Centre de recherche sur l'emploi et les fluctuations économiques (CREFÉ)
Center for Research on Economic Fluctuations and Employment (CREFE)

Université du Québec à Montréal

Cahier de recherche/Working Paper No. 101

Limiting Court Behavior: A Case for High Minimum Sentences and Low Maximum Ones *

Dominique Demougin
CREFE-Otto-von-Guericke University

Stéphane Pallage
CREFE-UQAM

January 2000

Otto-von-Guericke University, Universitätsplatz 2, D-39106 Magdeburg, Germany. Fax: (49) 391-671-1218. Email: demougin@ww.uni-magdeburg.de.
Department of Economics, University of Quebec at Montreal, CP 8888 succ. Centre-Ville, Montreal, QC, H3C 3P8, Canada. Tel: 514-987-3000 (8370#). Fax: 514-987-8494. Email: pallage.stephane@uqam.ca.

Abstract:
We model a simple justice system in which a court is mandated by society to assess the guilt and the punishment of an accused. The court takes prison facilities as given and neglects its impact on the cost to society of implementing the sentence. Clearly, the court, in this world, will condemn more often than society and assign higher penalties. Under these circumstances, society at large would necessarily benefit from having maximum sentences. We show, however, as a series of perverse results, that (1) maximum penalties need to be lower than the highest socially desirable penalty; (2) society would benefit from imposing high minimum sentences even though it is precisely the harshness of courts, which it wants to curb.

Keywords:
Maximum and minimum penalties, sentencing guidelines, social optimum

JEL classification: K41,K14

1 Introduction

The purpose of this paper is to find a rationale for the existence of maximum as well as minimum penalties associated with specific crimes. Though in most democracies, the duty of pronouncing the guilt or the innocence of a citizen is left to a court, it is also true that society limits sanctions, thereby suggesting that it does not entirely trust the court's choices.

Clearly, for society to restrict the behavior of courts, it must be that courts tend to sentence differently than would be socially desirable. In particular, if courts tend to be more severe than society, imposing maximum and minimum sentences might help align their decisions with the social optimum. Courts would indeed be more severe than society if, when choosing the sentence, they neglect the social cost of this sentence. We argue that this is very likely. In particular, we argue that the social cost of penalties is an externality.

This is equivalent to the following two assumptions: First, a court, when assigning a penalty believes its impact on the social cost of sentences is nil or very small. We believe this is a reasonable assumption. Indeed, when choosing a penalty, the court takes prison facilities as given. Because of the essentially fixed nature of the cost of prisons, the cost of sending one more person to jail is negligible with respect to overall cost of sentences. Hence, the court is likely to ignore that cost. Clearly, however, society does care about the fixed cost. Second, a court, when making its judgment, believes it has no impact on the choice of other courts. The court focuses on one case, whereas society at large deals with many cases. Society might be willing to make some cost arbitrage between cases, which courts are not capable of. For these reasons, the objectives of society and a given court will not typically lead to the same choice of penalty.

In this paper we show that it is socially optimal to restrict courts behavior by imposing not only a maximum sentence, but also a minimum sentence. Moreover, we show that the maximum sentence ought to be lower than the highest socially optimal penalty. Those results are somewhat perverse. A society unhappy with its courts, imposing too often penalties that are socially too high, might manage to refrain courts from doing so by setting a high minimum penalty and a low maximum one. The role of the minimum penalty is very important: with a minimum penalty, courts might decide to acquit those accused that society would not have convicted.

These limits to sentences, while helping society contain the cost of sentences, tend to make it easier for criminals. In an interesting article on the pattern of crime in 17th and 18th centuries England, Beattie (1974) provides an extreme example of the effects of minimum sentences. The author reports that in 1689 an estimate of 50 crimes, many of which petty crimes, were by law assigned the death penalty. This number was increased to 200 by the end of the 18th century. Beattie (1974) documents evidence that the increased harshness of the criminal code led courts to either dismiss cases for which there was sufficient evidence to condemn but the penalty was perceived to be too strong, or to underplay the description of some crimes in order to save petty criminals from capital punishment. In a more recent example, DiIulio (1996) states:

> Where “three strikes” laws have taken effect, prosecutors have begun to exercise their discretion in bringing charges in ways that spare many thrice-convicted violent felons one-way tickets to the big house. [DiIulio, page 9]

These examples of penalties clearly exceed the severity of the minimum penalty we advocate as optimal for society. But they illustrate nicely the mechanisms by which minimum penalties may end up reducing convictions. It is precisely because the minimum penalty is perceived to be too high for the case at hand that some accused are not penalized.

There exists a very rich literature on the optimal size of sentences, initiated by the seminal research by Becker (1968). The sense of optimality of a sentence in this literature is directed towards the deterrence of criminal behavior. Becker showed that the optimal deterrent was often a uniformly high penalty. When taking into account the principle of reasonable doubt that forms the basis of the US justice system, Andreoni (1991) established that the optimal deterrent was rather a penalty growing with the level of the offence. Using a similar reasoning, Rasmusen (1995) illustrated cases under which the penalty is not a continuous function of the level of harm.

Our emphasis is quite different. Though in our model preferences could be interpreted as resulting from the deterrence incentive, our main objective is to emphasize the externality caused by the penalty and the consequences of court decisions on social welfare. We show how minimum and maximum sentences, for a given crime, can be used efficiently to reduce the gap between the optimal choices of society and its courts.

Another branch of literature has been widely concerned about the optimal magnitude of fines and their use relative to imprisonment. See for example Polinsky & Shavell (1979, 1984), Friedman (1981) and Waldfogel (1995). These papers build on the fact that imprisonment being more costly than fines, should be used only when the criminal is unable to pay. Implicitly, they focus on “white collar crimes” such as fraud, or property crimes such as theft. In this paper, we are interested in crimes that require other types of penalties than fines.

The remainder of the paper is as follows. In the next section, we build a model of a simple justice system. We characterize the optimal sentence from the perspectives of a court and society respectively. In Section 3, we further highlight the differences between optimal choices by a court and society, and show the importance of reasonable doubt. In Section 4, we offer the main result on minimum and maximum sentences. We conclude in Section 5.

2 The model

In this section, we consider the following simple representation of a justice system. For simplicity, we focus on one court among many in similar situations. A crime has been committed. We assume that the seriousness of the offence can be measured by a scalar. Popular consensus has evaluated to x the extent of the offence. The crime is of a type that precludes monetary sanctions.

The police has made an investigation and identified a suspect who faces the court. In an adversary trial, the prosecution makes its case against the accused, the defence attempts to find weaknesses in the prosecution's argument, and as a result, the court[1] determines the probability q that the accused is indeed guilty of the crime. This probability reflects the amount of resources devoted by society to the case. We take these resources as exogenously given.

The court then chooses to condemn or not, and contingent on condemnation, the appropriate level of the sentence p. Initially, we assume that there is a continuous choice of penalties $p \in [0, \infty)$. Dismissal of a case implies $p = 0$. For parsimony, sentences and offences can be measured by the same metric and are expressed in similar units. Once an accused is found guilty and the penalty has been fixed at p, society bears a cost $C(p)$ for the implementation of the sentence. We assume that the social cost of the sentence is an externality for the courts. This has two implications. First, a given court believes its impact on overall sentencing cost is nil, which as we argued before is reasonable considering that it takes prison facilities as given. Of course, if from the perspective of the court, sending one more person to jail has no cost consequence, it is likely, however, to matter dearly to society. Second, the court believes its decision does not affect the decision of other courts, and arbitrage across cases is impossible.

Let x_a represent the actual offence of the accused. From the perspective of the court, two events

[1]In all that follows, we model the court as an entity, while actual courts are often made of a jury and a judge. We abstract from the question of optimally dividing tasks within a tribunal since we are interested in rationalizing constraints imposed upon both judges and juries.

can occur. Either the accused is guilty ($x_a = x$) or he is not guilty ($x_a = 0$). When choosing the sentence, the court only knows the probability associated with those two events. Hence it may turn out doing one of four more or less desirable things: condemning a culprit, condemning an innocent, not condemning a culprit or not condemning an innocent.

A court has preferences over the four outcomes. We assume that these preferences satisfy necessary requirements to be represented by a von Neumann–Morgenstern disutility function $u = u(\alpha)$, strictly increasing and convex in α, with $u(0) = u'(0) = 0$ (where $u'(.)$ is the first derivative of the function u). We make two assumptions about these preferences. First, we assume that courts dislike gaps between the offence of the accused x_a and the penalty ($|x_a - p|$). For example, they dislike sentencing to 25 years in jail someone who stole a piece of bread. Second, we assume that they get extra–disutility from condemning an innocent.[2] Hence we have

$$u = u\left(|x_a - p| + \Phi(x_a, p)\right) \quad \text{with } \Phi(x_a, p) = \begin{cases} \phi & \text{if } x_a = 0 \text{ and } p > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\phi > 0$ is an additional moral cost of an unjust penalty, stemming from empathy.[3]

It will be useful, though not necessary for the case of the court, to interpret the disutility associated with each outcome as a cost which could be expressed in monetary units. This will come particularly handy when we build the social planner's function.

If the court knew with certainty the status of the accused, sentencing a culprit ($x_a = x$) to a penalty p would leave the court with disutility $u(|x_a - p|) = u(|x - p|)$. After sentencing an innocent ($x_a = 0$) to p the court's disutility is given by $u(|x_a - p| + \phi) = u(p + \phi)$. Not condemning an innocent ($p = 0,\ x_a = 0$) is fine from the perspective of the court ($u(|x_a - p|) = 0$). Not condemning a culprit ($p = 0,\ x_a = x$) yields $u(|x_a - p|) = u(x)$ and the court is left with the unpaid crime.[4] Table 1 summarizes the disutility of the court in the four alternatives.

[2]These assumptions lead us to an indirect utility function consistent with Andreoni (1991). In particular the second assumption allows for "reasonable doubt".

[3]Alternatively, assuming courts have no empathy, one can interpret ϕ as the future cost of having a genuine criminal outside. Indeed, by closing the case, the court has stopped all investigations on that case with the undesirable effect that the real culprit remains unbothered and may recidivate.

[4]Note that with this formulation there is no extra moral cost of not condemning a culprit. In other words, there is no special satisfaction from social revenge.

Table 1: The court's cost

	q	$1-q$
	Guilty	Not Guilty
Condemn	$u(\|x-p\|)$	$u(p+\phi)$
Not Condemn	$u(x)$	0

2.1 The court

The court must decide on both guilt and penalty. This implies comparing the disutilities of condemning u^C and not condemning u^{NC} the defendant. The court thus solves the following nested problem:

$$[P] \quad \min \{ \underbrace{\min_p \; q\; u(|x-p|) + (1-q)\; u(p+\phi)}_{u^C} \; , \quad \overbrace{q\; u(x)}^{u^{NC}} \; \}$$

Note that $q \; u(x) + (1-q) \; u(\phi) > q \; u(x)$, thus the court would never condemn and set $p = 0$. Hence convicting the accused implies $p > 0$, while $p = 0$ is equivalent to finding the accused innocent.

Call p^T the optimal penalty from the perspective of an unconstrained tribunal. p^T has the following unsurprising features: It fits the crime only in the case of perfect information on the guilt of the accused. As the probability of guilt goes to 1, p^T goes to x. More serious crimes require more severe penalties, and the risk of convicting an innocent tempers the court. Proposition 1 provides a more thorough characterization:

Proposition 1 (court's behavior)

(i) Under perfect information on the guilt of the accused ($q = 0$ or $q = 1$), the court always sets $p^T = x_a$.

(ii) Under imperfect information ($q \in (0, 1)$), p^T is smaller than the crime x.

(iii) p^T is increasing in x and in the probability of guilt q, while it is decreasing in the moral cost ϕ of sentencing an innocent.

The proof of this proposition follows directly from the first order condition of the court's minimization problem u^C. (iii) is a direct application of the implicit function theorem. For the remainder of the paper we will be entitled to write $|x - p^T| = x - p^T$.

2.2 Society

As the court, society is concerned about the case of the accused. But it is also concerned with the management of the sentence chosen by the court. If the cost of carrying through the sentence is an externality for the court, it ought to be taken into account in the search for the social optimum. Consequently, the costs to society in the four alternatives differ from the costs to a given court. In Table 2, for example, the cost to society of condemning an innocent is higher than the cost to the court: the court cares about $|x_a - p| + \phi$ which, since the accused is innocent leaves it with the weight of an unjust penalty $p+\phi$. Society bears the same disutility, but is left with yet another cost, the cost induced by the sentence $C(p)$. This cost has to be borne in all cases where a conviction of level p takes place. We assume that $C(p)$ is positive, monotone, increasing and convex.[5]

Table 2: Society's cost

	q Guilty	$1-q$ Not Guilty
Condemn	$u(\lvert x-p\rvert) + C(p)$	$u(p+\phi) + C(p)$
Not Condemn	$u(x)$	0

Society's objective is to minimize the social cost associated with the judgment. This means comparing the cost of condemning w^C and the cost of not condemning the accused w^{NC}. A social planner would solve:

$$[P'] \quad \min \{ \underbrace{\min_p \; q \; u(|x-p|) + (1-q) \; u(p+\phi) + C(p)}_{w^C} \; , \quad \overbrace{q \; u(x)}^{w^{NC}} \; \}$$

Denote by p^S the socially optimal penalty.

[5]From Table 2, we see why it is important to view the disutility of the different outcomes as a cost. It allows us to relate the social cost of imprisonment to the disutility of sentencing.

Proposition 2 (Social optimum)

(i) $\forall\ q \in (0, 1]$, *the socially optimal penalty* p^S *is always lower than the crime:* $p^S \leq x$.

(ii) p^S *is increasing in* x *and in* q, *and decreasing in* ϕ.

Again the proof of this proposition is straightforward. Because of the social cost of implementing the sentence, society would prefer a penalty that *never* fits the crime. Society only agrees with the court in the case of unquestionable innocence. In other words, society's goal, is not perfectly met by a court. We can rewrite $|x - p^S|$ as $x - p^S$.

3 Society vs court

In a world of imperfect information on the guilt of the accused, the impact of the externality is most important. Proposition 3 characterizes the differences in the choices of the court and society.

Proposition 3 (Society vs court)

1. *If* $p^T(q) = 0$, *then* $p^S(q) = 0$.
2. *If* $p^S(q) > 0$, *then*

   (i) $p^T(q) > 0$,

   (ii) $p^S(q) \leq p^T(q)$.

Proof of this proposition can be found in the appendix. Its message is two-fold: First, the court condemns "at least as often" as society. Indeed, if the court finds it optimal not to condemn the defendant, so does society. If society finds it optimal to condemn, so does the court. The converse, however, is not true. In other words, if one is quicker to convict, such is the court. Second, the penalty imposed by the court upon conviction is higher than the penalty socially desirable. All in all, society's optimal penalty is always smaller than the court's, and the court convicts at least as often as society would.

The differing tendencies of courts and society extend to the notion of *reasonable doubt*. Let us denote by q^{*T} and q^{*S} the probabilities of guilt that make the tribunal and society respectively indifferent between condemning and not condemning.

Proposition 4 (Reasonable doubt)

(i) If $\phi > 0$ then both $p^T(q)$ and $p^S(q)$ are discontinuous at q^{*T} and q^{*S} respectively.

(ii) $q^{*T} < q^{*S}$

We prove these claims in the Appendix. Both q^{*T} and q^{*S} have a natural interpretation in the notion of reasonable doubt. If there is a moral cost ϕ of condemning an innocent, then for the court to condemn requires a certain minimum probability $q^{*T} > 0$ that the accused is guilty. Society is more stringent than the court on the reasonable doubt requirement. To put it simply, the court and society in our model do not agree on what constitutes a reasonable doubt.

Moreover, the discontinuity of p^T and p^S at respectively q^{*T} and q^{*S} implies that there is a *natural minimum sentence* for the court as well as for society. We illustrate this geometrically for the court in Figure 1. Let us define $\hat{p}^T(q)$ and $E(\hat{u}|q)$ as respectively:

$$\begin{aligned}&\hat{p}^T(q) \equiv \arg\min_p qu(x-p) + (1-q)u(p+\phi)\\&E(\hat{u}|q) \equiv qu(x-\hat{p}^T(q)) + (1-q)u(\hat{p}^T(q)+\phi)\ .\end{aligned}$$

$E(\hat{u}|q)$ represents the expected disutility of the court when condemning the defendant at the optimally chosen punishment $\hat{p}^T$. In Figure 1 we have represented $E(\hat{u}|q)$, which, from the envelope theorem, is decreasing in q.[6] The line $qu(x)$ represents the disutility of the court upon dismissal of the case. Since the court attempts to minimize the disutility of its decision, we can immediately see from Figure 1 that it will choose $p^T(q)$ such that:

$$p^T(q) = \begin{cases} \hat{p}^T(q) & q \geq q^{*T} \\ 0 & \text{otherwise}\ . \end{cases}$$

Note that $\hat{p}^T(q)$ is continuous. This has an important implication, namely that there are $q < q^{*T}$ for which $\hat{p}^T(q) > 0$ yet $p^T(q) = 0$. This means that if it had found the defendant guilty, the court would have been ready to impose a positive sentence. The disutility of actually convicting the accused, however, is too high, which results in a non-guilty verdict. Hence, the court considers that the defendant deserves a sentence but it is not willing to assign it to him. This could be interpreted as sentencing with remission. The court does not want to condemn to sentences that are too small, which explains the discontinuity of $p^T(q)$ at q^{*T}. Hence for a given crime, there

[6]This is natural since $E(\hat{u}|q) \to 0$ as $q \to 1$.

exists a “natural” minimum sentence self-imposed by the court. The key to understanding why this is so lies again in ϕ, i.e. the moral cost of possibly condemning an innocent. When the probability of guilt is small, the risk of convicting an innocent is relatively large. Consequently the disutility of condemning is bigger than the disutility of not condemning. If ϕ were nil, the court would be willing to impose arbitrarily small sentences for arbitrarily small evidence of guilt.[7]

Figure 2 summarizes the differences between the behaviors of society and the court. Note that except for the linearity, Figure 2 is fully general. Hence we will make use of this picture hereafter to illustrate the more general case.

4 Minimum and Maximum Sentences

If the court is unconstrained in its choice of p, for every level of q, the distance between $p^T(q)$ and $p^S(q)$ measures the gap between court’s actions and society’s desires. Naturally, since q is not observable *ex ante*, the point-wise loss is not a good measure. Instead, we must revert to expected loss which implies weighting the point-wise loss by its probability.

Let $\Psi(q)$ denote the relative frequency by which an accused exhibits the probability q of being guilty for given police resources and crime type. We can view Ψ as a probability distribution over q. If q was distributed uniformly, the surface between $p^T(q)$ and $p^S(q)$ would represent an adequate measure of the welfare loss. For parsimony, but without loss of generality, we will work with a uniform distribution, so that we can use geometric arguments. It is easy to generalize the results to any reasonable distribution.[8] Specifically, the welfare loss can be represented within a figure similar to Figure 2 with some distortion.

We now examine the following question: Could society constrain court behavior in such a way that it would reduce the welfare loss of delegating? What instruments are available to society? Since society delegates the choice of the verdict to the court, society does not observe q and cannot force

[7]Clearly, the natural minimum sentence is an artifact of our model. It arises from the assumption that ϕ is fixed. This assumption is not unrealistic considering that there is a fixed component to sentences: the shame that hits an accused for spending even one night in jail. Because of this fixed component, a court in reality could not assign arbitrarily small penalties even if it wanted to.

[8]Of course, the uniform distribution, though useful, is by no means reasonable, since it suggests that the police selected the accused randomly. By reasonable, we mean an upward-sloping distribution. Hence, the higher q the higher the number of accused of that type.

the court to reveal its actual value.[9] Hence society can only restrict court behavior by imposing limits to sentences.

In Figure 3 we show that imposing a *maximum sentence* p^M would reduce the welfare loss. For example imposing for maximum sentence the worst sentence society would want to inflict on the accused, i.e. $p^M = p^S(1)$, would reduce the welfare gap by the area of triangle abc. This does not mean, however, that $p^S(1)$ is the best maximum sentence that we can design. In fact, the optimal maximum sentence is strictly lower than this one, as we will show below.

Could *minimum sentences* help? A minimum sentence such as p^m in Figure 3 has two different effects. First, it pushes further the conviction threshold of the court. The latter will indeed require a higher probability of culpability to convict the accused to the minimum penalty. Hence, it reduces the area between p^T and p^S by the area of trapezium $efdq^{*T}$. Second, the minimum sentence imposes a social cost stemming from the fact that some accused will now be sentenced to a higher pain than desired by the court, which was already too high from the point of view of society. This second effect increases the gap between social goal and court's actions by the area of triangle fgh. The tradeoff between these two effects will determine whether a minimum sentence is desirable or not. Call p^*_M and p^*_m the optimal maximum and minimum sentences. We now introduce our most important result: it is indeed in society's interest to set upper and lower limits to sanctions. The upper-bound ought to be *small*, and the lower-bound *high* in the following sense:

Proposition 5 (Limits to sentences)

(i) *There exist a minimum sentence p^*_m and a maximum sentence p^*_M which are welfare improving.*

(ii) $p^*_m > p^T(q^{*T})$.

(iii) $p^*_M < p^S(1)$,

The intuition for the proof goes as follows. That there exists a welfare improving maximum sentence is clear from Figure 3: $p^S(1)$ is one such maximum sentence. That the same holds for a minimum sentence can be established using the following argument: suppose we impose a minimum sentence p^m, a small value ϵ above $p^T(q^{*T}) = e$, in Figure 3, and we let $\epsilon \to 0$. Trapezium

[9] One could think of solving a mechanism design problem whereby the court would be asked to make a report on q. However, to induce truthful revelation, such a mechanism requires side payments, which, for obvious reasons, are precluded in any justice system.

$efdq^{*T}$ has strictly positive side eq^{*T} while its height $q^{*T}d$ converges to zero; whereas triangle fgh has both height fh and base hg converging to 0. Hence, the welfare loss converges to zero faster than the welfare gain. This establishes (i) and (ii). To see why (iii) holds, note that the welfare gain of lowering the maximum sentence by a small value ϵ below $p^S(1)$ exceeds the welfare loss of doing so. The welfare gain can be measured in Figure 3 by the area of the trapezium of height $c - \epsilon$ and length cb, whereas the welfare loss is measured by a triangle of smaller base and same height.

These results are striking for the following reasons. First, a society wishing to refrain its courts from being too harsh on criminals, would find optimal to set high minimum penalties. It would do so, not to increase severity, but rather to decrease the rate of convictions. The response of courts in our model, will indeed be a reduction in this rate of convictions.

Historical examples of such responses of courts to high minimum penalties are described by Beattie (1974). In 17th and 18th centuries England, the number of crimes punishable by death was on the rise. By the end of this period, about 200 offences, many of them minor property crimes, were by law assigned the death penalty. As the criminal code became harsher, Beattie (1974) shows evidence of "an increasing tendency over the period for prosecutors and the courts alike deliberately to understate the nature of the crime in order to save the accused from the gallows,"[10] or to simply dismiss cases for which the death penalty was perceived by the court to be excessive. This last effect was reinforced by another one, directly coming from the victims whose dislike for the death penalty made them more reluctant to prosecute offenders. Hay (1975) reports contemporary opinion that "the gibbets and corpses paradoxically weakened the enforcement of the law: rather than terrifying criminals, the death penalty terrified prosecutors and juries, who feared committing judicial murder on the capital statutes."[11] In essence, the extremely severe minimum penalty in this era had all potential to make it easier for criminals. According to DiIulio (1996), a similar reaction of courts to high minimum penalties took place recently in the United States after "three strikes" laws were passed in some states. These laws were designed to discourage repeat offences: upon third conviction, a felon was to be sentenced to life behind bars. DiIulio (1996) suggests that courts started to find ways of avoiding such drastic conclusion. Clearly, both these examples are extreme cases of minimum penalties and far exceed the minimum penalties we advocate in this paper.[12] Yet, they illustrate the mechanism by which minimum penalties may affect the behavior

[10] Beattie (1974), page 83.

[11] Hay (1975), page 23.

[12] Though Hay (1975) argues that the use of capital punishment in England and the reaction of courts might have

of courts.

A second striking aspect of our result is that to achieve its goals, at the risk of seeming inconsistent, society would also find optimal to set low maximum penalties. This time, it is not the rate of convictions, which it would target, but rather the size of the penalty courts would assign to defendants found guilty.

5 Conclusion

In this paper, we show that minimum and maximum penalties may be used optimally to reduce the rate of convictions and the magnitude of sentences. Our theory is based on the assumption that the cost of implementing a sentence is an externality to courts. As a result, the latter are inclined to be more severe than is socially optimal. A mix of high minimum penalties and low maximum ones is shown to align courts with societal goals.

References

Andreoni, J. (1991). Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?. *RAND Journal of Economics*, *22*, 385–395.

Beattie, J. M. (1974). The pattern of crime in England, 1660-1800. *Past & Present*, *62*, 47–95.

Becker, G. S. (1968). Crime and punishment: an economic approach. *Journal of Political Economy*, *76*, 169–217.

DiIulio, J. J. (1996). Help wanted: economists, crime and public policy. *Journal of Economic Perspectives*, *10*(1), 3–24.

Friedman, D. (1981). Reflections on optimal punishment; or should the rich pay higher fines?. *Research in Law and Economics*, *3*, 185–206.

Hay, D. (1975). Property, authority and the criminal law. In *Albion's Fatal Tree: Crime and Society in Eighteenth-Century England*. D. Hay, P. Linebaugh, J.G. Rule, E.P. Thompson, C. Winslow Editors, Pantheon Books, New York.

Polinsky, M. & Shavell, S. (1979). The optimal tradeoff between the probability and magnitude of fines. *American Economic Review*, *69*, 880–891.

Polinsky, M. & Shavell, S. (1984). The optimal use of fines and imprisonment. *Journal of Public Economics*, *24*, 89–99.

Rasmusen, E. (1995). How optimal penalties change with the amount of harm. *International Review of Law and Economics*, *15*, 101–108.

Waldfogel, J. (1995). Are fines and prison terms used efficiently? evidence on federal fraud offenders. *Journal of Law and Economics*, *38*, 107–139.

adequately served the goals of Parliament.

Appendix

A Proof of Proposition 3

To prove claim 1, note that $p^T = 0$ implies

$$\min_p q\ u(x-p) + (1-q)\ u(p+\phi) > q\ u(x).$$

Now, we know that

$$u^C = \min_p q\ u(x-p) + (1-q)\ u(p+\phi) < \min q\ u(x-p) + (1-q)\ u(p+\phi) + C(p) = w^C.$$

Hence $w^C > q\ u(x) = w^{NC}$, which implies $p^S = 0$.
Claim $2i$ is the logical contraposition of claim 1. Finally, for claim $2ii$, let $p^S = p^*$, for given q. We have

$$-q\ u'(x-p^*) + (1-q)\ u'(p^*+\phi) + C'(p^*) = 0$$

Hence, using the fact that $C(.)$ is an increasing function,

$$-q\ u'(x-p^*) + (1-q)\ u'(p^*+\phi) < 0$$

So p^* does not solve the court's problem. Given the strict convexity of u, there exists $\Delta \in (0, x)$ such that

$$-q\ u'(x-(p^*+\Delta)) + (1-q)\ u'(p^*+\Delta+\phi) = 0$$

Hence, $p^T = p^* + \Delta = p^S + \Delta \geq p^S$, which is satisfied for any q. ■

B Proof of Proposition 4

(i) Let $\phi > 0$. For the court to be indifferent between convicting or not requires that:

$$q^{*T} u(x-p^T) + (1-q^{*T}) u(p^T+\phi) = q^{*T} u(x)$$

This in turn implies $p^T > 0$ at q^{*T} and $p^T = 0$ at $q^{*T} - \epsilon\ ,\ \ \forall\ \epsilon > 0$. A similar reasoning applies to p^S.
We prove claim (ii) by showing that at q^{*T} society would not condemn the defendant. This follows from a series of inequalities:

$$\begin{aligned}
q^{*T} u(x) &= q^{*T} u(x - p^T(q^{*T})) + (1-q^{*T}) u(p^T(q^{*T}) + \phi) \\
&< q^{*T} u(x - p^S(q^{*T})) + (1-q^{*T}) u(p^S(q^{*T}) + \phi) \\
&\leq q^{*T} u(x - p^S(q^{*T})) + (1-q^{*T}) u(p^S(q^{*T}) + \phi) + C(p^S(q^{*T}))
\end{aligned}$$

The first inequality stems from the fact that $p^T(q)$ is the penalty that minimizes $qu(x-p)+(1-q)u(p+\phi)$. The second inequality follows from the fact that $C(p) \geq 0$. Hence the claim. ■

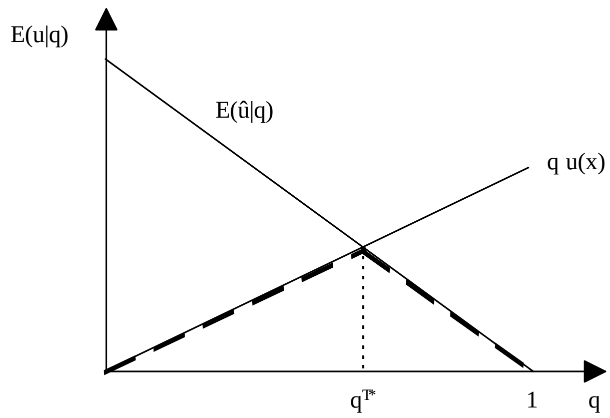

Figure 1: On reasonable doubt

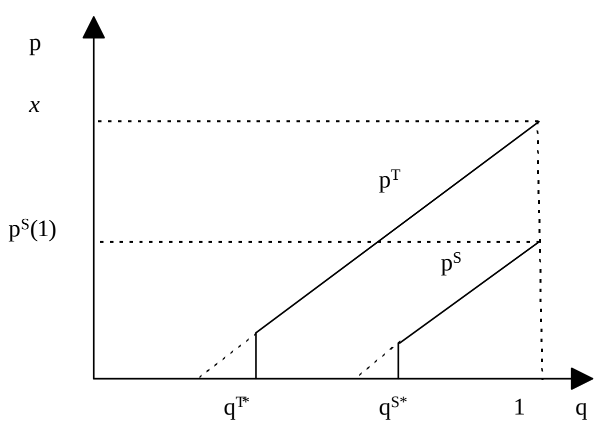

Figure 2: Optimal sentences, society vs court

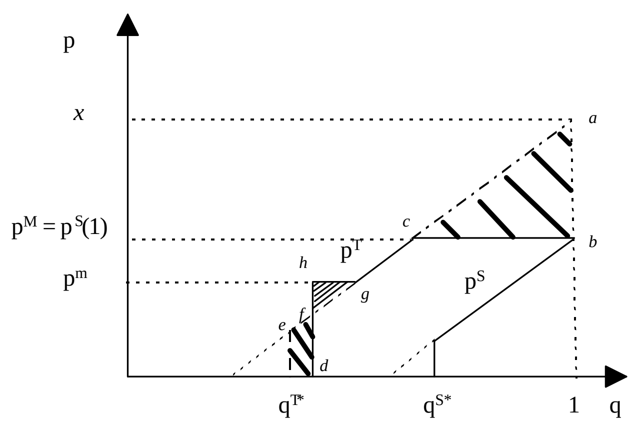

Figure 3: On the use of limits to sentences