Plea Bargaining with Budgetary Constraints and Deterrence

Joanne Roberts[1]
Department of Economics
University of Toronto
Toronto, ON M5S 3G7
Canada
jorob@chass.utoronto.ca

March 23, 2000

Abstract

In this paper, I construct a simple model that illustrates conditions under which increased criminal sanctions can lead to increased levels of crime. This finding is derived from the interaction of binding budgetary constraints and plea bargaining, given that the costs of trial are assumed to be increasing in the size of sanction. In an environment with these institutional features, maximal sanctions are not optimal when resources are limited, and increased sanctions cannot generally be viewed as a substitute for increased monitoring. In this framework, increasing sanctions for different offences proportionally can lead criminals to substitute between offences. In fact, increased sanctions can lead to more severe crime. This effect is unambiguous when the marginal cost of trial is constant or increasing. The increasing cost of trial can imply that even when a proportional sanction increase implies a reduction in total crime levels it may imply an increase in severe offences, since some minor criminals will substitute into more severe crimes. This model also suggests that increased resources for prosecutors deter crime.

[1]Special thanks to Dan Bernhardt and Patrick Francois for encouragement and supervision. Chris Ferrall, Steeve Mongrain, John Spicer, and Dan Usher provided helpful comments. I am grateful to the Social Science and Humanities Research Council of Canada for financial support. All errors are my own.

1 Introduction

Increasing penalties has long been believed to be a crucial strategy for decreasing the incentives to commit crime, and thereby, reducing the level of crime. In Becker's (1968) model of crime, he identified two variables that could be adjusted to increase deterrence by affecting the certainty and the severity of punishment: the probability of apprehension and the length of sentences.

Empirical evidence appears to indicate that the level of crime is more responsive to changes in the probability of apprehension than to changes in severity.[2] Although empirical issues make these results difficult to judge, the findings do suggest an asymmetry in the effects of increased certainty of apprehension and increased severity of punishments on the crime level.[3] Anecdotally, this asymmetry has also been observed in some jurisdictions. For example, in New York City increasing the probability of apprehension by increasing the police presence appears to have drastically reduced crime[4] where threats of increased punishments had appeared unsuccessful. A similar asymmetry can be generated by an institutional feature: Prosecutors with restricted resources may make plea bargains with defendants.

In the present U.S. judicial system, few criminal cases are determined by trial. In fact, approximately 90% of cases are resolved by guilty pleas.[5] Given the severe budgetary pressure on prosecutors, this mechanism of resolving cases is viewed as an essential tool for managing large case loads. However, the implications of plea bargaining for the other objectives of the judicial system are not often considered.

Within a model which incorporates a constrained prosecutor, increased sanctions may lead to reduced deterrence. When sanctions are *increased*, the costs of trial *increase*, and this places additional pressure on the prosecutor's resource constraint.

[2]Ehrlich (1996) discusses the findings of the empirical literature on deterrence. He also discusses many issues in this empirical work including identification issues, mismeasurement of data, and difficulties in separating incapacitation and deterrence effects.

[3]See for example Ehrlich (1975, 1977), and Wolpin (1978).

[4]Although causality is difficult to ascertain, an increased police presence has been conjectured to have been an important factor in crime reductions.

[5]This figure is from United States Sentencing Commission Data. Resolved cases are defined as those dealt with by guilty plea, dismissal or trial.

As a consequence, prosecutors find themselves required to make more frequent and more generous plea bargains, which may result in a *decreased* expected sentence for any given potential criminal.

In a model with offences of varying severity, the prosecution cost associated with avoiding trial is larger for more severe offences. As a result, increasing the probability of getting a plea bargain, when a plea bargain divides the surplus generated by avoiding trial, makes more severe crimes relatively more attractive than less severe crimes. Given the cost of trial function is not very concave in the sanction level, this can lead to a substitution from less severe crime into more severe crime. This effect is driven solely by the fact that more severe crimes are more costly to try. In this model, I illustrate that in an environment with plea bargaining, increased sanctions may lead to a reduction in deterrence and/or to a substitution into more severe offences.

Becker (1968) argued that given the substitutability of certainty and severity of punishment in deterring crime, the most efficient way to reduce crime levels, if capturing criminals is costly, is to set maximal sanctions and small probabilities of apprehension. This paper suggests that in the presence of resource–constrained prosecutors who enter into plea bargains the use of maximal sanctions may not be optimal.

There is a long literature attempting to explain the non–optimality of maximal sanctions. Stigler (1970) and Mookherjee and Png (1994) argue that maximal sanctions reduce marginal deterrence, and thus, encourage criminals to commit more severe crimes or crime more intensely. Costly sanctions (Polinsky and Shavell (1979), Shavell (1987), Kaplow (1990)), errors in apprehension (Ehrlich (1975)), or imperfect information about the probability of apprehension (Bebchuk and Kaplow (1992)) can imply non–optimality. Maximal sanctions can imply overdeterrence of socially desirable activities (Mookherjee and Png (1992) and Shavell (1991)). Malik (1990) argues that criminals increase expenditure on socially wasteful avoidance activities as sentences rise. Andreoni (1991) shows that the probability of conviction may fall as sentences rise if jurors use a reasonable doubt test. Boadway *et al.* (1993) demonstrate that time–inconsistency can imply that maximal sanctions, if threatened, may not be imposed after the crime is committed.

In this paper, I abstract from all of these issues. I construct a simple model, in which prosecutors have no commitment problems, defendants' guilt is observable, and the probability of conviction is constant across sanction levels. Within this environment, I demonstrate conditions under which increasing sanctions can lead to more crime. This result relies on two assumptions. First, the prosecutor is resource constrained. By entering into plea bargains, the prosecutor can resolve cases in a less costly manner than going to trial. Secondly, the cost of going to trial is increasing in the size of the sanction. As the sanction increases, defendants exert more effort in their defense;[6] and therefore, their cases require more prosecutorial effort. In Andreoni's (1991) argument, increasing the sanction increases the standard of proof required by the jury, and this, therefore, may require more prosecutorial effort.

In this environment, I illustrate that as long as sanctions cannot be raised to the point where no one commits a crime (avoiding the limiting case), in the presence of plea bargaining under budgetary constraints, raising sanctions can lead to an increase in crime levels. I also show that increasing sanctions can lead to a substitution between offences. In fact, even a proportional increase in sanctions can imply substitution. The effect of this is that, even if increased sanctions reduce the total number of crimes being committed, it can actually imply a higher incidence of severe crime. Instead of increasing the legislated sanction, a better strategy for reducing crime may be to increase the expenditure on prosecutorial services. Increasing this expenditure would reduce the pressure to offer attractive plea bargains for administrative reasons.[7] Another method of increasing deterrence may be to increase the probability of apprehension.

In the next section of the paper, I discuss the basic model with one offence. In that environment, I consider the effect of increasing the legislated sanction and the probability of apprehension. I, then, generalize the model to one with two offences,

[6]Defensive effort is shown to be increasing in the penalty in Rubinfeld and Sappington (1987) and in the model presented in the previous paper in this thesis.

[7]This paper does not address many other reasons that plea bargains are given. For example, bargains are often given as a means of acquiring information about other criminals, or as an avenue through which prosecutors can act as advocates for defendants.

and consider the incentive of agents to substitute between these offences. Lastly, I discuss policy implications and conclude.

2 Model with One Offence

There is a measure of agents who are indexed by their criminal aptitude θ, which is uniformly distributed on the unit interval. More able agents are assumed to extract more rents from criminal activities. There is one possible criminal activity which has a payoff of $\pi(\theta)$, where $\pi'(\theta) > 0$. I assume $\pi(\theta)$ is a concave and differentiable function. If agents do not commit a crime, they receive a reservation utility which is normalized to zero.

If an agent commits a crime, she will be caught with probability $\rho > 0$.[8] For simplicity, I assume that no innocent agents are charged.[9] The prosecutor is assumed to not observe θ; and therefore, no sentences or plea bargains are conditioned on the agent's ability.[10] The expected sanction from going to trial is s_T, which is the product of the probability of conviction p_c and the legislated sanction s.[11][12] With some probability $\psi \in [0, 1]$, the defendant will be offered a plea bargain. The resulting sentence s_P is the outcome of an arbitrary bargaining game between the defendant and the prosecutor.

[8] This assumption can be generalized to $\rho(\theta)$, where the probability of apprehension is a function of the aggregate crime level. However, this does not qualitatively affect the results.

[9] Beyond simplifying this exposition, this assumption serves to reduce the incentive of agents to commit crimes. If there were a positive probability that an agent would be charged and convicted of a crime he did not commit, then the incentive for that agent to commit a crime would increase.

[10] Or equivalently, the prosecutor is directly assumed to not be able to condition plea bargains on agent ability.

[11] Andreoni's intuition (1991) can be translated into this problem as follows. If we assume that p_c is decreasing in s, then as s rises, s_T does not necessarily rise, and in fact, could fall. Therefore, increasing the legislated sanction s could reduce the agent's expected penalty from trial. As a result, in the bargaining game, the defendant's position could be improved. However, throughout this paper, I assume that the probability of conviction is fixed; and therefore, the legislator can directly affect s_T by altering the legislated sanction s. If the probability of conviction is a function of prosecutorial effort, then the increasing cost of trial can be viewed as a consequence of increased effort required to offset either a direct effect of the increased sanction on the probability of conviction or an indirect effect operating on the probability through the defendant's increased trial effort. These effort dynamics are explicitly considered in the previous paper in this thesis.

[12] Since in this model all arrested agents are guilty, it is simplest to suppose $s_T = s$ directly.

Agents are risk neutral. Their expected payoff to crime can be written

$$\pi(\theta) - \rho[\psi s_P + (1-\psi)s_T].$$

Denote as θ^*, the agent who is indifferent between committing a crime and receiving his reservation utility. This agent is implicitly defined by

$$\pi(\theta^*) = \rho[\psi s_P + (1-\psi)s_T]. \qquad (1)$$

All agents $\theta \geq \theta^*$ choose to commit a crime, and all agents $\theta < \theta^*$ do not. Therefore, $1-\theta^*$ represents the total quantity of criminals. I assume that $\pi(1) > 2\rho s_T$ guaranteeing that some agents always commit crimes (*i.e.* $\theta^* \leq 1$). I also assume that $\pi(0) = 0 < \rho s_P$, which implies that the least able agent will never commit a crime.

The prosecutor aims to maximize average expected sentences.[13] In previous work, prosecutors have been modeled as maximizing total expected sentences[14] or social welfare.[15] However, if a prosecutor maximizes total sentences, he may have an incentive to encourage more crime, and thus, to have more criminals to prosecute. A prosecutor maximizing average expected sentences can be thought of as having one of two motives. First, the prosecutor may be maximizing deterrence, and therefore, acting in the social interest. Second, the prosecutor may be self–interested and associate professional status and electoral success with being 'tough on crime'. The prosecutor is constrained by a resource constraint, which can be interpreted as the financial resources of the prosecutor's office, or the time of the prosecutor. The prosecutor's problem can be written as

$$\max_{\psi} \ \psi s_P + (1-\psi)s_T$$

$$\text{subject to} \quad M \geq \rho(1-\theta^*)(1-\psi)c(s_T) \ ; \quad 0 \leq \psi \leq 1$$

[13]Note that with this objective the prosecutor has no time–consistency problem. This is due to the fact that the prosecutor values imposing high sanctions. However, the legislator may wish to alter the punishments once the crime has been committed as in Boadway *et al.* (1993), in order to avoid imposing socially–costly punishments. If this game is repeated, this problem can be avoided. For that reason, I assume that the legislated sanction is not altered after crimes are committed.

[14]For example, the prosecutor maximizes total expected sentences in Landes (1974).

[15]As in Grossman and Katz (1983) and Reinganum (1988).

where $\rho(1-\theta^*)$ represents the number of captured criminals, M is the total budget, and $c(s_T)$ is the cost of a trial.[16] For expositional ease, the cost of a plea bargain is zero. As the sanction associated with trial increases, the cost of prosecuting the case increases (*i.e.* $c'(s_T) \geq 0$). This increased cost can be thought of as an increased effort cost resulting from either an increased standard of proof,[17] or increased defensive effort.[18] Either of these factors would induce prosecutors to exert more effort in a case going to trial in response to an increase in the sanction level.[19] Prosecutors are assumed to be 'small'. So, in deciding to plea bargain, they do not internalize the consequences of the aggregate level of plea bargaining on individual plea bargain outcomes.

3 Plea Bargaining

A plea bargain is the outcome of a bargaining game between a criminal defendant and a prosecutor. Entering into a plea bargain results in the defendant pleading guilty and receiving an agreed upon sentence which I denote $s_P(s_T)$.[20] A plea bargain divides the surplus generated by foregoing a trial. In particular, the defendant gains from having his sentence reduced by the quantity $s_T - s_P$. The prosecutor loses the same

[16] This could also be modelled with the total cost being a function of the total number of trials without qualitatively affecting the results. Although this seems a natural way to capture congestion, it significantly complicates the analysis. If the cost function is very convex, then as the number of cases taken to trial rises the plea bargaining outcome becomes more lenient (since the marginal cost of taking an additional case to trial is rising). This, in turn, can imply that given a fixed level of crime taking a larger proportion of cases to trial actually reduces an agent's expected sanction. Although this can be parameterized such that it yields the following results, I focus on a simpler construction that does not have this convex congestion cost.

[17] An increased standard reduces the probability of conviction giving an incentive for more prosecutorial effort.

[18] In particular, if prosecutors and defendants each chose trial effort levels, then increasing the sanction would always induce more effort by defendants. If this enhanced the marginal productivity of prosecutorial effort (*ie. if* $\frac{d^2 p_c}{de_p de_d} > 0$*), prosecutors would always choose to increase their effort levels in response.*

[19] These effort decisions can be modeled more explicitly; however, they considerably complicate the analysis without qualitatively changing the findings. For a more explicit treatment of these issues issues, see Roberts (1999).

[20] Often plea bargains do not dictate the exact sentence, only a charge and a sentence range. In that case, s_P can be interpreted as the expected sentence over this range. Since all agents are risk–neutral, this is an equivalent interpretation.

quantity in total sentences. However, he saves the cost of a criminal trial, $c(s_T)$.

I assume that s_P is the outcome that is reached in this bargaining game. Although I do not model this game explicitly, it is helpful to consider the sign of $\frac{ds_P}{ds_T}$. In particular, what is the effect of changing the legislated sanction on the outcome of the bargaining game? An increase in s_T can be interpreted as an increase in the bargaining power of the prosecutor, or rather as a decrease in the threat point of the defendant. As a result, the sign will generally be positive. However, the entire magnitude of the increase in s_T may not be translated into an increase in s_P. When s_T increases, the cost of taking a case to trial also increases, and as a result, the benefit of obtaining a plea bargain for the prosecutor rises. This effect works to improve the relative position of the defendant in bargaining.

For example, consider the Nash Bargaining Solution to this game where the parties evenly divide the surplus in equilibrium. The resulting plea bargain is given by[21]

$$s_P = s_T - \frac{c(s_T)}{2}.$$

In this case,

$$\frac{ds_P}{ds_T} = 1 - \frac{1}{2}\frac{dc(s_T)}{ds_T}.$$

As can be seen directly in this example, not only the magnitude of the change in the plea bargain outcome, but also its sign, is determined by the effect of the increased sanction on the cost of trying a case.[22] Throughout this paper, I focus on situations in which increasing the trial sanction increases the plea sanction. Therefore, I assume $\frac{ds_P}{ds_T} \in [0, 1]$.

4 Behaviour of Prosecutors

The prosecutor's objective function is decreasing in the proportion of people offered plea bargains, ψ, and increasing in the plea settlement, s_P. So, in equilibrium, he

[21] The simplicity of this expression follows from the linearity of agents' objectives.

[22] Recalling Andreoni's intuition (1991) that increasing the legislated sanction can reduce the probability of conviction then the expected trial penalty s_T would not necessarily rise, and in fact, could fall. Given this implied improvement in the bargaining position of the defendant it is immediate to see that $\frac{ds_P}{ds}$ could be negative.

will minimize ψ. The prosecutor minimizes the number of plea bargains in order to increase the expected sentence of each captured criminal. Suppose that his budget constraint always binds implying that $M < \rho(1-\theta^*)c(s_T)$, for any $\theta^* < 1$. Therefore, ψ cannot be set to zero, and is determined by the budget constraint. The equilibrium proportion of cases taken to trial becomes

$$(1-\psi) = \frac{M}{\rho(1-\theta^*)c}.$$

Substitution into (1) implicitly defines the critical agent θ^*,

$$\pi(\theta^*) = \rho s_P + \frac{M(s_T - s_P)}{(1-\theta^*)c}. \qquad (2)$$

4.1 Consequences of Increasing the Legislated Sanction

Suppose the legislative body wants to increase deterrence of the crime modeled above by increasing the trial sanction. Abstracting from any effect of an increased sanction on the probability of conviction (as addressed by Andreoni (1991)), the expected sanction resulting from trial s_T increases. In the following proposition, I consider the impact of such an increase on the level of crime $1-\theta^*$. When $\frac{d\theta^*}{ds_T} < 0$, an increase in the trial sanction s_T encourages more agents to commit crimes and leads to an increase in the overall level of crime $1-\theta^*$.

Proposition 1 *The effect of an increase in the legislated sanction, s_T, on the crime level, $1-\theta^*$, is determined by the following sign.*

$$\text{sign}\left[\frac{d\theta^*}{ds_T}\right] = \text{sign}\left[\frac{ds_P}{ds_T}\left[\rho - \frac{M}{c(1-\theta^*)}\right] + \frac{M}{c(1-\theta^*)} - \frac{M(s_T-s_P)}{c^2(1-\theta^*)}\frac{dc}{ds_T}\right]$$

Although increasing the trial sanction weakens the position of the prosecutor by increasing the cost of taking a case to trial, this effect may be partially offset by the increased sanction improving the bargaining position of the prosecutor. In the following corollary, I highlight two cases. First, the bargaining position of the prosecutor improves dramatically from the increase in trial sanction. As a result, the plea bargain outcome is increased by the entire amount of the increase in the

trial sanction. Second, the plea bargain outcome is unaffected by the increase in trial sanction. In this case, the bargaining benefit of an increased trial sanction can be thought of as being entirely offset by the increased cost of going to trial which decreases the bargaining power of the prosecutor. If the sign of $\frac{d\theta^*}{ds_T}$ is negative in the first case, implying the crime level has increased, then it will also be negative in the second case. If crime is increased when the bargaining power of the prosecutor is increased, it will also increase when the bargaining power of the prosecutor is unchanged.

Corollary 1 *If* $\frac{ds_P}{ds_T} = 1$, *then*

$$\text{sign } [\frac{d\theta^*}{ds_T}] = \text{ sign } [\rho - \frac{M(s_T - s_P)}{c^2(1-\theta^*)}\frac{dc}{ds_T}].$$

If $\frac{ds_P}{ds_T} = 0$, *then*

$$\text{sign } [\frac{d\theta^*}{ds_T}] = \text{ sign } [1 - \frac{s_T - s_P}{c}\frac{dc}{ds_T}].$$

In the large bargaining power effect case (where $\frac{ds_P}{ds_T} = 1$), when $\frac{M}{c(1-\theta^*)}$ is large relative to ρ, there are few plea bargains (this follows from the budget constraint). Therefore, any increase in the cost of trial has a large effect because of the large number of cases being tried. The larger this effect the higher the probability of the increase in sanction leading to an increase in the crime level.

The probability that increasing the sanction will increase the crime level is increasing in the distance between the trial and plea bargain outcomes, increasing in the marginal effect of the increased sanction on the cost of trial, and decreasing in the cost of trial. The expression $\frac{s_T - s_P}{c}$ is a measure of the sacrifice made by a prosecutor in agreeing to a plea. The larger $s_T - s_P$ is the larger the sentence reduction received. The larger c is the larger the benefit from avoiding trial. When this sacrifice term is large, the effect of an increased sanction is more likely to lead to an increased level of crime.

4.2 Consequences of Increasing the Probability of Apprehension

Increasing the probability of apprehending criminals is often considered a substitute for increasing sanctions. However, in the context of this model, they can produce drastically different results. Assuming that prosecutors cannot enter into plea bargains to acquire information, prosecutors cannot affect the probability of capture. Given this and if the probability of apprehension has no impact on the plea bargain outcome, prosecutors' budgetary considerations have no impact on this probability. As a result, increased monitoring does not interact with these constraints in the same way that increased sanctions do. Consequently, the intuition that increased monitoring increases deterrence is maintained.

Proposition 2 *Increasing ρ, the probability that a criminal is caught, reduces the level of crime.*

However, the increased number of apprehensions may have an impact on the plea bargaining outcome. In particular, increasing apprehensions may result in congestion which forces prosecutors to offer more appealing plea bargains. Therefore, $\frac{ds_P}{d\rho} < 0$. If this is the case, then

Proposition 3 *If increasing the probability of apprehension results in more generous plea bargains, then the effect on the crime level is determined by the following expression.*

$$\text{sign}\Big[\frac{d\theta^*}{d\rho}\Big] = \text{sign}\Big[s_P + \frac{ds_P}{d\rho}[\rho - \frac{M}{c(1-\theta^*)}]\Big]$$

This expression is positive implying increased deterrence unless the bargaining effect is very large. In fact, $\frac{d\theta^*}{d\rho} > 0$ if $\frac{d\rho s_P}{d\rho} > 0$. So, the crime level will be reduced by increasing the probability of apprehension only if the bargaining effect is large enough to entirely offset the direct effect of the increased probability of conviction.

This suggests that given the institutional features that are observed in the current judicial system; increased sanctions and increased monitoring may not be thought of

as close substitutes. This observation can perhaps partially explain the success of increased police monitoring in New York City and the relative failure of increased sentences.[23]

4.3 Consequences of Increasing Prosecutorial Resources

Increasing prosecutorial resources, can deter crime by slackening the constraints on prosecutors. As a result, prosecutors enter into fewer plea bargains. This implies a higher expected sentence for potential criminals. If $\frac{ds_P}{dM} > 0$, there is also a bargaining effect. Given the prosecutor has to offer fewer pleas for budgetary reasons, his bargaining power or threat point is increased in any bargain he does undertake. Therefore, he can offer less attractive pleas.

Proposition 4 *Given* $\frac{ds_P}{dM} \geq 0$. *Increasing the prosecutorial resources* M*, reduces the level of crime.*

Although increasing both sanctions and monitoring may have important effects on the level of crime, this proposition also suggests an alternative strategy for reducing the level of crime: increase the resources and bargaining power of prosecutors.

4.4 Discussion of Optimal Policy

If the social planner chooses a value of M sufficiently large that the prosecutor's budget does not bind, then the prosecutor would take every case to trial, and the model would collapse to the basic Becker framework. The planner could then achieve the desired level of deterrence by adjusting the probability of apprehension and the sanction. If the sanction is costless to impose, the optimality of maximal sanctions re–emerges. Since maximal sanctions will deter if every case goes to trial, the budget required is falling in the sanction. However, given the current judicial system, where less than 10% of cases go to trial, the budget necessary to take every case to trial seems infeasible.

[23] In particular, the increased narcotics sentences implied by the mandatory minimum provisions have not appeared to reduce narcotics offences.

Alternatively, one could reduce the cost of trial directly. For example, by reducing the burden of proof. However, this possibility has other obvious disadvantages in a system where guilt is to be ascertained.

If increasing the prosecutorial budget is too costly relative to altering the probability of apprehension, then the government would leave the prosecutor constrained and focus deterrence efforts on making punishment more certain. In this constrained environment, maximizing sanctions may not be optimal.

5 Model with Two Offences

Increasing sanctions may cause more severe crimes. This is even the case if the total crime level falls. Since different crimes imply different plea bargaining concessions, the effect of increasing sanctions is not uniform across the plea bargaining outcomes associated with different offences. The implications of this are most easily illustrated in a simple model with two crimes, $\alpha \in \{H, L\}$, where H denotes the more severe crime. The sanctions for these offences are denoted s_H and s_L, where $s_H \geq s_L$. The cost associated with taking a case to trial is $c(s)$, where $c'(s) > 0$. Therefore, it is more costly to try more severe offences.

I denote the payoff to crime as $\pi_\alpha \theta$, where $\pi_H > \pi_L$. Each agent chooses either to commit crime H or L or to not commit any crime. For simplicity, I assume that the probability of capture, ρ, is constant across crimes.[24] Assume an interior solution so that agents $1 - \theta_2$ commit the more severe crime, and agents $\theta_2 - \theta_1$ commit the less severe crime. In equilibrium, the prosecutor chooses the probability of plea bargaining with each type of case, ψ_H and ψ_L, subject to a budget constraint given by

$$M = \rho(1 - \theta_2)(1 - \psi_H)c(s_H) + \rho(\theta_2 - \theta_1)(1 - \psi_L)c(s_L),$$

where M is the total prosecutorial budget. Throughout this section of the paper, I assume that plea bargains are determined by Nash bargaining for simplicity. As a

[24] This assumption does not qualitatively affect the results. However, if the probabilities of apprehension are very different, it is unlikely that both crimes will be committed in equilibrium.

result, the plea bargain outcome is $s_\alpha^p = s_\alpha - \frac{1}{2}c(s_\alpha)$. The *ex ante* expected sentence is given by $Es_\alpha = (1-\psi_\alpha)s_\alpha + \psi_\alpha s_\alpha^p = s_\alpha - \frac{1}{2}\psi_\alpha c(s_\alpha)$.

The prosecutor maximizes the average expected sentence as before. I assume that the prosecutor does not internalize the effect of his policy on the aggregate crime level. Therefore, the prosecutor's problem is given by

$$\max_{\psi_L,\psi_H} \frac{\rho(\theta_2-\theta_1)Es_L + \rho(1-\theta_2)Es_H}{\rho(1-\theta_1)}$$

$$\text{subject to } M = \rho(1-\theta_2)(1-\psi_H)c(s_H) + \rho(\theta_2-\theta_1)(1-\psi_L)c(s_L).$$

The associated first order conditions imply that

$$\psi_H = \psi_L.$$

5.1 Proportionally Raising Sanctions

If the penalties for both offences are increased proportionally, the crime level for either offence may fall or rise. These implications are easily seen as a simple generalization of the results from the previous one offence case. However, adding a second offence adds some new insights. In particular, with more than one crime, agents can substitute not only in and out of crime but also between different offences. In fact, increasing sanctions proportionally can reduce the total amount of crime, but still increase its average severity. Some agents committing the less severe crime may choose to not commit an offence and others may choose to substitute into the more severe offence. This is driven by the fact that the plea bargaining process may serve to make the less severe offence less attractive relative to the other two options.

The binding budget constraint implies that when sentences are increased more plea bargains will be offered unless the effect of the increased sentences is to drastically reduce the level of crime in which case the number of plea bargains will fall.

Lemma 1 *Proportionally increasing sanctions always leads to an increase in the probability of receiving a plea bargain, ψ, if the following conditions hold:*

$$\pi_H - \pi_L \leq \frac{1}{2}c(s_L)\frac{c'(s_H)-c'(s_L)}{c(s_H)-c(s_L)}, \text{ and}$$

$$\pi_L c(s_H)(c(s_H) - c(s_L)) \geq \pi_H(c(s_H)^2 - c(s_L)^2)$$

This condition is sufficient to guarantee that the cost effect is not outweighed by the deterrence effect. For the remainder of the argument, I restrict consideration to the case where $\frac{d\psi}{ds} > 0$.

The effect of a proportional increase in sanctions on the total level of crime (*ie.* on the θ_1 margin, where $1 - \theta_1$ represents the total crime level) is ambiguous as in the above case with only one offence. However, the effect on the level of severe crime, $1 - \theta_2$, is clearer.

Proposition 5 *Proportionally increasing sanctions leads to more severe crime if*

$$[\frac{d\psi}{ds}(c(s_L) - c(s_H)) + \psi(c'(s_L) - c'(s_H))] < 0.$$

The sign of this expression is primarily determined by the curvature of the cost function. If the cost function is very concave, then the rate of increase of plea bargaining surplus may be sufficiently slower for the more severe offence to discourage substitution. In fact, with a sufficiently concave function, an increase could encourage substitution into the less severe offence. However, if the cost function is weakly convex then as the sanction rises the surplus associated with the severe offence relative to the less severe offence also rises encouraging substitution.

Corollary 2 *Given that a sentence increase has led to more plea bargains, proportionally increasing sanctions leads to more severe crime if the trial cost function is weakly convex.*

As long as the cost function is not too concave, increasing sanctions will lead to an increase in crime severity as agents substitute from the less severe crime into the more severe crime. This result is again driven by the increasing cost of punishment. The foregone cost of trial is greater for the more severe offence. This surplus is then divided, making the more severe crime more appealing to potential criminals. In this manner, the increased sanction results in a more appealing plea bargain for the severe crime. This effectively reduces the expected sanction associated with that offence and entices more agents to undertake that offence.

5.2 Policy Implications

Stigler (1970) argued that marginal deterrence, deterring people from more severe crimes, was one argument against the universal use of maximal sanctions. In a similar vein, Mookherjee and Png (1994) show that marginal deterrence is also important on the intensity of crime margin. However, in both cases, the implied optimal schedule of sanctions, sanctions which increase in offence severity or intensity, should then be shifted upward maintaining the marginal relationship between sanction levels for different offences. This simple model shows that proportional increases in sanctions cannot guarantee that these marginal relationships are maintained. Rather, the above result suggests that when increasing sanctions, the appropriate marginal relationships to consider are those between the expected penalties as opposed to the legislated sanctions for different offences.

6 Conclusion

Given the observation that some agents receive plea bargains,[25] this model suggests that increasing sanctions may lead to increased incentives to commit crime. This reduction in deterrence is driven by increasing trial costs and binding budgets.

A more effective method of deterring crime may be to increase the budgets and number of prosecutors or to even decrease the legislated criminal sanctions. Another effective way to increase deterrence may be to increase the probability that criminals are apprehended.

Mandatory minimum sentences, which are currently used for many drug and violent crimes, may serve as a method of increasing criminal sanctions without increasing

[25] Although it is difficult to discern how many plea bargains are made solely for the purpose of conserving resources, the large proportion of cases resolved in this manner suggest that this is an important factor. This model lacks important features of the judicial process which also motivate the use of plea bargains. Most importantly, I do not incorporate risk aversion which is a primary reason why prosecutors and defendants reach agreements. In this model, I also do not consider the problems of asymmetric information about agents' guilt or innocence. If some agents are innocent, they might want to go to trial in order to separate themselves from guilty defendants as in Grossman and Katz (1983). Prosecutors and defendants could also possess different information about the strength of the prosecutor's case, as discussed in Reinganum (1988).

the costs of trial. In these cases, the cost of trial may not be increasing in the sanction since the quantity of evidence that must be prepared and presented is limited to only a few dimensions (*i.e.* the quantity of drugs possessed).

"Three strikes" laws, like those in California, where a third felony conviction results in life–imprisonment may reduce the benefits of plea bargaining to career criminals. Even if the defendant is risk–neutral, and the expected sentence from plea bargaining is less than that of going to trial, the defendant may not wish to plead guilty to a felony, and may rather risk being sentenced to a much more severe sanction in the hope of being found innocent. The benefit to being found innocent is now much higher than in a system where punishment increases more slowly following previous guilty decisions.

7 References

Andreoni, J. (1991). "Reasonable Doubt and the optimal magnitude of fines: should the penalty fit the crime?," *RAND Journal of Economics*, 22(3): 385-395.

Bebchuk, L.A. and L. Kaplow (1992). "Optimal Sanctions when the Probability of Apprehension Varies Among Individuals," NBER Working Paper, 4078.

Becker, G.S. (1968). "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76: 169-217.

Boadway, R.W., N. Marceau and M. Marchand (1993). "Time–Consistent Criminal Sanctions," Institute for Economic Research Working Paper, Queen's University, 883.

Ehrlich, I. (1996). "Crime, Punishment, and the Market for Offenses," *Journal of Economic Perspectives*, 10(1): 43-67.

Ehrlich, I. (1975). "The Deterrent Effect of Capital Punishment: A Question of Life and Death," *American Economic Review*, 65(3): 397-417.

Grossman, G.M. and M.L. Katz (1983). "Plea Bargaining and Social Welfare," *American Economic Review*, 73: 749-757.

Landes, William M. (1974). "Legality and Reality: Some Evidence on Criminal Procedure," *Journal of Legal Studies*, 3: 287-336.

Lewis, D.E. (1986). "The General Deterrent Effect of Longer Sentences," *British Journal of Criminology*, 26(1): 47-62.

Malik, A.S. (1990). "Avoidance, screening and optimum enforcement," *RAND Journal of Economics*, 21(3): 341-353.

Mookherjee, D. and I.P.L. Png (1992). "Monitoring vis-a-vis Investigation in Enforcement of Law," *American Economic Review*, 82(3): 556-565.

Mookherjee, D. and I.P.L. Png (1994). "Marginal Deterrence in Enforcement of Law," *Journal of Political Economy*, 102(5): 1039-1066.

Polinsky, A.M. and S. Shavell (1979). "The Optimal Tradeoff between the Probability and Magnitude of Fines," *American Economic Review*, 69(5): 880-891.

Reignamun, J.F. (1988). "Plea Bargaining and Prosecutorial Discretion," *Ameri-*

can Economic Review, 78: 713-728.

Roberts, J. (1999). "Plea Bargaining and 'Truth-Telling'," mimeo.

Rubinfeld, D.L. and D.E.M. Sappington (1987). "Efficient awards and standards of proof in judicial proceedings," *RAND Journal of Economics*, 18(2): 308-315.

Shavell, S. (1987). "The Optimal Use of Nonmonetary Sanctions as a Deterrent," *American Economic Review*, 77(4): 584-592.

Shavell, S. (1991). "Specific versus General Enforcement of Law," *Journal of Political Economy*, 99: 1088-1108.

Stigler, G.L. (1970). "The Optimal Enforcement of Laws," *Journal of Political Economy*, 78: 526-536.

8 Appendix

Proof of Proposition 1: Totally differentiating (2) yields

$$\Big[\pi_\theta - \frac{M(s_T - s_P)}{c(1-\theta^*)^2}\Big]\frac{d\theta^*}{ds_T} = \frac{ds_P}{ds_T}\Big[\rho - \frac{M}{c(1-\theta^*)}\Big] + \frac{M}{c(1-\theta^*)} - \frac{M(s_T - s_P)}{c^2(1-\theta^*)}\frac{dc}{ds_T} \quad (3)$$

Since the budget constraint binds, $M < \rho(1-\theta^*)c$. This implies that

$$\Big[\pi_\theta - \frac{M(s_T - s_P)}{c(1-\theta^*)^2}\Big] > \Big[\pi_\theta - \frac{\rho(s_T - s_P)}{(1-\theta^*)}\Big] = \Big[\frac{\pi_\theta(1-\theta^*) - \rho(s_T - s_P)}{(1-\theta^*)}\Big].$$

Concavity of π implies that the above term is greater than

$$\Big[\frac{\pi(1) - \pi(\theta^*) - \rho(s_T - s_P)}{(1-\theta^*)}\Big]. \quad (4)$$

Recalling that $\pi(1) = \rho[\psi s_P + (1-\psi)s_T]$, and that $\pi(1) > 2\rho s_T$, implies (4) is greater than zero. This implies that

$$\Big[\pi_\theta - \frac{M(s_T - s_P)}{c(1-\theta^*)^2}\Big] > 0.$$

Therefore, sign$\Big[\frac{d\theta^*}{ds_T}\Big]$ is equal to the sign of the right hand sign of (3).

Proof of Proposition 2: Totally differentiating (2) yields

$$\Big[\pi_\theta - \frac{M(s_T - s_P)}{c(1-\theta^*)^2}\Big]\frac{d\theta^*}{d\rho} = s_P > 0$$

The result follows from the concavity of $\pi(\theta)$ and the binding budget.

Proof of Proposition 3: Totally differentiating (2) yields.

$$\frac{d\theta^*}{d\rho} = \frac{1}{\pi_\theta - \frac{M(s_T - s_P)}{c(1-\theta^*)^2}}\Big[s_P + \frac{ds_P}{d\rho}[\rho - \frac{M}{c(1-\theta^*)}]\Big]$$

Proof of Proposition 4: Totally differentiating (2) yields

$$\Big[\pi_\theta - \frac{M(s_T - s_P)}{(1-\theta^*)^2 c}\Big]\frac{d\theta^*}{dM} = \Big[\rho - \frac{M}{c(1-\theta^*)}\Big]\frac{ds_P}{dM} + \frac{(s_T - s_P)}{(1-\theta^*)c}$$

Since $\frac{ds_P}{dM} \geq 0$, $\frac{d\theta^*}{dM} > 0$ given the concavity of $\pi(\theta)$ and the binding budget.

Proof of Lemma 1: Substituting equilibrium crime levels into the prosecutor's budget constraint and differentiating with respect to sentence levels yields that

$$\frac{d\psi}{ds}\Big[(1-\theta_2)c(s_H)+(\theta_2-\theta_1)c(s_L)+$$

$$\frac{\frac{1}{2}\rho[\pi_L c(s_H)(c(s_H)-c(s_L))-\pi_H(c(s_H)^2-c(s_L)^2)]}{\pi_L(\pi_H-\pi_L)}\Big]$$

$$=(1-\psi)\Big[(1-\theta_2)c'(s_H)+(\theta_2-\theta_1)c'(s_L)+$$

$$\frac{\rho[(\pi_H-\pi_L)(\frac{1}{2}\psi c'(s_H)c(s_H)+c(s_L)-c(s_H))+\frac{1}{2}\psi c(s_L)(\pi_L c'(s_H)-\pi_H c'(s_L))]}{\pi_L(\pi_H-\pi_L)}\Big]$$

Given

$$\pi_H-\pi_L\leq\frac{1}{2}c(s_L)\frac{c'(s_H)-c'(s_L)}{c(s_H)-c(s_L)},$$

the numerator is positive.

That the denominator is also positive is guaranteed by

$$\pi_L c(s_H)(c(s_H)-c(s_L))\geq\pi_H(c(s_H)^2-c(s_L)^2).$$

Proof of Proposition 5: The agent indifferent between the two crimes is given by

$$\theta_2=\rho[\frac{s_H-s_L+\frac{1}{2}\psi(c(s_L)-c(s_H))}{\pi_H-\pi_L}].$$

Proportionally increasing sanctions yields

$$\frac{d\theta_2}{ds}=\frac{\frac{1}{2}\rho}{\pi_H-\pi_L}[\frac{d\psi}{ds}(c(s_L)-c(s_H))+\psi(c'(s_L)-c'(s_H))].$$