good effect of reducing crime but the bad effect of increasing the penalty costs on those criminals it still fails to deter. This is the idea that Louis Kaplow (1990a, 1990b) has used to show why optimal costly penalties will sometimes be either zero or maximal. Using a model based on Polinsky & Shavell (1984), he showed that the social optimization problem is not convex and so may have corner solutions. The penalty might, for example, have very weak deterrence value, so increasing it beyond zero would increase social expenditure without much reducing crime. Or perhaps as the penalty becomes more severe, it needs to be carried out so much less often that the total cost falls and the optimal penalty is maximal. Because of these corner solutions, there will be some harm level at which the corner solution jumps from zero to maximal, but for the most part the crime's penalty should be unrelated to its harmfulness.

The model in the present article will be somewhat different. Standard assumptions will be made to rule out the corner solutions of zero or maximal optimal penalties, the social cost of penalties will arise from false convictions, which can be reduced by greater court care, and the focus will be on how penalties and court care change with harm.

2. The Model

Let a certain type of crime cause harm h.[1] The amount $n(p)$ of this crime depends on the expected penalty, p. The probability of conviction will be assumed to be exogenous, so p represents the penalty itself.[2] Let the deterrence function $n(p)$ satisfy $n_p < 0$, where $Lim(n_p) = -\infty$ as $p \to 0$, and $n > \varepsilon > 0$ for some constant ε. Under these assumptions, additional punishment always deters more crime but never drives it below ε, and the marginal deterrence is infinitely high starting from zero punishment.[3]

For each harm level h, society chooses the punishment, p, and the care that courts take to avoid false convictions, c. Depending on the punishment and court care, society incurs a false-conviction cost $f(p, c)$ per crime, where $f_p > \delta > 0$ for some constant δ, $f_c < 0$, $Lim(f_c) = -\infty$ as $c \to 0$, and $f_{pc} < 0$ for all $p > 0$. Under these assumptions, greater punishment increases the cost of false conviction by at least δ, greater court care reduces false conviction and has infinite marginal benefit starting from a level of zero care, and the marginal benefit from care is greater if the penalty is greater.[4] The cost of

[1]The harm h could be the direct harm to the victim, or it might add the precautions of potential victims and the effort of the criminal and subtract the crime's benefit to the criminal. See Ehrlich (1982) for a discussion of theories of the objectives of criminal punishment, and Friedman (1981) and Shavell (1985, p. 1244) for discussions of whether penalties should increase with the benefit to the criminal or the harm to the victim.

[2]If p were assumed to have a direct social cost, it could represent the enforcement level as well, and the model would be little changed.

[3]It is assumed that $n > \varepsilon$ because otherwise the optimum might specify infinite penalties that eliminate crime completely— in which case there would be no false convictions either. The assumptions that n_p and f_c become infinite when p and c are zero exclude the opposite extreme: that punishment and court care are so ineffective that they should be abandoned altogether.

[4]For discussions of court error, see Png (1986), Posner (1973), Shavell (1987), and

false conviction, $f(p,c)$, includes such things as the disutility of those falsely convicted, the risk borne by those who fear they might be falsely convicted, the deterrence to efficient behavior created by fear of false punishment, and the public's discomfort in knowing that some punishment is mistaken.[5]

Society's problem is to choose p and c to minimize S, the sum of the social costs from h, the direct harm from the crime; f, the cost of false conviction; and c the cost court care to prevent false conviction. Summing these yields equation (1).

$$S = n(p)(h + f(p,c) + c). \tag{1}$$

Deterrence has two costs here: mistaken punishment and court care. Because of diminishing returns to court care, not enough will be spent to completely eliminate false convictions. Given this, the deterrence benefit from heavier penalties must be weighed against the false-conviction cost. The deterrence benefit is greater if the crime causes more harm, so for more harmful crimes the tradeoff will lead to greater penalties. This is a simultaneous system, so the greater penalties lead in turn to more spending to avoid false convictions when the crime's harm is greater. It can be shown, though I will not do so here, that not only does the most severe crime not receive an infinite punishment, but there is no "bunching" of penalties at that most serious crime: penalties and court care rise with harm.[6]

Rasmusen (1994); for more specific discussion of error avoidance as a goal of justice, see Ehrlich (1982), Rubinfeld and Sappington (1987), and Kaplow (1994). Note that in the present paper, unlike some of these studies, "court care" is the result of the court's decision, not of the litigants' decisions on how much to spend.

[5]To focus on the cost of false punishment, the model assumes that the public expense of correct punishment is zero— the punishment takes the form of Beckerian fines with zero transactions cost. This assumption is easily relaxed by adding another function increasing in $n(p)$ and p, but this would make little difference to the results.

[6]The proof that penalties and court care rise with harm rather than staying constant is available from the author. The analysis has implicitly assumed that each crime is independent, avoiding the issue of marginal deterrence: a higher penalty for crime h causing criminals to substitute to lesser offences, as described in Stigler (1970). Since marginal deterrence provides an incentive to steepen the punishment schedule rather than

Mathematically, the difficulty with this optimization problem is that the minimand is not convex. The government is trying to minimize the cost of crime, $S(p, c; h)$. If $S(p, c; h)$ were convex, optimization theory tells us that $p^*(h)$ and $c^*(h)$ would be continuous, but one condition for the convexity of $S(p, c; h)$ is that S_{pp} be positive, i.e., that

$$n_{pp}[h + f(p)] + nf_{pp} + 2n_p f_p > 0 \tag{2}$$

If inequality (2) is false, then the second-order-condition for the problem is not satisfied and the implicit function theorem cannot be used to show that the derivative p_h exists and is positive.[7] Inequality (2) can easily be false under the model's assumptions, which said nothing about the sign of n_{pp} or f_{pp}. There is no good reason for supposing that either n or f is concave or convex; they are more like demand functions, which might have any curvature, than like cost functions, which are usually convex. Adding arbitrary convexity assumptions to n and f would require adding the assumptions that $n_{pp} > 0$, $f_{pp} > 0$, $f_{cc} > 0$, and $f_{pp}f_{cc} - f_{pc}^2 > 0$. These assumptions imply that there are diminishing returns to deterrence and care, that false-conviction costs rise more than linearly with the penalty, and own-effects are stronger than cross-effects. But even these extra assumptions would not guarantee the validity of inequality (2), because though the first two terms would then be positive, the last term would still be negative. The first two terms would be positive because the marginal deterrence effect would weaken as the penalty increased, and the marginal cost of each false conviction would become greater. The last term would be negative because when crime fell, there would be less false conviction, which might outweigh the fact that with a higher penalty each instance would be more costly. As a result, the cost function would still fail to be convex, even though its component functions

smooth it, discontinuities would very likely continue to exist in the punishment schedule. The jury response to higher penalties is still another consideration (see Andreoni [1991]). A somewhat different issue is how the severity of harm should be related to the amount of enforcement effort; for analysis of this, see Shavell (1991) and Mookherjee and Png (1992).

[7]Thus, the approach used in Becker (1968) for comparative statics, which relied on special assumptions, would fail here.

would be well-behaved under the additional assumptions. To guarantee a continuous optimum it would be necessary to assume directly that (2) is true, which has no justification.

The next section of the paper will construct a simple example to show by construction that the optimal penalty and court care can be discontinuous in harm and to develop the intuition behind the outcome.

3. An Example with Discontinuous Penalty and Court Care

The following example will show why the optimal penalty and court care might be discontinuous in the crime's harm. Let the social cost of false conviction be

$$f(p,c) = p^2 + \frac{p^2}{c}, \tag{3}$$

which satisfies all the assumptions made earlier: the cost of false conviction is increasing at an increasing rate in the penalty, it falls in court care, and it cannot be completely eliminated, no matter how much court care is used.

The total social cost of crime, using equations (1) and (3), is

$$S = n(p)\left(h + p^2 + \frac{p^2}{c} + c\right). \tag{4}$$

The full optimization problem is to minimize S with respect to p and c, penalty and court care. It will be convenient to solve this in stages. Minimizing (4) with respect to c yields the first order condition

$$\frac{\partial S}{\partial c} = n(p)\left(-\frac{p^2}{c^2} + 1\right) = 0, \tag{5}$$

which implies that

$$c^* = p^*. \tag{6}$$

The optimal penalty depends on the specific functional form for the deterrence function $n(p)$. Let it be shaped as in Figure 1, which is drawn using the following function:

$$n(p) = 10 + 180/p \quad if \ p \leq 6$$
$$= 10 + 180/p + .1(p-6)^2 \quad if \ 6 < p \leq 12$$
$$= 10 + 180/p + .1(p-6)^2 - .5(p-12)^2 \quad if \ 12 < p \leq 20 \qquad (7)$$
$$= 3.1 + 2.1/(p - 19.4) \quad if \ 20 < p$$

This deterrence function says that some crime is deterred by a small penalty, very little more is deterred by a moderate penalty, and almost all crime disappears when the penalty is high.

FIGURE 1 GOES HERE

Equations (5) and (6), still apply, so $c^* = p^*$. Substituting this into the objective function (4) gives an objective function already optimized for c and now stated only in terms of p:[8]

$$S(p) = n(p)(h + p^2 + 2p) \qquad (8)$$

Figure 2 shows the shape of the objective function given the deterrence function from (7).

FIGURE 2 GOES HERE

The $S(p)$ function has two local minima. For very low harm ($h = 2$), the global minimum is clearly at a low penalty, and for very high harm ($h = 50$) it is clearly at a high penalty. For $h = 15$ and $h = 16$, the two minima are very close in their social cost. If $h = 15$, the two minima are at (3.20, 2,096) and (21.90, 2,121) for *(penalty, social cost)*, so 3.20 is the optimal penalty. If $h = 16$, the two minima are at (3.29, 2,162) and (21.91, 2,125), so 21.91

[8]A substitution of this kind would mathematically incorrect if the analysis then proceeded to take a derivative with respect to p, but here the minima will be found by numerical methods.

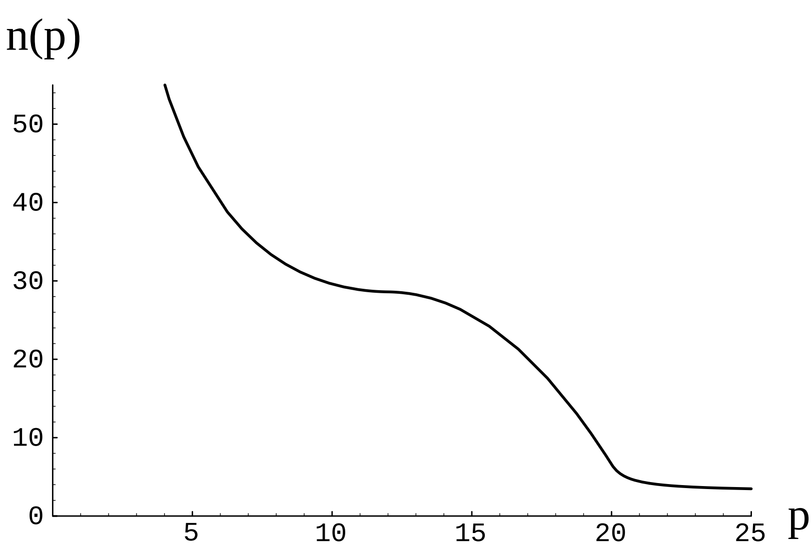

Figure 1: Crime as a Function of Expected Penalty

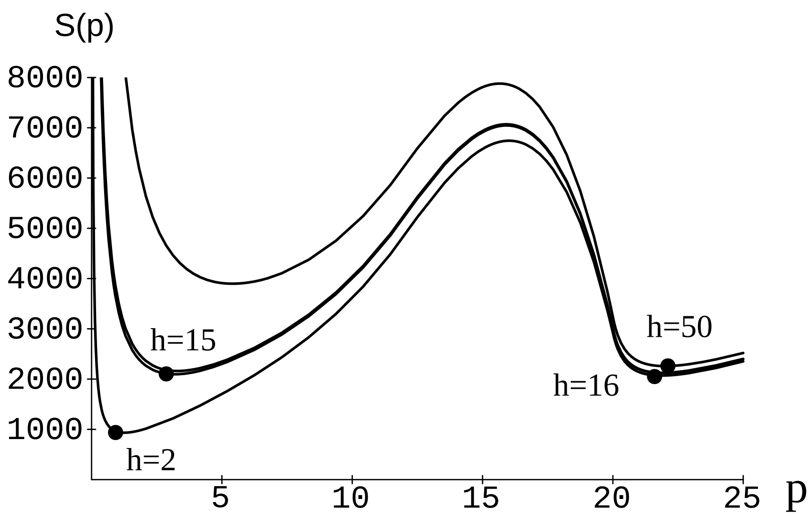

Figure 2: The Cost of Crime as a Function of the Expected Penalty

is the optimal penalty. A small change in the harm induces a jump in the optimal penalty from 3.20 to 21.91, and the care to avoid false conviction also jumps, since $c^* = p^*$ by (6). Despite these jumps, the social cost of a crime is a continuous function of its harm. The social cost increases only from 2,096 to 2,125 going from $h = 15$ to $h = 16$. The more harmful crime has much higher false-conviction and court costs, but is offset almost exactly by a decline in the amount of crime.

A plausible general phenomenon lies behind the example. For low levels of harm, what matters most is to have a small penalty that deters some crime but keeps the cost of false convictions low. Increasing the penalty beyond that low level has little additional deterrent effect. If the crime's harm becomes great enough, however, the penalty is drastically increased, because high enough penalties result in another large drop in the amount of crime. Optimal court care is discontinuous because it is based on the penalty: if the penalty is low, care has little benefit and should also be low, but once the penalty jumps, the benefit from increased care also jumps. The system is simultaneous but recursive: increased harm does not increase court care directly, but it increases the optimal penalty, which increases optimal court care and feeds back to the optimal penalty.

This story can be based on either of two interpetations of the deterrence function in Figure 1.

In the first interpretation, there always exist marginal offenders, indifferent about committing the crime, but most potential offenders fall into one of two criminal groups: casual criminals who are deterred by a relatively low penalty, and serious criminals who are deterred only when the penalty becomes very high. When the harm is small, the optimal penalty and care are small but positive, deterring a large number of offenses without incurring much cost from each false conviction. The number of offenses and false convictions is high, but since the harm per crime is small, these are tolerated. As the harm increases, the penalty and care increase slightly, but there is still no attempt to deter the serious offender. At some point, however, the crime has become harmful enough that it worth making the jump to increasing penalties so drastically as to deter the serious offender also. Since the penalty increases drastically, it is also worth increasing court care drastically.

In the second interpretation, a single individual decides how many crimes to commit. He will always commit more crimes if the penalty declines, but his crimes generally fall into two categories: crimes of opportunity, which require little thought, and planned crimes, which require considerable effort. A small penalty will deter the planned crime, but not the crime of opportunity. If the crime inflicts little harm, then the optimal penalty is set low and crimes of opportunity are tolerated, but beyond some threshold, the jump is made towards deterring both types of crime.

4. Concluding Remarks

Proportionality and optimal deterrence are incompatible aims of punishment. Optimal deterrence may require that a slightly more serious crime receives a much more severe sentence, if a sizeable increase in the penalty would have little additional deterrent effect, but would still increase the costs from false conviction. As the crime becomes more and more harmful, it is eventually worth increasing the penalty substantially for the sake of the additional deterrence, however, and there must exist some pair of almost identical crimes that spans the break-point.

Tort law is also concerned with choosing penalties to deter harmful behavior. As Png (1986) and Shavell (1987) note, excessive imposition of liability not only penalizes the innocent, but deters them from useful behavior that courts might confuse with wrongdoing. Tort law balances the false-liability costs of penalties against the benefit from deterring tortious behavior. This balancing can result in optimal judgements either bigger or smaller than the harm inflicted. Polinsky and Che (1991), for example, suggest that the penalty should be greater, but so should be the burden of proof, to reduce the number of cases brought and thus the transactions cost. The question remains of whether the judgement should increase smoothly with the harm.

The present model applies to civil liability insofar as deterrence, not compensation, is its aim. Tort law uses cash transfers as penalties to deter harmful behavior, but high penalties are costly because when wrongly inflicted they deter innocent behavior. The model shows that the optimal schedule for liability need not be continuous. Figure 1 can be relabelled as the relationship between liability and amount of harmful behavior, where a small amount of attention will eliminate a large number of accidents, but further large reductions require substantial changes in behavior. A small penalty will prevent many accidents, but the penalty must become much greater if the number of accidents is again to be substantially reduced. High liability for lower levels of harm would needlessly deter innocent behavior, but at some point the harm becomes great enough for deterrence of harmful behavior to take precedence, and a jump occurs in liability.

Whether the offense be criminal or civil, a discontinuous jump in the optimal punishment is not, of course, a necessary conclusion, only a possible one, which depends on the deterrent relationship between crime and penalty. The example showed that such a jump would occur if there is a large group of potential offenders who are deterred by a small penalty and others who will not be deterred except by substantially higher penalties. Whether this is true depends on the particular time and place, but judges should pause before striking down statutes simply because similar offenses have very different penalties; a jump in penalties at some harm level may be appropriate, even though the particular threshold level may seem arbitrary.

REFERENCES

Andreoni, James, 1991, "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" *RAND Journal of Economics*, 22: 385-395.

Becker, Gary (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, March/April 1968, 169-217.

Ehrlich, Isaac (1982) "The Optimum Enforcement of Laws and the Concept of Justice: A Positive Analysis," *International Review of Law and Economics*, 2: 3-27.

Friedman, David (1981) "Reflections on Optimal Punishment, or: Should the Rich Pay Higher Fines?" *Research in Law and Economics*, ed. Richard Zerbe, 3: 185-205.

Kaplow, Louis (1990a) "A Note on the Use of Nonmonetary Sanctions," *Journal of Public Economics*, 42: 245-247.

Kaplow, Louis (1990b) "Optimal Deterrence, Uninformed Individuals, and Acquiring Information about Whether Acts are Subject to Sanctions," *Journal of Law, Economics, and Organization*, Spring 1990, 6: 93-128.

Kaplow, Louis (1994) "The Value of Accuracy in Adjudication: An Economic Analysis," *Journal of Legal Studies*, 33: 307-402.

Mookherjee, Dilip and Ivan Png (1992) "Monitoring vis-a-vis Investigation in Enforcement of Law," *American Economic Review*, (June 1992), 82: 556-565.

Png, Ivan (1986) "Optimal Subsidies and Damages in the Presence of Judicial Error," *International Review of Law and Economics*, June 1986, 6:101-105.

Polinsky, A. Mitchell and Y. Che (1991), "Decoupling Liability: Optimal Incentives for Care and Litigation," *Rand Journal of Economics*, Winter

1991, 22: 562-570.

Polinsky, A. Mitchell and Steven Shavell (1984), “The Optimal Use of Fines and Imprisonment,” *Journal of Public Economics*, 24: 89-99.

Posner, Richard (1973) “An Economic Approach to Legal Procedure and Judicial Administration,” *Journal of Legal Studies*, 2: 410-415.

Rasmusen, Eric (1994) “Damage Awards When Court Error is Predictable,” working paper, Indiana University School of Business, June 1994.

Rubinfeld, Daniel and David Sappington (1987) “Efficient Awards and Standards of Proof in Judicial Proceedings,” *Rand Journal of Economics*, Summer 1987, 18: 308-315.

Shavell, Steven (1985) “Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent,” *Columbia Law Review*, October 1985, Vol. 85, 1232-1263.

Shavell, Steven (1987) “The Optimal Use of Nonmonetary Sanctions as a Deterrent,” *American Economic Review*, September 1987, 77: 584-592.

Shavell, Steven (1991) “Specific Versus General Enforcement of the Law,” *Journal of Political Economy*, October 1991, 99: 1088-1108.

Stigler, George (1970) “The Optimum Enforcement of Laws,” *Journal of Political Economy*, May/June 1970, 78: 526-36.