

Jail or Fine - Let Them Choose*

Alexander Klein[†]
University of Munich

January 19, 2000

Abstract

We reconsider two standard results of deterrence theory. The first states that there is no need for jail terms until monetary fines are not exhausted. The second says that there is no marginal deterrence without joint production in law enforcement.

In a framework with asymmetric information about offenders' wealth however, neither result holds if one takes into account a commonly used instrument of law enforcement, namely to offer the convicted criminal the choice of either going to jail or paying a fine.

The basic idea of our paper is that the fine can be adjusted such that only one of two types prefers it over the jail term. Thus, fine-jail-options serve as a screening device because they allow to deter both types independently although their wealth is private information.

JEL Classifications: K14, K42

Keywords: criminal law, nonmonetary sanctions, marginal deterrence

*I am grateful for helpful discussions with Patrick D'Souza, Georg Gebhardt, Susanne Krehelmer, Klaus Schmidt, Henrik Thiele, and Achim Wambach. The usual disclaimer applies.

[†]Department of Economics, University of Munich, Ludwigstr. 28 (Rg), 80539 Munich, Germany, Phone: +49 (89) 2180-3907, Fax: +49 (89) 2180-3510, email: Alexander.Klein@lrz.uni-muenchen.de

1 Introduction

The economic analysis of crime has its starting point with Becker's (1968) seminal work: individuals rationally decide whether to engage in a criminal activity by comparing the expected punishment with the returns to crime. Hence, crime is less attractive if the government increases the probability and/or the severity of punishment. Becker's main result is that since imposing a fine is a costless transfer, this fine should equal an individual's entire wealth and the probability of apprehension should be as low as possible to optimally deter crime. Two important questions however remained open: First, why do we observe jail sentences at all¹, and secondly, why are sanctions not always maximal in the real world? Becker (1968) and Posner (1977) argue informally for the use of fines whenever possible in order to avoid the social costs of maintaining prisons, the private disutility of incarceration, and the social waste of idle human capital in jails. Polinsky and Shavell (1984) have extended the approach of Becker (1968) by a jail term. Under the assumptions of perfect information on behalf of the social planner concerning each individual's wealth, and of a costless means of enforcing fines once the criminal is convicted, the optimal fine is the maximal fine. It is optimal to

¹Levitt (1997) summarizes three common arguments in favour of prison terms: First, fines are somewhat unjust because only the rich can afford them and the poor need to go to jail. Secondly, criminals usually have too low wealth, and thirdly the incapacitation value. All those arguments, however, lack a justification why the social planner does not at least take advantage of the remaining assets or the human capital of the offender.

supplement a fine with an imprisonment term if the maximal fine is not very large and the marginal cost of imprisonment is sufficiently small.

Levitt (1997) introduced fine-jail-options into the literature. His fundamental premise is that the government cannot enforce fines except under the threat of further punishment, in particular prison sentences. He gives three reasons for why it can be difficult to collect fines. First, there is private information about the agent's wealth. Second, fines are difficult to collect if wealth consists of human capital, and third prohibitive costs of collection. Levitt (1997) considers two types of criminals, rich and poor, where the rich have a larger disutility of being in jail. First he characterizes a setting where only jail sentences are available as a sanction device. Since each individual can only choose between committing the crime or not, there exist four possible outcomes. Then, he analyzes, in which of those initial equilibria fine-jail-options lead to strict Pareto improvements. The main result of his paper is, that this happens indeed if in a jail-only world only the rich would commit the crime.

The intuition is that the rich save on disutility of being in prison by redeeming this option and society avoids maintenance costs of prisons. The poor continue to be deterred from the crime. However, the analysis hinges on the assumptions that offenders' utility receives a positive weight in the social welfare function and that expenses for maintaining prisons are strictly

positive. Particularly the latter one does not allow to apply the result on other nonmonetary sanctions like, e.g., withdrawal of the driving licence.

In this paper, we consider a similar framework but with multiple crimes and identify a new role for fines: They enable marginal deterrence for one type without affecting deterrence of the other type. As a result, the crime rate decreases. Thus, we show how incentive compatible fine-jail-options can be optimal and why the result of Polinsky and Shavell (1984) need not hold with asymmetric information.² Although our result itself resembles that of Levitt (1997), the underlying effect is new.

Another puzzle of deterrence theory is why sanctions are not always maximal. The most obvious explanation is that of marginal deterrence. When faced with the possibility of committing one of several harmful acts, offenders need incentives to choose the least harmful one. A standard result in this branch of the literature was established by Wilde (1992). He stated that jointness in production of law enforcement is a necessary condition to have marginal deterrence in the social optimum.

In this paper however, where individuals have private information about their wealth, we demonstrate that if apprehension probabilities for both crimes can be set independently, it may nevertheless be optimal not to raise

²Chu and Jiang (1993) have a similar result but it does hold only when individuals are risk averse.

both sanctions to its maximum, that is to allow for marginal deterrence.³

The reason is that the only means to make the less harmful crime attractive to those who would otherwise choose the more damaging one is to offer an optional monetary fine for the less detrimental offense, which changes incentives only for those individuals.

Lowering enforcement expenses for the less damaging offense however would attract the hitherto deterred criminals to commit this crime in the first place.

As a result, we do have marginal deterrence in the optimum although enforcement effort can be chosen independently for each crime.

The paper is organized as follows: In Section 2 we derive the optimal policy in a jail-only-world. In Section 3 we introduce fine-jail-options and show how they are Pareto-improving. Section 4 allows for different probabilities of apprehension. It is shown that marginal deterrence is optimal nevertheless. Section 5 concludes.

³Friedman and Sjostrom (1993) also implicitly allowed for different probabilities of apprehension. They have shown that the expected utility loss from punishment (effective punishment in their terms) may fall with severity of crime. However, they needed to assume risk aversion on behalf of the offenders and increasing marginal punishment costs.

2 Optimal Punishment in a Jail-Only World

There are two crimes, C^1 and C^2 , where the social harm associated with C^2 is larger. Agent i , assumed to be risk neutral, maximizes the following expected utility function

$$E(U_i) = \begin{cases} 0 & \text{if no crime } (C^0) \text{ is committed} \\ \beta_i^1 - \alpha J^1 \theta_i & \text{if } C^1 \text{ is committed} \\ \beta_i^2 - \alpha J^2 \theta_i & \text{if } C^2 \text{ is committed} \end{cases} \quad (1)$$

where β_i^j is the private benefit to agent i of committing crime j , α is the detection probability conditional on having committed one of the two crimes.^{4,5} J^j is the length of the jail term if detected, and θ_i is a measure of the disutility of jail to agent i ⁶. For algebraic simplicity, it is assumed that the individual derives all benefits from the crime before the detection occurs; the substantive results of the analysis are unchanged when this assumption is altered. β_i^j may also reflect that the type of crime influences the likelihood of detection.

The social planner commits to a level of α , J^1 and J^2 before the agent's decision about whether or not to commit a crime, so as to maximize social

⁴Here we suppose joint production of law enforcement. This assumption is relaxed in section 4.

⁵Note that the probability of conviction is one once detected. Some papers, e.g. Mookherjee and Png (1994), allow for separate probabilities for detection and conviction.

⁶For simplicity we assume the disutility of jail being linear in the length of imprisonment. Polinsky and Shavell (1999) took into account offenders' discounting of disutility.

welfare W given by

$$\begin{aligned} \max_{\alpha, J^1, J^2} W &= \gamma \sum_{i=1}^N E[U_i^*(\beta_i^1, \beta_i^2, \theta_i, \alpha, J^1, J^2)] - \sum_{j=1,2} c^j(\alpha, J^1, J^2) \cdot [L^j + \alpha J^j \psi] - e(\alpha) \\ \text{s.t. } E[U_i^*] &\in \arg \max_{j \in \{0,1,2\}} E[U_i] \quad \forall i \end{aligned} \quad (2)$$

where U_i^* reflects agent i 's utility maximizing choice of behavior, γ is the criminals' weight in social welfare, c^j are the numbers of agents committing crime C^j (which depend on the social planner's choice of policies), L^j is the social harm associated with the respective crime, ψ is the cost borne by society of incarcerating offenders that are detected (assumed to increase linearly with the number of years served in prison⁷) and e is the enforcement cost associated with detection probability α . The social planner maximizes W subject to the incentive compatibility constraint for each agent.

Assumption 2.1 *The following restrictions are made concerning the parameter values: (i) $0 < \beta_i^j < L^j \quad \forall i, j$, (ii) $L^2 > L^1$, (iii) $\beta_i^2 > \beta_i^1 \quad \forall i$, (iv) $0 \leq J^j \leq J^{max} \quad \forall j$, (v) $\psi \geq 0$, (vi) $e'(\alpha) > 0$, $e''(\alpha) > 0$, (vii) $i \in \{R, P\}$, (viii) $\theta_R > \theta_P$, (ix) An agent's type i is private information. (x) $0 \leq \gamma \leq 1$*

(i) excludes socially desirable crimes. The main conclusions of the paper remain unchanged if one allows for those acts. (ii) means that C^2 causes more harm than C^1 . For example, C^1 is a modest exceeding of the speed limit and

⁷Society may also discount future expenses. See again Polinsky and Shavell (1999) for an analysis.

C^2 dangerous speeding. (iii) reflects that private benefits are monotonically increasing in social loss. (iv) imposes an upper bound on jail sentences. This assumption is standard to avoid the Becker (1968) result of infinite punishments, and it can be motivated by moral obligations, constitutional restrictions, or simply by the fact that people have finite lives. (v) says that there are costs of maintaining prisons⁸ (vi) implies that enforcement is costly and the cost function is convex in the probability of apprehension. (vii) limits the analysis to the case of two types, R (ich) and P (oor). (viii) characterizes the difference between the rich and the poor type, that is the rich does suffer more when serving a jail sentence. (ix) implies that the social planner cannot tailor punishments directly conditional on the individual's type. (x) restricts the weight the criminals receive in social welfare All parameters are common knowledge.

The social planner adopts the two-step-procedure of Grossman and Hart (1983) to solve his maximization problem. First he looks for the cost minimizing punishment scheme to implement each of the nine possible optima (three pooling and six separating outcomes). Secondly, he compares the corresponding values of the welfare function and chooses the dominant one. We will not solve this problem explicitly because the solution hinges on the func-

⁸A possible interpretation is that these cost overcompensate any possible benefits of incapacitating potential repeat offenders, although the latter is not modelled explicitly.

tional form of the target function. Instead, we highlight some properties of the social optimum which will become useful later on.

The solution to the individual's maximization problem is straightforward. Crime j is committed only if the disutility of expected punishment is less than the private benefit to crime. If the latter condition is fulfilled for both crimes, the individual opts for the offense which entails larger net benefits.

$$E(U_i^*) = \begin{cases} 0 & \text{if } \alpha J^1 \geq \frac{\beta_i^1}{\theta_i} \text{ and } \alpha J^2 \geq \frac{\beta_i^2}{\theta_i} \\ \beta_i^1 - \alpha J^1 \theta_i & \text{if } \alpha J^1 < \frac{\beta_i^1}{\theta_i} \text{ and } \beta_i^1 - \alpha J^1 \theta_i \geq \beta_i^2 - \alpha J^2 \theta_i \\ \beta_i^2 - \alpha J^2 \theta_i & \text{if } \alpha J^2 < \frac{\beta_i^2}{\theta_i} \text{ and } \beta_i^2 - \alpha J^2 \theta_i > \beta_i^1 - \alpha J^1 \theta_i \end{cases} \quad (3)$$

Given (3), the planner maximizes social welfare (7). The following results follow directly from the setup of the problem. They characterize the solution and will become useful later in the paper.

Lemma 2.1 *The optimal policy in a jail-only-world involves*

- (i) *In the social optimum, $J^1 < J^2$.*
- (ii) *The social planner always sets $J^2 = J^{max}$.*
- (iii) *If both agents commit C^2 at the social optimum, $\alpha = 0$.*
- (iv) *If one agent commits C^2 and the other C^1 in the optimum or if both agents choose C^1 , $J^1 = 0$.*

Proof: See Appendix A.

Lemma 2.2 (i) *If in the social optimum the poor commits no crime and the rich commits C^2 , then*

$$\frac{\beta_P^2 - \beta_P^1}{\beta_R^2 - \beta_R^1} < \frac{\theta_P}{\theta_R} \quad (4)$$

(ii) *If in the social optimum the rich commits no crime and the poor commits C^2 , then*

$$\frac{\beta_R^2 - \beta_R^1}{\beta_P^2 - \beta_P^1} < \frac{\theta_R}{\theta_P} \quad (5)$$

Proof: See Appendix B.

From Lemma 2.2 (i), those crimes which should be committed only by the rich are those where their marginal benefit of committing the more severe variant is relatively large compared to the marginal benefit of the poor. Furthermore, the rich should not suffer too much when the sanction is non-monetary. As an example, consider again traffic violations like speeding. The nonmonetary sanction here can be interpreted as loss of the driving licence. The rich may employ a driver, which is not feasible for the poor. Nevertheless, if the rich is a busy professional, it may be particularly worthwhile for him to exceed the speed limit even more.

3 Fine-Jail-Options as an Additional Instrument in a Jail-Only World

We now introduce the following instrument into the model: Once convicted, the social planner offers the criminal the choice $\{F^j, J^j\}$ of paying a fine and foregoing the prison sentence.⁹ We continue to assume that the social planner has no means of mandating the criminal to pay the fine if the criminal refuses. However, the offender can be forced to go to jail. As before, an agents wealth is private information so that punishments cannot be made directly conditional on wealth. To keep the analysis simple, we assume $\gamma = 1$. Furthermore, we do not impose an exogenous upper bound on the monetary fine.

Assumption 3.1 *The fines $F^1, F^2 \in [0, \infty]$.*¹⁰

The individual's utility function becomes

$$E(U_i) = \begin{cases} 0 & \text{if no crime committed} \\ \beta_i^1 - \alpha \cdot \min\{J^1\theta_i, F^1\} & \text{if } C^1 \text{ is committed} \\ \beta_i^2 - \alpha \cdot \min\{J^2\theta_i, F^2\} & \text{if } C^2 \text{ is committed} \end{cases} \quad (6)$$

⁹This policy is very common for example in Germany. Sanctions for minor crimes often are expressed in daily rates. The offender can substitute each day in prison with the payment of his daily income. Of course, the problem of determining this figure remains.

¹⁰Note that this assumption does not reuscite Becker's result of infinite punishments. With asymmetric information, convicted offenders would deny the ability to pay such a fine.

The social welfare function is now given by

$$\begin{aligned} \max_{\alpha, J^1, J^2, F^1, F^2} W &= \alpha \sum_j f^j F^j + \sum_{i=1}^N E[U_i^*(\cdot)] - \sum_{j=1,2} c^j(\cdot) \cdot [L^j + (1-f(\cdot)\alpha J^j \psi)] - e(\alpha) \\ \text{s.t. } E[U_i^*] &\in \operatorname{argmax} E[U_i] \quad \forall i \end{aligned} \quad (7)$$

where f^j is the percentage of those agents who elect to pay the fine once detected.¹¹

Proposition 3.1 *(i) If the jail-only optimum involves the rich committing crime C^2 and the poor successfully deterred from any crime, there exist a fine-jail-option such that the rich commits C^1 and the poor is still deterred from any crime. In this case, social welfare increases unambiguously.*

(ii) If in the jail-only optimum the poor commits crime 2 and the rich is deterred from any crime, and if the poor has a larger marginal benefit of crime, $\beta_P^2 - \beta_P^1 > \beta_R^2 - \beta_R^1$ there exist a fine-jail-options such that the poor commits C^1 and the rich is still deterred from any crime. Again, social welfare increases unambiguously.

Suppose, as in Lemma 2.2, that in the social optimum of the jail-only-world, the rich commits crime C^2 and the poor does not commit any offense. Why does the social planner not lower J^1 ? Because a lower J^1 would attract the poor to commit crime C^1 as well (instead of neither crime). In this

¹¹To keep the analysis simple it is assumed implicitly that any proceeds from collecting fines are not distributed among potential or actual offenders in order to avoid any distortions in their decision.

particular case, fine-jail-options can be constructed for C^1 such that they will only be redeemed by the rich, who in response switch to C^1 . Thus, they allow the social planner to separately deter both groups and to adjust the lower punishment for the rich without violating the corresponding incentive constraint of the poor.

Social welfare increases for three reasons. First, the rich has higher net benefits, otherwise he would not change his behavior. Secondly, society saves on maintenance costs of $\alpha c^2 J^2 \psi$. Third, and this effect is new, there is less criminal activity. Those who caused a social Loss L^2 before, now are just responsible of harm $L^1 < L^2$.

What is the difference to the result of Levitt (1997)? He considers a one-crime model, where the introduction of fine-jail-options leads to a welfare improvement as well, if in a jail-only-world only the rich would commit the crime. However, the only driving force is that the rich, who opt to pay the monetary fine, save on disutility from being in prison and that society foregoes maintenance expenses. Each type still chooses the same offense as before. In our model, the option to pay a fine really serves as a screening device: The existence of this instrument changes the ex ante behavior (at least of the rich). Thus, the source of the welfare improvement stems not only from the 'rich' who need not go to jail but also from another (a better) allocation of crimes. Furthermore, Levitt's result breaks down if there are

no social costs associated with the nonmonetary sanction and if offenders' utility has no weight in the social welfare function, i.e. $\gamma = 0$ Particularly the second assumption is hard to justify. In this paper, Proposition 3.1 continues to hold if one relaxes both assumptions. This allows us to apply our argument to other nonmonetary sentences.

Nevertheless, there are other jail-only-equilibria with the potential to be improved by fine-jail-options. In those situations, offenders do not change their ex-ante-behavior, though.

Corollary 3.1 *If in the jail-only-optimum*

(i) the rich commits C^2 and the poor C^1 or

(ii) the rich commits C^1 and the poor neither crime,

then the introduction of fine-jail-options strictly improves social welfare

if and only if $\psi > 0$ or $\gamma > 0$

Proof: See Appendix D.

Corollary 3.2 *If in the jail-only-optimum*

(i) both individual commit C^2 or

(ii) both types commit C^1 or

(iii) both types commit neither crime or

(iii) the poor commits C^2 and the rich C^1 ,

then the introduction of fine-jail-options does not improve social welfare.

In case (i), $\alpha = 0$ by Lemma 2.1 (iii). But without apprehension there is no conviction and fine-jail-options cannot have any effect. In case (ii), $J^1 = 0$ by Lemma 2.1 (iv). Obviously, no individual would choose to pay a fine when convicted of crime 1. Again, fine-jail-options cannot have any effect. In case (iii), both criminals are already completely deterred. For the proof of case (iii), see Appendix E. Again, there is no potential for fine-jail-options to improve social welfare. Corollaries 3.1 summarizes those situations where the story of Levitt (1997) applies. Only in one of nine possible initial equilibria, namely that where the poor commits C^1 and the rich neither crime, one cannot make unambiguous predictions how fine-jail-options work in our model.

4 Relaxing the Assumption of Joint Production in Law Enforcement

Beside the question why we observe jail sentences despite their social cost, Becker (1968) posed a second puzzle: Why are sanctions not always maximal? One common argument is that of marginal deterrence¹². When faced with the possibility of committing one of several harmful acts, offenders need

¹²The seminal paper is Stigler(1970). See also Shavell (1992) and Mookherjee and Png (1994).

incentives to choose the least harmful one.

Wilde (1992) established the result that jointness in the production of law enforcement¹³ is a necessary condition to have marginal deterrence in the social optimum. The intuition is clear: Suppose one could set probabilities of apprehension independently for both crimes. Criminal choice still only depends on expected sanctions, though. Thus, if either sanction is less than the maximum possible sanction, social welfare can be increased by saving on the resources devoted to apprehension for either crime and raising the corresponding sanction so as to hold expected punishment constant. So a constant level of deterrence is achieved with lower apprehension expenses. Suppose now that, in our model, we allow for different probabilities of conviction, that is the social planner can set α^1 and α^2 independently.

Proposition 4.1 *(i) Suppose that in the jail-only-optimum with independent probabilities of apprehension the rich chooses C^2 and the poor is committing neither crime. Then there exists a fine-jail-option $\{F^1, J^1\}$ for crime C^1 such that the rich commits C^1 and the poor remains to be deterred from any crime.*

(ii) If in the jail-only-optimum the poor commits C^2 and the rich neither crime and if the poor has a higher marginal benefit of crime, then there exists a fine-jail-option $\{F^1, J^1\}$ for crime C^1 such that the poor commits C^1 and

¹³That is the planner can set only a uniform apprehension probability for all crimes.

the rich remains to be deterred from any crime.

In both circumstances social welfare increases unambiguously and the optimal fine F^1 associated with crime 1 is strictly less than the monetary equivalent of the jail term for the other crime, which is the maximal one.

Proof: See Appendix F.

The intuition is obvious. Consider part (i) of Proposition 4.1. Without fine-jail-options, both jail terms are maximal. Offering to pay a monetary fine F^1 must attract the rich into C^1 and still deter the poor. It is shown in the Appendix that this is always possible, but a sufficient condition is that $F^1 < \theta_R J^1 = \theta_R J^{max} = \theta_R J^2$. This means that the actual fine F^1 for crime C^1 is indeed less than what the offender has to expect when committing C^2 . That is, the intuition of Wilde (1992) may fail if one drops the assumption of perfect information regarding offenders' wealth. Why is it not possible to decrease α^1 further? Because this policy would attract the other type to commit C^1 in the first place. The only means however to make C^1 attractive to those who would otherwise choose C^2 is to offer an incentive-compatible monetary fine. Thus, in spite of independent apprehension probabilities we do have marginal deterrence.

The argument for part (ii) is analogous. In all other initial equilibria the introduction of fine-jail-options leads to very similar insights as in section 3,

for this reason we do not discuss them in detail.

5 Conclusion

In this paper we have reconsidered two standard results of the literature on optimal deterrence. As opposed to Polinsky and Shavell (1984) we have constructed a model where jail terms coexist with less than maximal monetary fines. Furthermore, if we allow for independent probabilities of apprehension for different crimes, the social optimum may nevertheless stipulate different sanctions for different crimes. This is in sharp contrast to Wilde (1992). These results are due to our assumption that the authority does not have access to offenders' wealth, either because it consists mainly of human capital or it is simply well concealed. This information asymmetrie can be overcome by incentive compatible fine-jail-options, which allow the convicted offender to choose between a fine and a jail term (or any other non-monetary sanction). Such an instrument allows screening between rich and poor offenders, where 'rich' individuals do suffer more when serving a prison term.

Appendix

A Proof of Lemma 2.1

A.1 Proof of part (i)

Assume $J^1 > J^2$. Now suppose J^1 is reduced to J^2 . It will be shown that this policy allows for a weak Pareto improvement.

If both individuals are deterred completely, nothing changes.

If both types commit crime 1, the principal can decrease $e(\alpha)$ without violating the other constraints.

If both types commit crime 2, either nothing is changed or at least one type can be attracted into C^1 instead. This would unambiguously increase social welfare: First, the respective individual receives higher net benefits, otherwise he would not switch to C^1 . Secondly, criminal impact is reduced and thirdly, maintenance costs stay constant (J^1 lasts as long as J^2 before). A contradiction of a social optimum.

If one type opts to commit C^1 anyway and the other is deterred from any crime, society can save on maintenance costs by lowering J^1 without violating the constraints of the other type.

If any type commits C^2 before and the other is completely deterred,

nothing changes.

If either type chooses C^2 and the other C^1 , the latter (and society) benefits from decreasing J^1 . The former either remains unaffected by this policy or prefers C^1 afterwards. In the latter case, social welfare increases unambiguously, a contradiction.

Now, suppose $J^1 = J^2$. Obviously, the net benefit of committing C^2 is strictly higher than that of C^1 . That is, an infinitesimal decrease of J^1 can neither attract offenders of C^2 into crime 1, nor those who are completely deterred. Thus, this solution is weakly dominated by $J^2 > J^1$. \square

A.2 Proof of part (ii)

The level of deterrence depends on α , J^1 and J^2 only through their respective product. Holding αJ^1 and αJ^2 fixed in (7), $U_i^*(\cdot)$, $c^1(\cdot)$, $c^2(\cdot)$ and the social costs of incarceration are unaffected as the ratio of α to J^2 changes. Enforcement costs $e(\alpha)$, however, are strictly increasing in α . Thus, for a given level of deterrence, social welfare is maximized by setting $J^2 = J^{max}$. Since $J^{max} \geq J^2 > J^1$ by part (i) of the lemma, it is always possible holding both products fixed. \square

A.3 Proof of part (iii)

Suppose that allowing both types to commit crime 2 is the solution of (2). Because enforcement is costly, and the only benefit is deterrence, if no deterrence occurs at the social optimum, the cost-minimizing enforcement level, $\alpha = 0$, must be chosen. \square

A.4 Proof of part (iv)

Suppose the rich commits C^2 and the poor C^1 . Then it must be true that

$$\beta_P^1 - \alpha\theta_P J^1 \geq \beta_P^2 - \alpha\theta_P J^{max} \quad (8)$$

Suppose, $J^1 > 0$. Obviously, lowering J^1 by ΔJ^1 , holding α constant, improves social welfare because the poor who commit C^1 anyway need not spend any time in jail and society saves $\psi\alpha c^1 \Delta J^1$. Next it is demonstrated that this is the only effect onto social welfare: Concerning the rich it must be true that

$$\beta_R^1 - \alpha\theta_R J^1 < \beta_R^2 - \alpha\theta_R J^{max} \quad (9)$$

Suppose there is some $\hat{J}^1 < J^1$ such that R is attracted by crime 1. Then welfare would increase unambiguously: R would have larger net benefits, maintenance costs decrease by $c^2\alpha\psi(J^{max} - \hat{J}^1)$ and social harm by $c^2(L^2 - L^1)$. Thus, a contradiction to an optimum and $J^1 = 0$.

If both agents choose C^1 , the incentive constraints look like

$$\beta_R^1 - \alpha\theta_R J^1 \geq \beta_R^2 - \alpha\theta_R J^2 \quad (10)$$

$$\beta_R^1 - \alpha\theta_R J^1 \geq 0 \quad (11)$$

$$\beta_P^1 - \alpha\theta_P J^1 \geq \beta_P^2 - \alpha\theta_P J^2 \quad (12)$$

$$\beta_P^1 - \alpha\theta_P J^1 \geq 0 \quad (13)$$

Obviously, no incentive constraint is violated, if J^1 is decreased, but maintenance costs become less, a contradiction to a social optimum. Therefore, $J^1 = 0$. \square

B Proof of Lemma 2.2

In the solution it must be true that

$$\beta_R^2 \geq \alpha\theta_R J^2 \quad (14)$$

$$\beta_R^2 - \alpha\theta_R J^2 \geq \beta_R^1 - \alpha\theta_R J^1 \quad (15)$$

$$\beta_P^2 \leq \alpha\theta_P J^2 \quad (16)$$

$$\beta_P^1 \leq \alpha\theta_P J^1 \quad (17)$$

First, note that $J^2 = J^{max}$ by Lemma 2.1. From (16),

$$\alpha \geq \frac{\beta_P^2}{\theta_P J^{max}} \quad (18)$$

Suppose (18) does not hold with equality. Then, social welfare can be improved by saving on enforcement cost, holding αJ^1 constant. (Since $J^1 < J^{max}$ by Lemma 2.1, this is always possible.) Thus,

$$\alpha = \frac{\beta_P^2}{\theta_P J^{max}} \quad (19)$$

Now suppose J^1 could be lowered such that the rich type commits to crime C^1 without violating (17). Then social welfare unambiguously rises. This contradicts the presumption of a social optimum. Thus, if J^1 is sufficiently low such that (15) holds with equality, (17) can not hold anymore.

To calculate the critical \hat{J}^1 , suppose (15) holds with equality:

$$\begin{aligned} \beta_R^2 - \alpha \theta_R J^{max} &= \beta_R^1 - \alpha \theta_R J^1 \\ \Leftrightarrow \hat{J}^1 &= J^{max} - \frac{\beta_R^2 - \beta_R^1}{\alpha \theta_R} \end{aligned} \quad (20)$$

(17) does not hold when $J^1 = \hat{J}^1$:

$$\begin{aligned} \beta_P^1 &> \alpha \theta_P \hat{J}^1 \\ \Leftrightarrow \beta_P^1 &> \alpha \theta_P \left(J^{max} - \frac{\beta_R^2 - \beta_R^1}{\alpha \theta_R} \right) \end{aligned} \quad (21)$$

Using (19) gives

$$\begin{aligned} \beta_P^1 &> \beta_P^2 - \frac{\theta_P}{\theta_R} (\beta_R^2 - \beta_R^1) \\ \Leftrightarrow \frac{\beta_P^2 - \beta_P^1}{\beta_R^2 - \beta_R^1} &< \frac{\theta_P}{\theta_R} \end{aligned} \quad (22)$$

The proof of part (ii) is analogous. \square

C Proof of Proposition 3.1

C.1 Proof of part (i)

Consider first a jail-only-world. In the particular solution of Lemma 2.2(i) it must be true that

$$\beta_R^2 \geq \alpha\theta_R J^{max} \quad (23)$$

$$\beta_R^2 - \alpha\theta_R J^{max} \geq \beta_R^1 - \alpha\theta_R J^1 \quad (24)$$

$$\beta_P^2 \leq \alpha\theta_P J^{max} \quad (25)$$

$$\beta_P^1 \leq \alpha\theta_P J^1 \quad (26)$$

Note that (19) still applies.

It is shown that there exist a fine-jail-option $\{F^1, J^1\}$ attracting the rich into C^1 and still deterring the poor. Then F^1 must fulfill

$$\beta_R^1 - \alpha F^1 \geq \beta_R^2 - \alpha\theta_R J^{max} \quad (27)$$

and

$$\alpha F^1 > \beta_P^1 \quad (28)$$

Combining (27) and (28) gives

$$\beta_P^1 < \alpha F^1 \leq \alpha\theta_R J^{max} - (\beta_R^2 - \beta_R^1) \quad (29)$$

Using (19)

$$\beta_P^1 < \alpha F^1 \leq \frac{\theta_R}{\theta_P} \beta_P^2 - (\beta_R^2 - \beta_R^1) \quad (30)$$

For a solution to exist we need to show that

$$\begin{aligned} \beta_P^1 &< \frac{\theta_R}{\theta_P} \beta_P^2 - (\beta_R^2 - \beta_R^1) \\ \Leftrightarrow \frac{\theta_P}{\theta_R} &< \frac{\beta_P^2}{\beta_P^1 + (\beta_R^2 - \beta_R^1)} \end{aligned} \tag{31}$$

Using Lemma 2.2(i) it is sufficient to show that

$$\frac{\beta_P^2}{\beta_P^1 + (\beta_R^2 - \beta_R^1)} > \frac{\beta_P^2 - \beta_P^1}{\beta_R^2 - \beta_R^1} \tag{32}$$

Simple algebra gives

$$\frac{\beta_P^2 - \beta_P^1}{\beta_R^2 - \beta_R^1} < 1 \tag{33}$$

But this condition is always fulfilled in the supposed solution, since $\theta_R > \theta_P$.

□

C.2 Proof of part (ii)

Similar reasoning leads to the condition

$$\frac{\beta_P^2 - \beta_P^1}{\beta_R^2 - \beta_R^1} \geq 1 \tag{34}$$

for a feasible fine-jail-option to exist. Using Lemma 2.2(ii) this is possible

but not ensured by the assumption that $\theta_R > \theta_P$. □

D Proof of Corollary 3.1

D.1 Proof of Part (i)

From Lemma 2.1 (iv), $J^1 = 0$. In the optimum,

$$\beta_R^2 - \alpha\theta_R J^{max} \geq \beta_R^1 \quad (35)$$

$$\beta_P^2 - \alpha\theta_P J^{max} \leq \beta_P^1 \quad (36)$$

Suppose (36) does not hold with equality. Then it would be possible to decrease α without violating (35), a contradiction. Thus,

$$\alpha = \frac{\beta_P^2 - \beta_P^1}{\theta_P J^{max}} \quad (37)$$

Obviously, there cannot exist a fine-jail-option $\{F^1, J^1\}$ attracting the rich into C^1 . Instead, it is shown that there exist an F^2 such that R chooses to pay the fine without reversing the incentive constraint of P . Such a fine must fulfill

$$F^2 \leq \theta_R J^{max} \quad (38)$$

$$\beta_P^1 \geq \beta_P^2 - \alpha F^2 \quad (39)$$

Combining (38) and (39) gives

$$\frac{\beta_P^2 - \beta_P^1}{\alpha} \leq F^2 \leq \theta_R J^{max} \quad (40)$$

Using (37) we need to show

$$\frac{\beta_P^2 - \beta_P^1}{\beta_P^2 - \beta_P^1} J^{max} \theta_P \leq \theta_R J^{max} \quad (41)$$

The last expression reduces to $\frac{\theta_R}{\theta_P} \geq 1$, which is always true. \square

D.2 Proof of part (ii)

In equilibrium, it must be true that

$$\beta_P^1 - \alpha\theta_P J^1 \leq 0 \quad (42)$$

To obtain a welfare improvement, F^1 must fulfill

$$F^1 \leq \theta_R J^1 \quad (43)$$

and

$$\beta_P^1 - \alpha F^1 \leq 0 \quad (44)$$

Thus, it must be true that

$$\theta_R J^1 \geq \frac{\beta_P^1}{\alpha} \Leftrightarrow \alpha J^1 \geq \frac{\beta_P^1}{\theta_R} \quad (45)$$

From (42), this condition always holds. \square

E Proof of Corollary 3.2

E.1 Proof of Part (iv)

The proof is equivalent to that of Corollary 3.2 (i), but instead of (41) we

have

$$\frac{\theta_P}{\theta_R} \geq 1 \quad (46)$$

as the condition for an F^2 to exist, which does not attract the rich into C^2 but is chosen by the poor when convicted. Since $\theta_R > \theta_p$, a contradiction. \square

F Proof of Proposition 4.1

Consider first a jail-only-world. In the solution of Lemma 2.2 it must be true that

$$\beta_R^2 \geq \alpha^2 \theta_R J^{max} \quad (47)$$

$$\beta_R^2 - \alpha^2 \theta_R J^{max} \geq \beta_R^1 - \alpha^1 \theta_R J^{max} \quad (48)$$

$$\beta_P^2 \leq \alpha^2 \theta_P J^{max} \quad (49)$$

$$\beta_P^1 \leq \alpha^1 \theta_P J^{max} \quad (50)$$

Note that, by standard arguments, $J^1 = J^2 = J^{max}$. First, (48) must hold with strict inequality. Suppose not. Then the rich is indifferent between committing C^1 and C^2 . From this it follows immediately that $\alpha^1 > \alpha^2$. Therefore, (50) must hold with inequality. Assume, in equilibrium, the rich commits C^1 . Then the social planner can lower α^1 a bit without violating (50). Social welfare increases unambiguously, a contradiction. Assume instead, in equilibrium, the rich chooses C^2 . As before, social welfare will be increased by lowering α_1 . Now the welfare effect not only depends on the saving on apprehension costs but also on the rich committing C^1 instead of C^2 . Again, a contradiction. Thus, (48) holds with strict inequality.

Next, (49) holds with equality. Suppose not. Then one can lower α^2 without violating (47) or (48). Therefore,

$$\alpha^2 = \frac{\beta_P^2}{\theta_P J^{max}} \quad (51)$$

Now, holding α^1 constant, introduce a fine-jail-option $\{F^1, J^1\}$ for crime C^1 , such that the rich switch to C^1 and the poor continue to be deterred completely. Using (51) the condition that such an F^1 exists, reduces to

$$\beta_P^1 < \alpha F^1 \leq \frac{\theta_R}{\theta_P} \beta_P^2 - (\beta_R^2 - \beta_R^1) \quad (52)$$

But this is the same condition as (30) which was shown to hold in the proof of Proposition 3.1. (It is easy to show that Lemma 2.2 still applies.)

Obviously, $F^1 < \theta_R J^{max}$. Otherwise it would not be possible to reverse the incentive compatibility constraint (48), which holds with strict inequality.

Is it possible to raise social welfare by lowering α^1 ? Suppose yes. Then this would have been possible before the introduction of fine-jail-options. A contradiction of a social optimum. \square

References

- [1] Avio, Kenneth L. (1998), "The Economics of Prisons", *European Journal of Law and Economics*, 6, 143-175.
- [2] Becker, Gary S. (1968), "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, 76, 169-217.
- [3] Chu, C.Y. Cyrus and Neville Jiang (1993), "Are fines more efficient than imprisonment?", *Journal of Public Economics*, 51, 391-413.
- [4] Friedman, David and William Sjoström (1993), "Hanged for a Sheep-The Economics of Marginal Deterrence", *Journal of Legal Studies*, 22, 345-366.
- [5] Garoupa, Nuno (1997), "The Theory of Optimal Law Enforcement", *Journal of Economic Surveys*, 11, No.3, 267-295.
- [6] Grossman, Sanford J. and Oliver D. Hart, "An analysis of the principal-agent problem", *Econometrica*, 51, 7-45.
- [7] Levitt, Steven D. (1997), "Incentive Compatibility Constraints as an Explanation for the Use of Prison Sentences Instead of Fines", *International Review of Law and Economics*, 17, 179-192.

- [8] Mookherjee, Dilip and I.P.L. Png (1994), "Marginal Deterrence in Enforcement of Law", *Journal of Political Economy*, 102, 1039-1066.
- [9] Polinsky, Mitchell A. and Steven Shavell (1999), "On the Disutility and Discounting of Imprisonment and the Theory of Deterrence", *Journal of Legal Studies*, 28, 1-16.
- [10] Polinsky, Mitchell A. and Steven Shavell (1999a), "The Economic Theory of Public Enforcement of Law", NBER Working Paper 6993.
- [11] Posner, Richard (1977), "Economic Analysis of Law", Little-Brown, Boston.
- [12] Shavell, Steven (1987), "The Optimal Use of Nonmonetary Sanctions as a Deterrent", *The American Economic Review*, 77(4), 584-592.
- [13] Shavell, Steven (1992), "A Note on Marginal Deterrence", *International Review of Law and Economics*, 12, 345-355.
- [14] Kaplow, Louis (1990), "A Note on the Optimal Use of Nonmonetary Sanctions", *Journal of Public Economics*, 42, 245-247.
- [15] Wilde, Louis L. (1992), "Criminal Choice, Nonmonetary Sanctions, and Marginal Deterrence: A Normative Analysis", *International Review of Law and Economics*, 12, 333-344.