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

Stephane Pallage* and Dominique Demougin† ‡

September, 2001

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 also benefit from imposing high minimum sentences even though it is precisely the harshness of courts, which it is trying to curb. [JEL. K0]

Keywords: Optimal penalty.

1 Introduction

In most democracies, the duty of pronouncing the guilt or innocence of a citizen is delegated to a court. However, it is also true that society often imposes sentencing guidelines, thereby limiting sanctions that can be imposed for specific crimes, and, suggesting that it does not entirely trust its own

*Université du Québec à Montréal, Department of Economics, CP 8888 succ. Centre-Ville, Montreal, H3C 3P8, Canada. Fax: 514-987-8494. Email: pallage.stephane@uqam.ca.

†Humboldt Universität zu Berlin, Spandauerstr. 1, 10178 Berlin, Germany. Fax: (49) 30-2093-1342. Email: demougin@wiwi.hu-berlin.de.

‡We are very much indebted to John M. Beattie and Ron Wright for helpful discussions and to two anonymous referees for detailed comments of an early version of the paper. Usual disclaimer apply and the authors alone are responsible for the views expressed and any remaining errors.

court system. For example, the United States developed in the 1980s sentencing guidelines as a response to two perceived 'judicial pathologies' (Wright, [1998]), both related to the way judges were observed to sentence. First, there was a large of a disparity in sanctions for seemingly similar offenses, and second, judges' tendency to view prison cells as a public good making overuse of this scarce resource (Wright and Ellis [1993]; Wright [1998]).[1] The first pathology is well known and extensively documented in the existing literature.[2] In this paper, we put our emphasis on the second type of judicial pathologies. Not withstanding alternative explanations for sentencing guidelines, we wish to show that they can successfully address this problem.

Clearly, for the state to restrict courts' behavior, it must be that under some circumstances these sentences differ from what is perceived to be socially desirable. We argue that courts might tend to impose verdicts that are more severe than society because they neglect some of the costs associated with their sentence. In particular, we contend that from the point of view of a court some of the social costs associated with penalties is perceived as an externality.

For example, when deciding on a prison term a court might believe its impact on the social cost of sentences is nil or very small. Indeed, when choosing to either penalize or not, and possibly when determining the length of imprisonment, a court is likely to take incarceration facilities as given, thereby ignoring the cost of sending one more person to prison.[3] Some might argue at this point that it is not reasonable for a court to overlook its impact on the overall detention costs. For example, by sentencing an accused to jail, any court should realize that food expenses are being raised. However, even if courts were to correctly estimate the true costs of sentencing, it would still remain reasonable for them not to care, simply because it is somebody else's budget. For example, in the United States, prosecutors and judges are all responsive, whether directly or indirectly, to the interests of the citizens of local jurisdictions and responsible to a local constituency. The biggest

[1]This led to the development of the parole system (Abadinsky, [1978]). Society endowed parole officers with full discretion to reduce prison terms and avoid overcrowding. The parole system, however, amounted to solve a problem of discretion by introducing another. The Sentencing Reform Act of 1984, establishing sentencing guidelines also aimed at restricting parole.

[2]See Wright [1998] and the references therein.

[3]One could argue that there may be an other external benefits associated with prison sentences. If serving time actually rehabilitates criminals, then courts may not sentence enough from society's point of view. According to Blumstein [1983], however, this effect, while present for many inmates, is cancelled by the opposite effect for many others, that prison in fact socializes criminals. In the remaining of the paper, we completely ignore these issues.

provider of prison facilities, however, is the state government. This service comes at virtually no cost to local governments. Giertz and Nardulli [1985] argue that this misalignment of responsibilities between local and state governments is a substantial factor in explaining current prison overcrowding.[4] In either case, the objectives of society and that of a given court will not typically lead to the same choice of penalty.

In this paper, we show that under these circumstances, it is socially optimal to restrict courts' behavior by imposing a maximum as well as a minimum sentence. Moreover, we prove that the maximum sentence ought to be lower than the highest socially optimal penalty. These conclusions are somewhat perverse second-best results: a society unhappy with its courts imposing too often too high penalties, prevents them from doing so by setting a high minimum penalty and a maximum punishment below what it itself would be willing to impose! The role of the minimum penalty is very important: faced with imposing either no punishment or the minimum sentence, the courts will often find it optimal to acquit. We illustrate this mechanism with some historical examples from England in the 17th and 18th centuries, when many petty crimes were by law assigned the death penalty.

Given that our main argument relies on an externality, it could be argued that taxes or quotas might serve as alternatives to sentencing guidelines. Of course, the administration of such tax would raise other difficult problems.[5] For example, in countries in which such principle is important, a tax would likely violate the principle of independence of the judges if the income of a judge or the budget of his office were a function of his sentencing practices. More importantly, both instruments imply non trivial administrative costs unlike limiting the court's behavior. Moreover, from a positive view point, we don't observe taxes or quotas for prison cells. Thus, in the remaining, we abstract from these instruments and constrain our analysis to show that the use of limits to sentences may rationally follow from the externality.

[4]Moreover, courts do not coordinate on the optimal allocation of jail time across all defendants, which also leads to prison overcrowding.

[5]In the United States, as we suggested above, judges' budgets depend on local governments whose interest in being tough on criminals is quite different from that of state governments running most of the prison system. As a result, these local governments would have no interest in taxing their judges. In a quota system, courts would each be allocated a fixed number of cells. Once a court has used up its allocation, it would have to decide which cell to vacate, before sentencing a new convict. Quotas suffer from various problems. Not all courts are faced with the same menu of crimes. The past menu is not necessarily a good predictor of the future one. On what basis should such quotas be allocated? On the basis of past sentencing practices? This would likely favor the most severe courts. There is a large element of arbitrariness in the original allocation, which no doubt entails a social cost.

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 offense. 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. While much attention has been given to the incentives of criminals, we focus here on the incentives of courts as law enforcers. We assume that courts and society are perfectly in line, when it comes to deterrence.[6] The difference between society and courts in our model lies in the use of information on the cost of sentences. Although the preferences in our model can 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 given crimes, 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 with the optimal magnitude of fines and their use relative to imprisonment. See for example Polinsky and Shavell [1979], Polinsky and Shavell [1984], Friedman [1981] and Waldfogel [1995]. These papers build on the fact that imprisonment being more costly than fines, it should be used only when the criminal is unable to pay. Implicitly, they focus on "white collar crimes" such as fraud, theft or other property crimes. In this paper, we are interested in crimes that require other types of penalties than fines.

The remainder of the paper is organized 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 highlight a series of findings from our model. In Section 4, we present a historical example: the frequency of the death penalty in 17th and 18th centuries England. We conclude in Section 5.

[6]We are inclined to believe that judges do take deterrence into account when sentencing. In fact, deterrence is one of the important goals of sentencing, explicit in every criminal law textbook: these goals are justice, protection of the public, deterrence and deserts.

2 The model

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

Police has made an investigation and identified a suspect who now faces a court. In an adversary trial, the prosecution makes its case against the accused, the defense attempts to find weaknesses in the prosecution's argument, and as a result, the court[7] determines the probability q that the accused is indeed guilty of the crime. This probability reflects the underlying evidence as well as the amount of resources devoted by society to that particular case. For our purpose, the evidence and the resources allocated to the investigation as well as the trial proceeding are taken as exogenously given.

The court then chooses to either penalize or not, and contingent on a condemnation selects the appropriate level of the sentence denoted by p. Initially, we assume that there is a continuous choice of possible penalties $p \in [0, \infty)$. Dismissal of a case implies $p = 0$. For parsimony, we assume that both the sentence and the offense can be measured by the same metric and 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. As discussed in the introduction, we assume that these social costs are from the point of view of a court a pure externality.

Let x_a represent the actual offense of the accused. From the perspective of society and that of the court, they know that only one of two events is true: 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 either 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 each of these four outcomes. We assume that the preference structure satisfies the necessary requirements to be represented by a von Neumann-Morgenstern disutility function $u = u(\alpha)$ where α is the outcome of the trial. We impose usual regularity requirement, assuming u to be strictly increasing and convex in α, with $u(0) = u'(0) = 0$ (where $u'(.)$ denotes the first derivative of the function u). We impose two additional

[7] 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.

requirements regarding preferences. First, courts are thought to dislike gaps between the offense of the accused, x_a, and the penalty he receives, $(|x_a - p|)$. For example, courts are assumed to dislike sentencing to 25 years jail term someone who stole a piece of bread. Second, courts are taken to get an extra-disutility from condemning an innocent person.[8] 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 the additional moral cost of an unjust penalty, for example stemming from empathy.[9]

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 objective function.

If the court knew with certainty the status of the accused, sentencing a culprit (i.e. $x_a = x$) to a penalty p would leave the court with disutility $u(|x_a - p|) = u(|x - p|)$. After sentencing an innocent (i.e. $x_a = 0$) 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 ($u(|x_a - p|) = 0$). Finally, 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.[10] Table 1 summarizes the disutility of the court in the four alternatives.

Table 1: The court's cost

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

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

[9] Alternatively, assuming courts have no empathy, one can interpret ϕ as the future cost of having a genuine criminal outside. Indeed, by wrongfully closing the case, the court would also stop further police investigations of the particular crime with the undesirable side effect that the real culprit would remain unbothered and free to recidivate.

[10] 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.

2.1 The court

The court must decide both on guilt of the accused and possibly the severity of punishment. From its point of view, given the trial proceeding and the ensuing probability q that the accused has, indeed, committed the crime, the court must compare its disutilities of either condemning the defendant – denoted hereafter by u^C – or pronouncing a non-guilty verdict with disutility u^{NG}. Using the foregoing notation, the court can be viewed as solving the following nested problem:

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

Note that with zero penalty, we always have $qu(x) + (1-q)u(\phi) > qu(x)$, which guarantees that no court will ever find it optimal to simultaneously pronounce a guilty judgement and, yet, set $p = 0$. Hence convicting the accused implies a positive sanction ($p > 0$), while no penalty ($p = 0$) is equivalent to the court finding the accused innocent.

Call p^T the optimal penalty from the perspective of an unconstrained tribunal. p^T has the following unsurprising features: (i) it fits the crime (meaning $p = x_a$) only in the case of perfect information on the guilt of the accused ($q = 0$ or $q = 1$); (ii) as the probability of guilt converges to 1, p^T goes to x; (iii) more serious crimes require more severe penalties and (iv) the risk of convicting an innocent tempers the severity of the court (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 analytical proofs of these claims are left to the reader as they follow immediately from the first-order condition of the court's minimization problem u^C and, the use of the implicit function theorem. Finally, since these results imply $x \geqslant p^T$, we omit in the remainder the absolute value and write $|x - p^T| = x - p^T$.

2.2 Society

As the court, society is concerned with the case of the accused. But it is also concerned with the management of the sentence chosen by the court. If, as we argue, 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 that of implementing the sentence, thus adding the expenses $C(p)$. 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.

Table 2: Society's cost

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

Society's objective is to minimize the expected social cost associated with the judgment. This means comparing the cost of sentencing the accused – hereafter denoted by w^C – with the cost resulting from a non-guilty verdict represented by w^{NG}. Thus, a benevolent social planner should solve:

$$[P'] \quad \min \{ \underbrace{\min_p \; q\, u(|x-p|) + (1-q)\, u(p+\phi) + C(p)}_{w^C} \;,\; \overbrace{q\, u(x)}^{w^{NG}} \} \qquad (2)$$

Denote by p^S the socially optimal penalty. As for the court, we can show that p^S is increasing in x and in q, and decreasing in ϕ. Moreover, 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. Henceforth, we will rewrite $|x - p^S|$ as $x - p^S$.

3 Results

In this section, we present three results that follow from the above set up. First, society is shown to always prefer lower punishments than unrestricted courts would implement. The claim obtains directly from the externality that leads courts to ignore $C(p)$. Second, if the cost of punishing an innocent has a fixed component (i.e. $\phi > 0$), we show that both society and courts apply the principle of reasonable doubt when choosing whether to convict. That is, there exists a strictly positive minimum threshold of certainty for the guilt of

the accused below which defendants are never penalized. Moreover, due to the externality the optimal threshold which society would like to implement is larger than the one an unconstrained court would use – i.e. courts punish too often. It is worthwhile to note that the result can be equivalently formulated as the existence of a natural minimum sentence. Third, since society feels its courts punish too much and too often, it welcomes any instrument that helps reduce the strength of sentences and their frequency. Naturally society likes maximum sentences. More surprisingly, we show that it also likes minimum sentences: the intuition is that faced with the choice of imposing the minimum sentence or to let the accused go free, the constrained court will often choose the latter.

Society vs courts

Proposition 1 *Society prefers lighter penalties than unconstrained courts.*

Proof: Analytically, we must show that for all $q \in [0,1]$, $p^S(q) \leq p^T(q)$. Due to the non-negativity of sentences, we already know that with $p^S(q) = 0$ the inequality is satisfied. For the case $p^S(q) > 0$, society's problem has an interior solution. Analytically, $p^S(q)$ is then characterized by the equation

$$-qu'(x-p) + (1-q)u'(p+\phi) + \alpha C'(p) = 0 \ ,$$

where α is set at 1. To examine the decision of unconstrained courts, i.e. when α is set at 0, we take implicit differentiation with respect to α:

$$\frac{\partial p}{\partial \alpha} = -\frac{C'(p)}{qu''(x-p) + (1-q)u''(p+\phi) + \alpha C''(p)} < 0 \ .$$

Thus, p would increase if α were set to 0. This, however, yields the first-order condition considered by unrestricted courts, thereby, proving the claim.■

Reasonable doubt and natural minimum sentence

The differing tendencies of courts and society extend to the perception of *reasonable doubt*. Let us denote by q^{*T} and q^{*S} the probabilities of guilt which would make the tribunal and society respectively indifferent between either condemning or not condemning the accused. Both variables have a natural interpretation in terms of reasonable doubt. Specifically, if there is a moral cost of condemning an innocent, $\phi > 0$, then for the court to pronounce a guilty verdict requires a minimum probability $q \geqslant q^{*T} > 0$ that the accused is guilty. This in turn implies a minimum sentence bounded away from zero. As can be expected, society is more stringent than an unrestricted court on the reasonable doubt requirement. Thus, though courts and the society agree on the concept of reasonable doubt, yet they do not agree on what constitutes such a doubt.

Proposition 2 *If the disutility of punishing the innocent has a fixed component $\phi > 0$, then there is a minimum sentence that a court would want to impose. The same holds for society. Furthermore, courts punish more often than is socially desirable.*

Proof: We proceed in two steps:

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

   With $\phi > 0$, for the court to be indifferent between convicting or not requires

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

   which 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, proving the first claim.

2. $q^{*T} < q^{*S}$. We prove the claim by showing that at q^{*T} society would not condemn the defendant. This obtains 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 $C(p) \geq 0$. Claim 2 implies that society would convict less often than its courts. Finally, the discontinuity of p^T and p^S at q^{*T} and q^{*S}, respectively, implies that there is a *natural minimum sentence* for the court as well as for society.■

The key to understand the notion of natural minimum sentences lies 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.[11]

[11] 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.

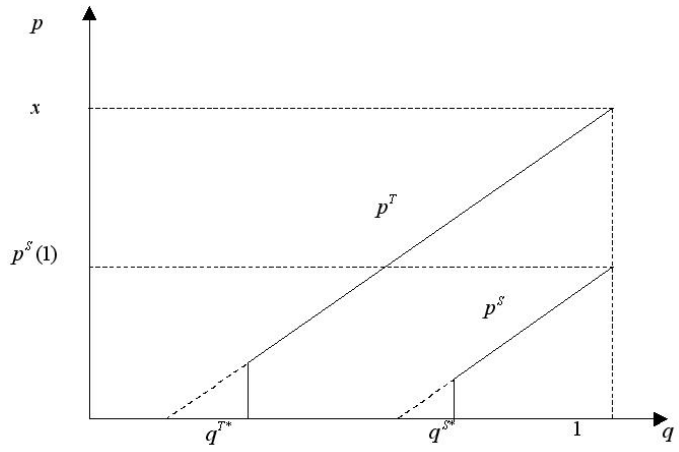

Figure 1: Optimal sentences: society versus courts

Figure 1 summarizes the differences between the behaviors of society and a representative court. Note that except for the linearity, Figure 1 is fully general. Hence we will make use of that representation for the ensuing analysis.

Minimum and Maximum Sentences

If courts are unconstrained in their choice of penalty, for every level of q, the distance between $p^T(q)$ and $p^S(q)$ measures the gap between courts' 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.[12] Specifically, the welfare loss can be represented within a figure similar to Figure 1 with some distortion.

We now examine the following question: Could society constrain courts

[12]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.

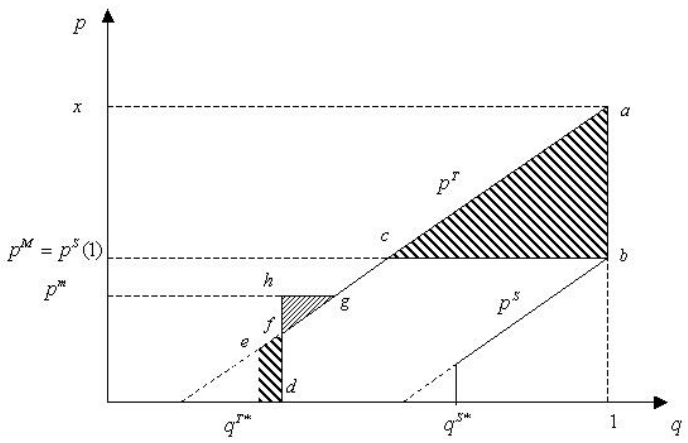

Figure 2: On the use of limits to sentences

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 a court, society does not observe q and cannot force the court to reveal its actual value.[13] Hence society can only restrict courts by imposing limits to sentences.

In Figure 2, it appears clearly that imposing a *maximum sentence* p^M reduces the welfare loss. For example a maximum sentence equal to the worst sentence society would want to assign, 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 show below.

Could *minimum sentences* help? A minimum sentence such as p^m in Figure 2 has two different effects. First, it pushes further the conviction threshold of the court. The latter will indeed require a higher probability of guilt to convict the accused to the minimum penalty. Hence, it reduces the

[13]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.

area between p^T and p^S by the area of trapezium $efdq^{*T}$, thereby increasing welfare. Second, the minimum sentence raises social cost due to 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 trade-off 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 state our main result: it is indeed in society's interest to set limits to sanctions. The upper-bound ought to be *small*, and the lower-bound *high* in the following sense:

Proposition 3 (Limits to sentences) *Society will want to constrain courts by imposing minimum and maximum sentences such that:*
(i) $p^*_m > p^T(q^{*T})$,
(ii) $p^*_M < p^S(1)$.

The proof goes as follows. That there exists a welfare improving maximum sentence is clear from Figure 2: $p^S(1)$ 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 2, and we let $\epsilon \to 0$. The trapezium $efdq^{*T}$ has a strictly positive side, eq^{*T}, while its width ,$q^{*T}d$, converges to zero; whereas the triangle fgh has both height fh and base hg converging simultaneously to 0. Hence, the welfare loss converges to zero faster than the welfare gain. This establishes the desirability of a minimum sentence satisfying (i) and characterized by the equality of marginal welfare gains and marginal welfare losses. To see why (ii) holds, we use a similar argument: 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 2 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 it optimal to set minimum penalties above the natural minimum penalty of courts. 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 an increase in the rate of acquittals. 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 maximum penalties below the highest penalty socially desirable – itself significantly lower than the highest one desired by courts.

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.

4 A historical example

A historical example of such responses of courts to high minimum penalties can be found in England in the 17th and 18th centuries. The number of crimes punishable by death was on the rise. By the end of this period, about 200 offenses, many of them minor property crimes, were by law assigned the death penalty. Capital crimes had very specific descriptions. They included arson, cattle-maiming, the destruction of trees, crimes known as sabotage, but also forgery, sheep-stealing, embezzlement of coal, poaching. The death penalty was applied to those breaking into houses with the goal of stealing linen or destroying it or more generally stealing goods worth forty shillings or more (Hay [1975]).[14]

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,"[15] or simply dismiss cases for which the death penalty was perceived by the court to be excessive. This last effect was further reinforced by the victims themselves, whose dislike for the death penalty made them more reluctant to prosecute offenders.[16] In case of conviction by the jury, judges could also use the royal pardon, by which a smaller pain could be substituted for hanging upon their recommendation. In fact, pardon was quite common: As estimated by Beattie [1986] and cited by Friedman [1995] about 60% of those sentenced to death were granted pardon by the judges.

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."[17] In essence, the extremely severe minimum penalty in this era, while helping society contain the cost of justice, had all potential to make it easier for criminals. In fact, Hay [1975] shows that the increasingly severe criminal code claimed surprisingly few lives: "At the beginning of the 17th century, for example, it appears that

[14] A typical monthly wage at the time was £2.

[15] Beattie [1974], page 83.

[16] Note, however, that Friedman [1999] argues that the death penalty, which benefits neither the victim nor the state, can be seen as a threat that facilitates rent-seeking through out-of-court settlements.

[17] Hay [1975], page 23.

London and Middlesex saw four times as many executions as 150 years later" and this, in spite of an increasing number of convictions for theft.

Changes in the "bloody" criminal code appear to have come for reasons partly supported by the model: the affordability of less morally painful punishments. Indeed, Friedman [1995] argues that the harshness of the 18th century English criminal code was driven by minimum cost considerations, and that pardons and out-of-court settlements increased the efficiency of the code, while not necessarily affecting deterrence. He claims that the subsequent reform of the criminal code, in the first half of the 19th century, substituting imprisonment to capital punishment, is a consequence of economic growth. As he puts it: "The preference for extreme punishments reflected the greater cost of the obvious intermediate punishment, and was relaxed as the society became richer." As prisons became affordable, English law makers turned to prison terms as the new minimum penalties.

England is not the only country in which courts respond to high minimum penalties by convicting less. 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 offenses: 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:

> 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]

Clearly, both these examples are extreme cases of minimum penalties and far exceed the severity of the minimum penalties we advocate as optimal in this paper.

Yet, they illustrate nicely the mechanism by which minimum penalties affect the behavior of courts and 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.

5 Conclusion

In this paper, we look for a socially optimal set of punishments, taking into account the preferences and incentives of courts. We show that minimum and maximum penalties may be used optimally to reduce the rate of convictions and the magnitude of sentences. Our argument relies 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 relatively high minimum penalties and low maximum ones is shown to help align courts with societal goals.

References

Abadinsky, H. [1978]. Parole history: an economic perspective. *Offender Rehabilitation, 2*(3). 275-278.

Andreoni, J. [1991]. Reasonable doubts 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, 1600-1800. *Past & Present, 62.* 47-95.

Beattie, J.M. [1986]. *Crime and the Courts in England, 1600-1800.* Princeton University Press. Princeton, NJ.

Becker, G.S. [1968]. Crime and punishment: an economic approach. *Journal of Political Economy, 76.* 169-217.

Blumstein, A. [1983]. Prisons: Population, Capacity, and Alternatives. In: *Crime and Public Policy*, J.Q. Wilson ed. ICS Press: San Francisco, California.

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.

Friedman, D. [1995]. Making sense of English law enforcement in the Eighteenth century. *University of Chicago Law School Roundtable, 2(2).* 475-505.

Friedman, D. [1999]. Why not hang them all: the virtues of inefficient punishment. *Journal of Political Economy, 107(6).* 259-269.

Giertz, J. F. & Nardulli, P. F. [1985]. Prison Overcrowding. *Public Choice, 46.* 71-78.

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 and C. Winslow Editors. Pantheon Books. New-York.

Polinsky, M. & Shavell, S. [1979]. The optimal trade-off 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, 69.* 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.

Wright, R. F. & Ellis, S. P. [1993]. A Progress Report on the North Carolina Sentencing and Policy Advisory Commission. *Wake Forest Law Review, 28*. 421-461.

Wright, R. F. [1998]. Three Strikes Legislation and Sentencing Commission Objectives. *Law and Policy, 18*. 429-463.